The Application of Intelligent Systems to Finance and Business

James Viner

Thesis submitted for the degree of
Doctor of Philosophy
of the University College London

October 1998

Department of Computer Science
University College London

Abstract

Intelligent systems have proved very effective in many business applications, however there is little understanding of the relationships between the techniques and the application domains. Through applying genetic algorithms and neural networks, two powerful general purpose techniques to the difficult problems of economic forecasting and financial trading, this thesis investigates the connections between the nature of the application and the chosen intelligent technique. Four experiments have been carried out to investigate this problem:

Residual Value Forecasting with Lex Vehicle Leasing: One method of setting vehicle hire charges requires accurate forecasts of the second hand value of vehicles in 3 or 4 years time. A synthetic depreciation series was constructed for proof-of-concept purposes. Neural network forecasting models were compared against linear regression benchmarks and it was found that they have comparable performance. Developments to this forecasting scheme are proposed.

Intelligent Trade Filtering with Sabre Fund Management: This project is an attempt to capture, reproduce and extend expert trading knowledge from a history of pattern based trading. Genetic algorithms were used to find rules that capture the symbolic relationships between the observed market state (pattern type, quality etc.) at trade entry and probable trade outcome. The GA found several rules that are deterministic to 95% confidence.

Continually Adaptive Trading Systems Design: Financial markets are in a constant state of change. Genetic algorithm-style operators were used on a population of trading system descriptions to generate new trading strategies on-the-fly which are evaluated on a rolling basis by their recent trading performance. In conjunction with the over-night loan market, the system could make super-LIBOR returns, although the impact of this result on theories of market efficiency is unclear. This system is unable to trade the FTSE index effectively, an observation consistent with the theory that the information set of the FTSE is too large for it to be out-performed.

Genetic Algorithm Trading System Induction with the European Bank for Reconstruction and Development, a bank that speculatively trades government bond futures markets. Genetic algorithm rule induction was used to automate trading system innovation and profitability tests were carried out in all relevant markets. The results are positive but mixed. The system makes higher returns on longer maturity markets, and the presence of this effect over the US Treasury bond markets is demonstrated to a confidence of 86%. The system was also tested on copper and gold futures. The system was then modified to give a new technique for assessing the change in the character of financial markets.

The principal scientific contributions to come from this thesis are: i) genetic algorithm rule induction is a powerful and effective technique for finding empirical models. It is particularly suitable for business problems as the experimenter has control over the representation and the resulting models are transparent; ii) novel results from the financial experiments present evidence both for, but primarily against, the Efficient Market Hypothesis; iii) that the decision to use a specific technology should be taken after the content of the available data has been investigated; iv) the proposition of a new technique for tracking changes in nature of financial markets using the trading rule induction engine; v) the design and operational analysis is given for the first known continuously adaptive trading engine; vi) the first known comprehensive operational analysis of a GA rule induction based trading system; vii) the first known public domain data intensive analysis of the vehicle resale market.

Acknowledgements

The story of this thesis has been an interesting tale, and many people have played important roles

UCL:
Many thanks to Phil Treleaven for the supervision, advice, guidance and help that he has given over the years.
Thanks also to Suran Goonatilake for his supervision and direction, and to Peter Bentley for his vital help in the final stages.
Thanks must go to my past and present colleagues for their friendship and contributions: Alasdair, Laura, Hugh, Pouneh, Raghbir, Paul, Adil, Sanj, Surayya, Tony, Jason and Konrad.
Thanks also to the Graduate School for a little extra money, and that busy legal team at UCLi.

Lex Vehicle Leasing:
Thanks to Nick, Chris, Steve, Martin, and Hemant for getting it all started.

Sabre Fund Management:
Thanks to Spish for the introduction to the financial markets.

European Bank for Reconstruction and Development:
Thanks to Mark Cutis, Asoka Selvarajah and Jaakko Karki for one of the most character building episodes of my life, and everyone else for the part they played: Olja, Dave, Benedicte, Serra, Andrey, Daniil, George P., George H. and Bart.

Thanks to my family and friends for their support and companionship: Mum, Dad, Helen, Grandma, the Islington (& related) lot, innumerable diving people and all my old pals from Warwick.

And Finally…
Thank you to the directors of Lex Vehicle Leasing for making it all possible.

Table of Contents

List of Figures

List of Tables

Chapter 1:

Introduction

This chapter presents the motivations for researching intelligent systems, and once the key features of intelligent systems have been described and discussed, the four experiments presented in this thesis are introduced. The prerequisites that exist for the successful deployment of intelligent systems are also examined. The chapter concludes with the motivations, goals and contributions of this work and the thesis structure.

1.1 Introduction

This thesis investigates four *intelligent system* experiments designed to convey competitive advantage from the analysis of data. There are a range of intelligent technologies available and a myriad of possible applications within business contexts. This thesis examines the relationships between the nature of the business problem, the intelligent system solution and the effectiveness of the resulting implementation, with a view to discovering the knowledge that enables intelligent system engineers to construct the most effective systems.

The technologies that will be focussed upon are genetic algorithms and neural networks. These have been the basis of a great deal of research and have been demonstrated to be extremely effective and powerful techniques. They will be tested in economic forecasting and speculative trading applications. These are some of the most difficult and taxing business problems that exist.

Pitting powerful, non-linear techniques against difficult problems will strain both the technologies and conventional understanding of markets. It is in these situations where much can be learned about the applications, the techniques and the relationships that exist between them.

1.2 Five Key Features of Intelligent Systems

The field of intelligent systems is diverse: there are a broad range of techniques that have been used for addressing thousands of problems, but a number of recurrent themes emerge: they are often capable of learning and adapting, they are flexible, can often be attributed with explanatory power and be capable of discovering new knowledge. As it is these five attributes that make intelligent systems valuable and worthy of research, a brief overview of each will be presented, before introducing the actual experiments carried out in the course of this thesis.

1.2.1 Learning

Learning is the process of knowledge acquisition. This is arguably the most important strength of many intelligent systems deployed in financial and business applications. Many intelligent systems, but especially those based on neural networks and genetic algorithms, can learn by example. This involves giving the system examples to learn from, and then the system makes internal modifications to reflect the new data it has been exposed to. This can be thought of as building an internal model that is consistent with the submitted examples. For this reason, is it vital to give the machine *representative* samples, or the model that the system constructs will not correspond strongly to reality. One of the main goals of intelligent systems research is to construct systems that can generalise well once the learning or training has been completed.

Particular problems exist when trying to automate expert knowledge. Experts are often unwilling or unable to concisely and consistently express their knowledge. Indeed, their knowledge may not be complete, consistent, or even correct. However, from records of expert behaviour, it is often possible to construct intelligent systems that can learn the knowledge implicit in the decisions the expert has made. Depending on the type of intelligent system, it may then be possible to decode the machine's internal state to analyse the information content of the expert's knowledge that has been transferred to the machine.

1.2.2 Adaptation

A further benefit of intelligent systems is that they can *adapt* to new situations. If the model the intelligent system has constructed begins to slide out of date, then it is often a simple matter to bring the system back up to date by simply re-training the system with more up-to-date information. Moreover, it is possible to design intelligent systems so that they are constantly absorbing new information and modifying their models accordingly. This is another important benefit of intelligent systems: they have no loyalty to the models they find, and so they will not retain aspects of models that are out of date simply because they have been a component of the model for a considerable period of time.

There can be less drastic reasons for shifting the knowledge base: for instance, in credit scoring, the risk-adversity of lending institutions and the desires of those seeking credit are very different in times of prosperity to those in recession.

This adaptability is of particular relevance in the financial markets. For instance, immediately prior to the end of a primary price trend, there is a buying frenzy where the price is bid up by those attracted to a security that is rapidly rising in price. When there are no market participants left who wish to buy, the price drops rapidly as there are no bids to push the price up, and at the same time, there are lots of dealers who are trying to sell securities that are (now) dropping rapidly in price. This has the effect of pushing the price down lower. When this transition from buying to selling occurs, it is clearly important for an intelligent system to react rapidly and appropriately. In extreme cases, it need only take a few seconds for the securities' price to go 'limit-down' and for trading to be suspended in an effort to control the rate of collapse of price.

1.2.3 Flexibility

The flexibility of intelligent systems shows itself in two ways: many of the techniques are intrinsically general purpose, and they are robust in the face of incomplete, noisy or inconsistent data.

In overview, many intelligent techniques are very similar. Once the problem has been specified, some part of the intelligent system constitutes an optimisation procedure of either maximising fitness or minimising errors, or a search for good solutions. A consequence of this is that if a business or financial problem can be expressed in terms of the value of solutions or the errors generated by solutions, then intelligent systems can be used to address the problem - hence their general purpose nature. Indeed, this is one of the few prerequisites for the use of intelligent systems.

The robustness of intelligent systems stems from several sources. Traditionally, computers have been associated with concepts such as “yes” and “no”, “black” and “white”, and not with an inability to reason with imprecise data. Fuzzy logic can not merely cope with concepts such as “fairly tall” or “too fast”, but also reason and infer from such statements. This is of significant value if, for instance, a potential customer in a database can be assessed as being “likely” or “very unlikely” to respond to a directly mailed advert.

1.2.4 Explanation

Some intelligent systems can be accredited with having explanatory power. It is possible to use genetic algorithms to find rules that can explain business and financial phenomena. In credit scoring, if a set of intelligently induced rules reject a loan application, and if the customer demands to know why, then the train of reasoning that led to the application’s rejection can easily be examined and possibly conveyed to the applicant. There are also applications in fraud detection. The daily volume of trades in any single exchange is far greater than could easily be checked by any reasonable sized watchdog. However, if a series of trades appear suspicious to the intelligent system, then exchange officials can be alerted, and the line of reasoning that led to the alert being raised can be examined.

1.2.5 Discovery

One of the strengths of intelligent systems is that they can construct their own models of business and finance that may be more appropriate to the problem than one designed by hand. Knowledge discovery is the process of examining data to find non-trivial relationships that are capable of adding value.

Wal-Mart[Veri94] is a chain of hyper-market stores in the USA. It used parallel database mining techniques on gigabytes of P.O.S.(point of sale) information to examine whether there were any particularly strong correlations between the sale of one good or commodity and the sale of another[Veri94]. There were some obvious facts to come out of this, such as if someone buys baby-food then they probably buy nappies. However, the machine also made the perhaps not terribly alarming, but nevertheless interesting and unexploited discovery that if young males bought nappies then they also bought beer with high probability. This observation led to the placing of beer in the child-care section of some Wal-Mart stores, and this was soon followed by a significant increase in beer sales in those stores. After this trial and a second analysis of the P.O.S. data on the sales of nappies and beer, it was decided to move a rack of crisps and other condiments into the child-care section next to the beer. Unsurprisingly perhaps, revenues were increased again.

1.3 Business Problems

Due to these five factors, intelligent systems have (rightly) been perceived as general purpose technologies, and as a result of this, their range of application is very broad. This is confirmed by the range of the applications that have been used as illustrative examples so far.

The research described in this thesis has been undertaken in conjunction with a number of industrial partners to gain broad, direct and thorough exposure into intelligent systems that are developed for use in industry. As a result of this, four separate intelligent system experiments have been carried out. Each of these will now be introduced.

1.3.1 Neural Networks for Residual Value Forecasting

This work was carried in association with Lex Vehicle Leasing. The problem was to forecast the level of depreciation of vehicles over the hire period - in essence, to forecast its residual value. This is an important company process because the single largest component of the total hire charge of a vehicle is the recouping of the (estimated) depreciation of the vehicle over the hire period. The company was keen to use neural networks for this task, and so neural network software was developed to utilise the company's large vehicle pricing data-base for forecasting vehicle's residual values. These neural network models were compared and contrasted with linear regression benchmarks in order to assess them.

1.3.2 Genetic Algorithms for Intelligent Trade Filtering

This research was undertaken in conjunction with Sabre Fund Management. The task was to assess which types of charting trades, if any, work most reliably. Charting is one method of trading in financial markets that consists of trying to find and exploit repeating patterns of market movement. Debate exists whether this approach to trading is worthwhile or not, but this is not an issue here. The data consisted of a record of the company's trading performance over recent years, and a comprehensive record of which patterns the expert trader thought were active at the time of trade entry. This project involved the use of genetic algorithms to find rules that can connect sub-sets of expert knowledge to market conditions. In other words, the genetic algorithm is trying to find rules that describe those parts of the expert's knowledge that are validated through their performance in the markets.

1.3.3 Continually Adaptive Trading Engine

This system again uses genetic rule induction for finding trading rules, but unlike the previous experiment, it does not attempt to find pockets of deterministic behaviour in the markets. Instead, the system attempts to evolve along with changes in the nature of the equity markets it faces, by rapidly assembling appropriate models for the current markets with genetic operations. This is very similar to the situation faced by living creatures - as the environment changes, whether due to climatic change, predator-prey arms races or industrial waste, species must evolve in order to survive.

1.3.4 Bull-Bear Trading Engine

This work was funded by the European Bank for Reconstruction and Development. It used genetic rule induction to find trading rules that describe exploitable pockets of predictability in the Government bond futures markets. The work later developed into an autonomous rule based trading system and a new method of assessing the stationarity of markets. Moreover, results from experiments with the Bull-Bear engine have proved inexplicable in terms of standard economic theory, but are consistent with a non-linear view of economics.

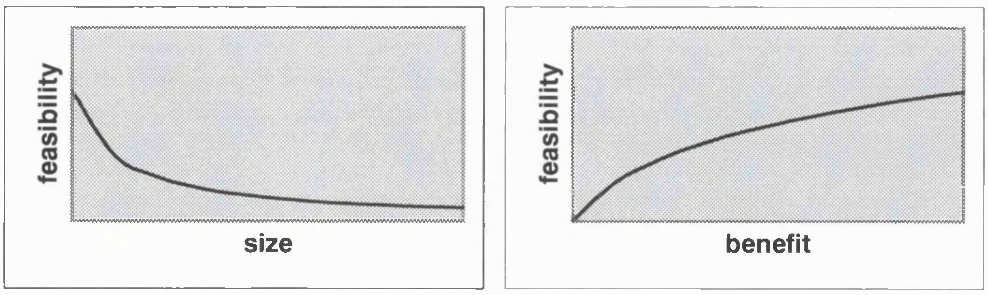
1.4 Prerequisites

For intelligent systems to be a likely candidate as a solution to a business problem, a number of prerequisites exist that must be fulfilled before it is likely that the solution will be not merely effective, but viable at all. Some of the these prerequisites will be examined in the next section.

1.4.1 Information and Data

The viability of each of these projects is partially determined by the availability of data. If data is unavailable, many of these projects simply could not have been undertaken. The availability of data pertinent to a company's operation in a computer-friendly format directly enables the analysis of such data with tools that were previously unavailable owing to their data intensive nature. The feasibility of computerising past paper records diminishes as the required data set grows, but is also clearly a function of the benefits of applying the computational analysis in the first place. See Figure 1.1. It is also dependent on the tolerance of the system to errors in the entered data set, as a law of diminishing returns exists between that money spent on data entry and the number of errors that escape detection and correction. As more core business processes are computerised, the records that historic data sets are built from are already in a computer friendly format. This also nearly eliminates errors in the data due to incorrect data entry.

Figure 1.1: Feasibility of computerising records



1.5 Use of Intelligent Techniques

Many intelligent techniques, but especially neural networks, are now sufficiently mature for their widespread application within the business environment. However, the technology may be mature enough - and indeed there are a number of very usable and effective packages for PCs around, such as 4-Thought, but the field of neural networks is complex and there is a very real danger that lay-people will attempt to use these tools in a way that is inappropriate without realising it. This exposes the technology to criticism that is misplaced, due to the user's unrealistic expectations, or reasonable and achievable, but *unrealised* expectations. Examples of inappropriate use are a failure to specify the model in a sensible manner in the first place, trying to extract information that simply isn't there, or simply using data in an inappropriate way. From the author's experience, it is common for users of neural network software to have a number of erroneous beliefs, or simply not be aware of the following important points:

- It can be important to represent the problem in a manner that is easy for the system to find solutions. Two representations of musical tones could be a sequence of wave samples every 1/10000th of a second, or a coding of musical notation on a stave. Which was more appropriate would be dependent on the application.
- Pre- and/or post- processing of input information can be important to enhance the information density of the input data. It is often through this route that domain knowledge can be incorporated that has no other easy route into the model. For example, in a fraud detection application, the start and claim dates could be inputs to a system. However, there is nothing intrinsically fraudulent about individual dates - it is the intervals that are important. Moreover, this reduces the complexity of the problem space, reduces the input complexity and prevents the system making spurious relationships between these and other inputs.
- It is important to have reasonable performance expectations, given a certain data set. It is not uncommon for managers to have wildly optimistic expectations of what is achievable with a certain data set, both in terms of accuracy and what information can be extracted from data. It is often difficult to form reliable judgements of what the likely performance of the system will be when one is conducting exploratory research.
- Specifying an appropriate network size for the amount and stability of the data. There are simple rules of thumb for estimating the appropriate size of network for a certain amount of data. Experiments can be conducted to find the optimal network size and topology - i.e. construct the network that gives the best out-

sample performance, but it is simply not possible to use and then re-use out-of-sample data to find the best network in this way. This leads on to the next point.

- How (rapidly) out-sample data can migrate in-sample data. If one repeatedly uses out-sample data to assess the progress of the system's development, then the once out-sample data becomes an implicit part of the model building process, and can no longer be considered as genuinely out-of-sample.
- The dangers of over fitting. The more training an intelligent system undergoes, the better it will know the *training set*. This is not what is required - what is required are systems that can act appropriately and generalise in the future, not ones that work well on training data and then fail to operate thereafter. The usual cause of over-fitting is the use of a model that is too complicated for the amount of data available.
- The dangers of over extrapolation. This too is related to the point above - if the model is applied to areas beyond the boundaries of where it is applicable, then the results inferred from such modelling will clearly be unreliable.

1.6 Thesis Aims and Contributions

This thesis has a number of goals:

1. The primary research goal is to identify causal links between the nature of the business problems, individual techniques and the effectiveness of the resulting implementations. This will take two forms: i) comparing and contrasting various techniques for specific problems; ii) exploration of individual techniques across a range of applications. Through the development of new intelligent system applications, understanding is sought as to what it is that makes certain projects successful and others unsuccessful. This information can then be used to help determine in advance which approaches are most likely to work well for any given business problem. This in turn is part of a greater research goal to contribute to the understanding of the successful deployment of intelligent systems.

2. Much of the financial intelligent systems application development that takes place is in-house and is correspondingly not in the public domain. One of the goals of this thesis is to bring proprietary intelligent systems research to the public domain. There are two main sections to this: i) analysis of trade/confidential/unusual or rare

data; ii) insight into the operational details of bespoke intelligent systems that are actually commissioned by companies for commercial advantage, as opposed to systems developed in research institutions for the purpose of research.

3. Another of the goals of this thesis is to contribute to the development of the field of intelligent systems through the development of new intelligent methods or the refining existing techniques. This is part of an on-going drive to build suites of effective, robust and understood techniques.

4. Intelligent systems offer a capacity to take a fresh look at old problems, while making few assumptions about the problem domain. From examining empirically justifiable models, as opposed to abstract, hypothesised (linear) analytic models, insight is sought into the nature of (for instance) financial markets. Financial economics is largely based on the linear paradigm and assumes that prices converge to their equilibrium level. The amount of evidence that suggests that this is not the case is increasing and intelligent systems, due to their inherent capability to cope with non-linear relationships, offer a prospect of contributing to a new paradigm of non-linear economics.

1.6.1 Contributions

More specifically, the principal contributions of this thesis are as follows:

1. That genetic algorithm rule induction is extremely useful for business applications as it is a powerful means of extracting non-linear value from data, the researcher has control over the representation, and the resulting induced models are transparent. This has been a constant criticism of neural networks.

2. This work on trading systems presents new information concerning the debate over the validity of the Efficient Market Hypothesis. This is a theory that divides the investment community.

3. Valuable insight into the value of exploratory data analysis, the necessity of benchmarking if possible, and matching of the application to the requirements.

4. Through the research of genetic rule induction systems, a new means of assessing the stationarity of financial markets has been developed, in a manner that removes some of the economist's defence against non-linear analyses.

5. The first known public domain design and analysis of a comprehensive genetic rule induction based trading system.
6. The first known public domain design and analysis of a continually adaptive trading engine.
7. The first known public domain data-intensive modelling of the depreciation of vehicles in the UK.

1.7 Thesis Structure

The overall thesis structure is as follows: after the introduction and an intelligent systems literature survey, one chapter is spent on each of the four projects, the implications for theories of market efficiency forms chapter seven and the final two chapters critically assess this work and draw conclusions. Each chapter begins with an introduction detailing the chapter's purpose and relevance, and concludes with a bullet-point summary. In greater depth, the content of each chapter is as follows:

Chapter Two is a survey of the field of intelligent systems. Neural networks and genetic algorithms are introduced as each of the four projects presented later use one of these techniques. Given the commercial focus of this thesis, existing applications of these technologies are reviewed.

Chapter Three presents the residual value forecasting project for Lex Vehicle Leasing. An overview of the business is given along with a statement of the problem. The data is reviewed and has some basic analysis performed on it. A synthetic data series is constructed for proof of concept purposes, and linear regression modelling benchmarks are established as test controls. Two neural network models are used for time-series forecasting and their results compared and contrasted against the benchmarks. The chapter concludes with a method for forecasting individual vehicle residual values from a general purpose time-series forecasting device.

Chapter Four details the intelligent trade filtering project for Sabre Fund Management. An overview is given of chart trading and some univariate analysis conducted on the data-set. Genetic algorithm rule induction is reviewed, and the representation presented. Experimental details are given of the genetic algorithm multivariate rule induction and the results are presented and discussed.

Chapter Five introduces the Continually Adaptive Trading Engine. The background and inspiration for this work are given, along with an overview of the ideas behind basic technical analysis. The genetic algorithm-style adaptive trading scheme is described in detail and profitability experiments are carried out on UK equities within the FTSE index. The results are presented and discussed.

Chapter Six documents the trading system induction project for the European Bank for Reconstruction and Development. It begins by introducing the government bond futures market and then presents summaries of the markets that are traded in the course of this experiment. Technical indicators are reviewed and the rule framework and evaluation scheme used by the genetic algorithm are presented. The presence of unexpected effects are identified and these are investigated further. The system is also applied to commodity futures. A new methodology is introduced for tracking the changes in the character of financial markets. The chapter concludes with a discussion and summary.

Chapter Seven is a discussion of the nature of financial markets. A pair of important works are presented and related to the results of the financial experiments documented in chapters 4,5 & 6. An overview is given of empirical studies of market efficiency, and conclusions are drawn about the validity of the Efficient Market Hypothesis.

Chapter Eight gives an assessment and conclusions of the work presented in this thesis. The thesis objectives are reviewed, and summaries and conclusions of each project are given along with an assessment of whether they have individually met their research goals. Following this is a discussion of broader thesis conclusions.

Chapter 2:

A Survey of Intelligent Systems

In this chapter, neural networks and genetic algorithms, the intelligent technologies relevant to this thesis are introduced. The operation of each technique is described together with an overview of existing business and financial applications.

2.1 Introduction

The primary motivation underlying this thesis is to test various existing intelligent technologies against a range of business problems, and then account for variations in the effectiveness of the implementations. In order to do this, it is not only necessary to have an understanding of how the techniques work, but also to examine how other researchers have tackled their problems. Consequently, this chapter will present an overview of the intelligent technologies used in this thesis, and critically assess a number of existing neural network and genetic algorithm applications that are directly relevant to this research.

As declared previously, the third thesis objective is to develop new techniques or enhance existing methods, and to do this it is clearly necessary to have some appreciation for the work that has already been carried out. Finally, the operation of the systems must be understood if their autonomous behaviour is to be correctly decoded to yield new information about the nature of the problems they are given.

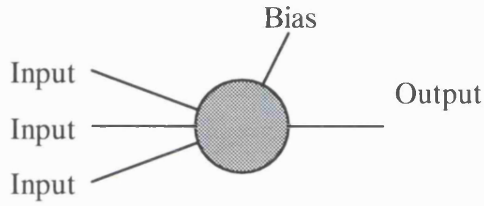
2.2 Neural Networks

There are a range of entities that come under the banner of neural networks, with established network models numbering over 100[Lipp87]. These include probabilistic networks, functional link networks, adalines (adaptive linear elements), autoassociative memories, Kohonen networks and more. The best known are Hopfield networks[Hopf82], the Self-organising map (SOM)[Koho84] and the Backpropagation model[RuHW86]. However, study will be focused upon this final group as this is the type of network that will be used in the course of this thesis. They are a general-purpose technology that can be used for a wide range of applications from prediction, classification and noise filtering. The following subsections will cover the operation, design, and application of these feed-forward neural networks.

2.2.1 Physiology of Neural Networks

The fundamental element that neural networks are built from is the *neuron*[Mast93]. It is a simple object and the operation of an individual neuron is straightforward:

Figure 2.1: A Neuron



Each neuron has a number of *inputs*. These are often positive real-valued numbers, and in addition each input has a signed, real-valued significance or *weight* associated with it. Positive weights are excitatory; negative weights are inhibitory. Each neuron also has a bias that can be thought of as an input that is held at 1.0. The neuron is then completely specified by the set weights for the inputs and the *activation function* that defines how the neuron reacts to the weighted sum of the inputs. Therefore, the net input is calculated[Mast93]:

$$net = w_n + \sum_{i=0}^{n-1} w_i x_i \qquad \text{Equation 2.1}$$

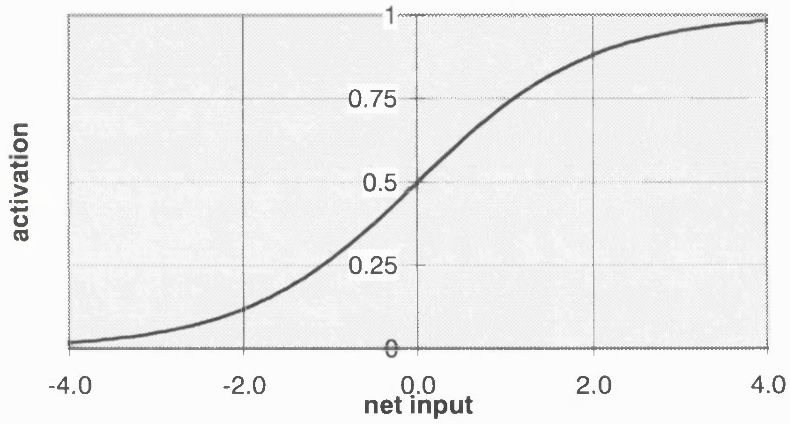
where i indexes the inputs, **x** is an n element input vector, **w** is the weight vector and w_n is the value of the bias. The neuron's *activation* is calculated from the net input using the activation function. A number of activation functions are in common use, but the majority of the models use sigmoid functions. These have the properties of being continuous, smooth, bounded and with positive derivative. The most common activation function is the *logistic* equation (Equation 2.2), and is shown graphically in

Figure 2.2. Other activation functions are in common use, such as the hyperbolic tangent [KaKw92].

$$f(x) = \frac{1}{1 + e^{-x}} \qquad \text{Equation 2.2}$$

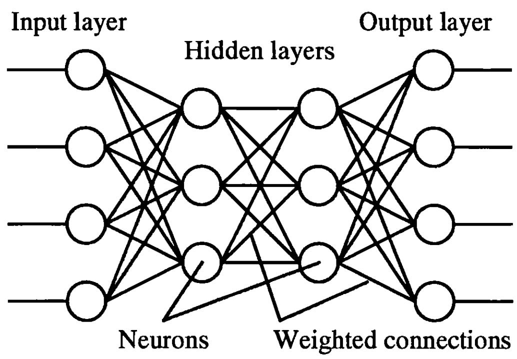
It is clear from Figure 2.2 that the activation is approximately linear with the net input in the central region of the graph, but that as the net input takes on progressively more extreme values, the progressive impact on the neuron's activation diminishes. It is for this reason that sigmoid functions are sometimes referred to as *squashing* functions[WeGe93].

Figure 2.2: The Logistic Equation



Neurons are typically arranged in layers, with the outputs from one layer feeding into the inputs of the following layer. See Figure 2.3. Modifications to this basic network topology concept are possible with the inclusion of feedback loops and links between non-adjacent layers. Notionally, the input layer performs no processing and is merely an abstraction to present the information to the first hidden layer. The output of the network is then given by the activations of each neuron in the output layer.

Figure 2.3: Structure of a feed-forward neural network



2.2.2 Training

It is the combination of network topology and the weight set that characterises a specific network. The topology is usually chosen, or refined through experiment, so for a network of any specific topology to perform useful processing requires the network's weights to have specific values. The usual method is by *training*.

Multilayer perceptrons are usually trained by example, which proceeds along the following lines: the network is given a series of training pairs <input vector, output vector>. The input data from each training example is propagated through the network to yield an output from the as yet untrained network. An *error* can then be calculated from the difference between the actual outputs and the desired outputs for each of the training examples. The network's weights are then adjusted to reduce this error. This training procedure is then repeated to progressively reduce the training errors. This can be thought of as attempting to find low points in a multidimensional error space surface, where each weight in the network contributes another dimension to the hyper-surface. Once the training data, network topology and transfer function have been selected, the function Error(**w**) has been implicitly defined. The training process can then be thought of as a search for a good point in the space.

Training must be stopped at some stage or the network will simply have memorised the training set, including the noise. This is discussed in greater depth in Section 3.6.1. Often the objective is to train the network so that it will be able to *generalise* when confronted by data that is not in the training set.

2.2.3 Conjugate gradient descent

The most common training algorithm is *backpropagation.* First presented by Rumelhart et. al.[RuHW86], this basic algorithm is covered in virtually any book[Lipp87] on neural networks and so will not be dealt with here. Instead, a more sophisticated algorithm *conjugate gradient descent* will be described, as this is the learning algorithm used later in this thesis.

Most neural network training techniques operate by attempting to find locations in the weight-space that have a low error. Unfortunately this is not as simple as it sounds as the surface is highly multidimensional and often characterised by a series of "ravines" in otherwise nearly flat error surfaces[Mast93]. Any naïve algorithm for finding minima in such spaces will find that the gradient is an extremely local pointer to a local minima, and hence will either become trapped in nearly flat areas where the gradients are small, or will overshoot wildly when the gradients are steep.

Many methods have been proposed for coping with this, such as the use of momentum [TsUh97] where the new search direction is a combination of the previous search direction and the current error-surface gradient, in an effort to prevent the network being overly influenced by local perturbations in the error surface. It is common to use a combination of momentum with a gradual reduction in step size, so that the system "homes in" on an acceptable minima. If the step size is reduced too rapidly, the network will become trapped in a local minima, whereas if it is reduced too slowly the search vector will simply thrash about over the error surface and never converge.

One method that attempts to use gradient information in a more sophisticated manner is *conjugate gradient descent*[Bren73]. Given a weight set $\mathbf{W}_0$ and a direction $\mathbf{W}_d$, the algorithm iteratively finds t such that the error E is minimised along the line through the weight space $\mathbf{W}_0+t\mathbf{W}_d$. There are a number of key parts to the conjugate gradient descent algorithm:

1. Very fast algorithms exist for finding minima in locally quadratic 1 parameter functions[PFTV88]: i.e. given a section through the space, the algorithm can rapidly find a good minima.

2. The search direction $\mathbf{W}_d$ is chosen intelligently[Pola71], a combination of recent and current search directions, with special provision for the search becoming caught in saddle points.

This algorithm is similar in many respects to backpropagation with momentum which was briefly mentioned earlier. However, it differs in two respects:

1. The algorithm has a continuously variable step size and so can both explore the space rapidly and at fine granularity.

2. Instead of being fixed, the momentum term varies in an optimal manner.

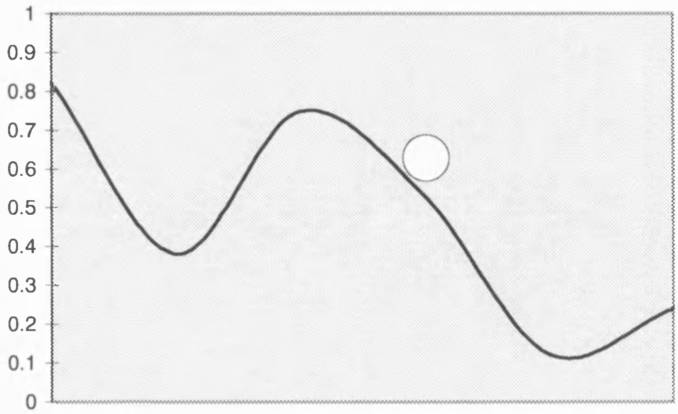
2.2.4 Simulated Annealing

Simulated annealing is a stochastic function optimisation technique that uses no surface gradient information which can also be used for training networks. The definitive text on this topic is [AavL87]. It takes its name from the analogous metallurgical process of annealing: Metals have a crystal microstructure. When this microstructure is random, the metal is weak as there are no well-defined crystal interfaces (dislocations) to prevent cracks and fractures spreading.

When metals are heated, the high temperatures make the atoms shake which disrupts the crystal structure. If a hot metal is quenched then the atoms are frozen into this random pattern which gives a hard but brittle ingot. If the metal is cooled slowly (annealed), then crystals form. The size and connections of the crystals and hence the resulting mechanical properties of the metal, are dependent on the temperature profile of the cooling process.

This model can be used to optimise functions. Consider a ball thrown into the error function shown in figure 2.4. When the system is shaken, the ball can end up anywhere. As the temperature (i.e. violence) of the shaking is reduced, a point is reached where the ball cannot be shaken out of the global minimum, but can be shaken out of the local minimum on the left hand side. As the temperature is reduced further, the ball will settle closer and closer to the global minimum.

Figure 2.4: Hypothetical error function



If this process is to be used to optimise neural networks, it is necessary to make a useful definition of "shaking at a certain temperature". For neural networks, this is a random perturbation to the weight set, where the random perturbations are scaled so that they have a specific standard deviation. The error function is evaluated after each shake. A series of shakes is performed at a specific temperature to find the best weight set at the current temperature. The standard deviation (temperature) of the random perturbations can then be reduced to progressively converge on a hopefully acceptable minimum.

2.2.5 Hybrid Learning Systems

Complementary technologies can be used together to build robust hybrid systems[GoKh95,Goon94]. In this thesis, conjugate gradient descent[Bren73] will be used in conjunction with simulated annealing[AavL87] to speed the learning process[Mast93]. This process has the following stages:

1. Simulated annealing is used to find candidate starting points.
2. Conjugate gradient descent is then used to rapidly find the best local minima.
3. Simulated annealing is then used again to attempt to escape from this local minima to a better region of the error surface.
4. Go to step 2 until the termination criteria have been met.

This is a very useful combination of technologies as the strength of conjugate gradient descent is the sophisticated processing of local gradients to find minima, while simulated annealing is powerful despite the fact that it does not use gradient information and is a stochastic search process.

2.2.6 Business and financial applications

Neural networks are a powerful, non-linear, general purpose intelligent technology, and as a result have received much interest from those seeking competitive advantage in business and finance[Goon94]. However, in addition to whatever worth they may have, any application that uses leading edge technology will gain more interest and attention than a solution that doesn't, all other things being equal.

The discussion of the operation of neural networks has so far been theoretical. Some financial applications of neural networks will now be presented, to illustrate how they can be used for tasks such as time-series forecasting, market categorisation and tactical asset allocation.

2.2.7 Time-series Forecasting

Mozer [Moze93] conducts experiments with a range of neural network models on high frequency currency data. He explores a taxonomy of neural networks based on a model of a network as a *memory* to hold on to relevant past information and an *associator* that uses short-term memory to classify or predict.

The taxonomy has the divisions along the orthogonal directions of form, content and adaptability, giving 36 distinct subclasses of neural networks, some of which have not been explored by the academic community.

The *form* of short-term memory is divided into the classes:

1. Tapped Delay-Line memory: This is the simplest form of memory, that is a buffer that contains the n most recent samples of the time-series. A small extension to this scheme is the use of non-uniform delays eg {t-1, t-2, t-4, t-8}.
2. Exponential Trace Memory: Consider a binary bit time-series x and a memory process $m_t = x_t + m_{t-1}/2$. This memory is capable of memorising the series. However, if the memory has finite resolution, or is noisy, then the less significant (oldest) information will be lost.

3. Gamma Memory: This is a combination of the previous two memory structures: a tapped delay line of exponential trace memories.

A memory does not have to simply record the input sequence: some processing can be carried out. This gives rise to various classes of the *content* of short-term memory.

1. Input memory: This is where no processing takes place before the data is "committed to memory".

2. Transformed Input memory: This is generally the standard neural network activation function - a squashed weighted sum of the inputs.

3. Transformed Input memory + State: This can be implemented in a recurrent network with 2 hidden layers. Consequently, the effects of the transformation that is applied to the inputs is in some way dependent on the current state of the network.

Instead of applying the memory functions to the inputs, they can be applied to the outputs instead. This gives rise to three more content classes: Output memory, transformed output memory and transformed output + state memories.

Finally, there is the *adaptability* of memories:

1. Adaptive Memories: These are memories that are constantly learning from the series as it progresses.

2. Static Memories: This is akin to a taught neural network whose weights have been frozen after training and is performing static processing on the series.

In summary, there are three dimensions that describe types of memory: form (delay line, exponential trace, gamma trace), content (I, TI, TIS, O, TO, TOS) and adaptability (static, adaptive). Mozer then conducts a comparative experiment on high frequency currency data with 3 different types of memories: I-delay, TIS-0 (a recurrent neural network with hidden units) and a hybrid approach combining these, for the problem of forecasting currency rates 1, 15 and 60 minutes ahead. 25 networks of each architecture were trained from random initial weights. The ten that performed best on a disjoint test set were applied to the out-sample data, where they had their forecasts averaged to produce a final forecast.

Performance is reported in terms of *normalised mean squared error*. The normalisation is done with the error involved if the forecasted value is assumed to be the same as the current value. An NMSE value less than one indicates that the system is performing some useful forecasting, and the lower the value, the greater the forecast's accuracy. Mozer[Moze93] reports an NSME of .859 on the 15 minute forecast and .964 on the 60 minute prediction, although the out-sample data was reported not to have a sufficient number of points to ascertain whether the results are simply an artefact of the data or if the system is actually doing something that adds value. Mozer found that all non-trivial models performed approximately as well as each other on the test and out-sample data:

Table 2.1: NSME for test data

Architecture	**1-minute prediction nmse** (57773 data points)
I-delay, 0 hidden	.999
I-delay, 5 hidden	.985
I-delay, 10 hidden	.985
I-delay, 20 hidden	.985
TIS-0	.986
hybrid TIS-0 and I-delay	.986

It is not disputed by the economic community that structure exists in the financial markets, but that once trading costs are taken into account, no useful structures will remain. Mozer does report that the I-delay, 10 hidden network correctly predicts the direction of market movement in 58.5% of cases, conditional upon a change having occurred. However, the information as to whether a change will occur is not available and the impact of trading costs on non-moves of this type is debatable. Also, the mean win-loss ratio is not reported - this information is required to determine whether the system is "tradable". It is simply not inferable from Mozer's paper whether this system would make profits in the market, which is the ultimate benchmark of whether the system makes useful forecasts or not.

2.2.8 Market Categorisation

Much of the research and development into the application of intelligent technologies for speculative trading is carried out in-house by banks and investment institutions that are in a position to benefit (or otherwise) from such research. However, it is interesting to attempt to assess the motives for disseminating information that appears to confer some competitive advantage.

This section will be concerned with just such an example: ABN Amro investigated the use of neural networks for characterising market behaviour[Eman96]. The starting premise is that there are a number of market price behaviours that are products of different market moods, and each of these requires the use of different market indicators to be traded successfully. The 3 stages identified were:

1. Trading: the price oscillates within a limited range;
2. Trending: the price makes progressively higher highs or lower lows;
3. Changing: the transition between trading and trending, or vice-versa.

The neural network is then used to assess which of these market characters is currently occurring, and hence select the set of predefined indicators that should be used.

Emanuels[Eman96] experiments with a range of feed-forward neural network architectures for forecasting prices of 5 stocks over various time-horizons using 1 and 2 hidden layers and between 2 and 12 hidden units. The inputs are price, quarterly price, monthly returns, and trend, volatility and trading volume indicators. After testing 1440 networks, it is concluded that the network's precise topology is unimportant, provided there are at least 5 hidden units for these inputs and data. Experiments are also conducted to find the best ratio of training set size to validation set size, and this is "much larger than expected" at 9:1.

Interestingly, a number of null results are reported: there is little correlation between the in- and out-sample root mean square error(RMSE), nor between the number of training epochs and out-sample RMSE. Also, it is reported that in-sample RMSE has little bearing on whether the network correctly forecasts the direction of the next market movement on the relevant time-frame, although this is to be expected - the

RMSE error will be small if the forecasted value is simply equal to the current price. According to financial economists, this is the best that it is possible to do.

Nevertheless, after approximately 6 months of live trading, Emanuels reports average returns of 24%, although it is not clear how or which of the 1440 tested networks were selected for trading, or if any risk adjustment has taken place. It is important to acknowledge that 6 months is not a sufficient period of time to assess a system of this nature[Debo94].

2.2.9 Tactical Asset Allocation

Tactical asset allocation is concerned with the allocation of funds between different types of assets, such as stock, bonds and cash. The objective of this is to obtain a portfolio that has the required properties: this might be maximum return for minimum risk or immunity to exchange or interest rate shifts. Neural networks are used to forecast the expected difference in returns between equities and cash and bonds and cash, so that appropriate modifications to the current portfolio can be made.

Refenes et. al.[RZCB95] document a comprehensive series of experiments into the application of neural networks for tactical asset allocation. Arbitrage pricing theory states that an assets' return is a linear function of its exposure to economic (and other) factors. Refenes et. al. have chosen variables such as expected returns, market valuation and state of the business cycle as potential inputs for the model. As many of these are highly correlated the most independent are extracted: multiple linear regression techniques are used to discover which of the inputs (factors) have the greatest explanatory power, and hence iteratively find the least useful input and remove it. From this method of "backwards stepwise regression", the initial 17 data series are pruned down to fewer, more independent inputs. 5 experiments are carried out with results that range from the "disappointing" for an all factor (unpruned) model to 4 and 1 factor models that "did much better, explaining what in financial engineering terms is a satisfactory 14.5%-17.5% of the variability (correlation squared) in [the differential returns between equities and cash]".

The 5 multiple linear regression models are then compared to a range of simple feed-forward neural networks trained with the backpropagation algorithm presented earlier.

The main parameters that were varied were the topology, and the training time. The networks are kept small as monthly data is used and hence little is available.

After conducting a series of training time experiments[RZCB95], it is concluded that the networks “exhibit a surprisingly consistent relationship between training time and training period, irrespective of the topology and [inputs]". In addition, one interesting point is made concerning the duration of training data. Networks trained on short periods with few inputs explained much of the variability - in the range 28.5 - 39.5%! This out-performed the linear models to a significant degree, and so the authors conclude that few inputs should be used so that the network can be thoroughly trained with little data. However, the authors also acknowledge that the performance of the models is clearly related to the period that is chosen for training. The group also found that although the networks failed to predict large deviations from the cash-equity differential returns series, they predicted the direction much more accurately than any of the regression models. However, the tradability of this signal is never assessed.

2.3 Genetic Algorithms

Genetic algorithms are a class of stochastic, evolutionary techniques to problem solving, developed by Holland[Holl75] to model adaptive biological processes. They work with a number of solutions to the problem simultaneously. Each candidate solution is an individual, and all the individuals in the GA constitute a population. Genetic algorithms operate by repeatedly combining the coded information from different individual solutions (genotypes) in new ways, to give rise to new sets of actual solutions (phenotypes) that are better answers to the problem. This is known as the *evolutionary cycle*, and consists of a number of phases[Gold89]:

1. Build the starting population, usually by generating a random set of individuals: this is effectively making a number of uniformed guesses about what the solutions to the problem are.

2. Evaluate the fitness of the current solutions: It must be possible to work out how good each candidate solution is, often by examining data and seeing how well it "works".

3. Preferentially select current good solutions: The better a solution is, the more likely we are to want it to survive, reproduce and influence the subsequent behaviour of the GA.

4. Combine good existing individuals to generate a new set of solutions: *Generational replacement* is the process of building a new population from the existing solutions. This is usually done with *crossover*, a process where new individuals are created by splicing information from multiple existing solutions to create new solutions, or *mutation*, where small random changes are made to current individuals.

5. Goto step 2 until some condition is met, such as a suitably low level of error, high enough fitness or a maximum number of cycles have passed.

There are a number of points to note about how the operation of a genetic algorithm differs from many other optimisation techniques[Gold89]:

1. GAs use payoff (fitness) functions, not derivatives or other additional knowledge.
2. GAs work with a number of solutions simultaneously.
3. GAs work with an encoding of the parameters, not the parameters themselves.
4. GAs use probabilistic transition rules.

In summary, the GA works by generating a (usually) random set of individuals, and this has the effect of manufacturing by chance, the fragments of a good, final solution to the problem. With repeated application of the evolutionary cycle, the GA then assembles these fragments into a good final solution.

Each of these stages will now be covered in greater depth.

2.3.1 Evaluation

For it to be possible to apply genetic algorithms, it must be possible to qualitatively assess the fitness/value/worth of any possible solution that the GA finds. This is clearly dependent on what the problem actually is. One important detail of the evaluation of individuals, is that the fitness is a property of the phenotype(the decoded solution), not the genotype (the encoding of the solution).

If the objective is to find a structure that has certain physical properties then this will require a very different fitness function to that used by a GA whose function is to find rules for forecasting volatility or optimise time-tables. Nevertheless, for any of these problems, it must be possible to make an assessment of how good any individual solution is.

Interestingly, once the representation(encoding) has been designed, the space of all possible solutions is defined, and the task of the GA can then be thought of as navigating in this space to find good solutions. This is very similar in many respects to the space searching that is done in neural network training. GAs search for high points in the fitness space whereas neural networks training algorithms search for low points in the error space.

2.3.2 Selection

There are a number of means of "preferentially selecting fit individuals"[Gold89] for parenting new candidate solutions.

1. **Roulette Wheel Selection**: an individual's probability of selection is the fraction of its fitness to the sum of the fitnesses of the total population.

$$p(i) = \frac{f_i}{\sum_{j=1}^{n} f_j} \qquad \text{Equation 2.3}$$

2. **Tournament Selection**: This selection method involves pairing off existing individuals (usually randomly) and then eliminating the less fit of the two.

3. **Ranking**: The individuals are ranked by fitness and then roulette wheel selection is performed on the rankings, rather than the individual's fitness.

Each of these schemes can be carried out either with or without replacement, although making such a change is unlikely to transform the operation of the GA. Many GA implementations use elitism. This is where the best individual(s) in the current generation automatically survive through to the subsequent generation.

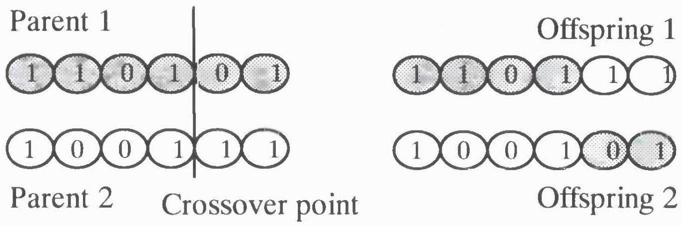
2.3.3 Operators

Once individuals have been selected for reproduction, genetic operators are used to generate new individuals from the existing genetic stock of the selected "parents". There are two main operators crossover and mutation.

2.3.4 Crossover

Crossover is the splicing or recombining of information to produce new candidate solutions[Davi91]. It involves a partial exchange of genetic information between a pair of individuals. The figure below shows *1-point crossover*. (See Figure 2.5). The position of the single crossover point is selected randomly along the length of the chromosome.

Figure 2.5: Crossover

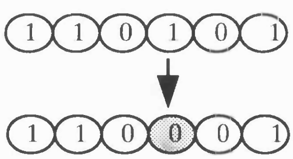


Other variants exist such as multi-point and uniform crossover[Gold89]. If the chromosome length is not fixed, then more advanced crossover strategies must be used. The *Messy GA*[Deb91] is one type of GA that uses variable chromosome lengths.

2.3.5 Mutation

Mutation is the other important genetic operator[Davi91]. It is a small random change to an individual's chromosome. It is illustrated in Figure 2.6:

Figure 2.6: Mutation



Mutation can be thought of as innovation. Whereas crossover attempts to combine existing information in new and better ways, mutation can give rise to individuals that no amount of crossing-over from the current population can achieve.

2.3.6 Variations on a Theme

GAs suffer from some common problems; research is on-going to tackle them[Bent98]. Two examples are *premature convergence*, where the population converges to a local fitness maximum, and coping with *deceptive* or *noisy functions*[Deb93]. In addition, research is conducted into the technical issues such as the efficient implementation of GAs on parallel computers.

1. Parallel GAs: multiple processors are either used to perform fitness evaluations concurrently, or Distributed GAs where practically separate population pools are evolved with few interactions between populations[AdCh94].
2. Niching/Speciation/Crowding: These promote the growth of stable sub-populations, in an attempt to focus effort on useful sections of the genotype space [Mahf95,TCRR94], or evolve multiple solutions with a single population.
3. Messy GAs: These focus on the use of "exotic" techniques, such as variable length encodings and two-stage evolution processes[Deb91].
4. Multi-objective GAs: Quite simply the evolution of solutions to problems that have multiple objectives.[SrDe95].

Two other important research drives related to genetic algorithms are:

1. Alife: This is the study, through simulation, of the basic phenomena observable in living systems. These include self-replication, evolution, adaptation, self-organisation, parasitism, competition and co-operation [Tesf96].
2. Genetic Programming: This is the construction of symbolic expression trees with genetic algorithms. These expression trees are usually LISP expressions, and hence are executable computer programs[Koza92]. The motive behind this work is to get the GA to evolve programs, instead of having human programmers write them.

2.3.7 Business and financial applications

To better illustrate the application of genetic algorithms to business problems, a number of case studies are presented from existing intelligent systems literature. It should be stressed that the application domain is in fact far greater than might be inferred from the selection of applications presented here[Koza98] - genetic algorithms are an extremely general purpose optimisation technique, and financial applications have been chosen simply for reasons of relevance.

2.3.8 Trading Rule Induction

Oussaidène et. al.[OCPT97] used a genetic programming approach to inducing rules for trading exchange rates. The main thrust of their paper is concerned with the efficient parallelisation of genetic programming on a distributed system, and then this tool is applied to trading rule inference.

The trading rules that the system finds are all of the form:

$$\text{IF } |\text{EXPR}| > \text{K THEN S=SIGN (EXPR) ELSE S=0}$$

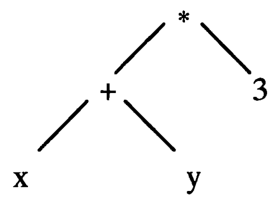
where if the modulus expression EXPR, evaluates to greater than a activation threshold K, then the system generates a trading signal. If EXPR is positive, a buy signal is generated, if EXPR is negative, a sell signal is generated. The value K is chosen to reflect the aggression of the system: the lower its value, the more frequently the system will trade. The expressions are constructed from a grammar that consists of a function set with basic logical and arithmetic operators {AND, OR, NOT, IF, *, /, +, -<, >, Min, Max, Abs} and a set of terminals that contain some preoptimised momentum based indicators, a 16 day volatility indicator, a random number generator and the values +1 and -1. Consequently an example of a currency trading rule might be:

$$\text{IF } M_1{*}V > \text{Min}(+1,M_2) > \text{K THEN} \ldots$$

where M_i is preoptimised momentum indicator i, and V is the volatility indicator.

Programs evolved by genetic programming (GP) techniques are usually represented by trees (see figure 2.7) and since this approach uses GP for rule induction, the expressions formed can potentially be arbitrarily complex. The justification of this is trivial, as any sub-clause of the expression (sub-tree) can be replaced with a more complex expression (a larger sub-tree). The crossover operator must be modified for genetic programming: instead of splicing bit strings as before, it exchanges sub-trees of the symbolic expressions being “crossed-over”, to produce two new S-expressions. The value of a genetic programming approach to rule induction is that the GA then has great flexibility in the complexity of the solutions that are found.

Figure 2.7: Tree for Symbolic Expression (x+y)*3



It was found that the best of the trading rules found by the system averaged an annual return of approximately 5% on out-sample data, but the authors acknowledge that further work is required to enhance the robustness of the solutions. There are two problems with this result:

1. It is unclear what a pre-optimised indicator is - and it is never revealed which data this preoptimisation takes place on. If the out-sample data was used in their construction, even in a separate experiment, then the integrity of any test must be questioned.
2. Also, the results are grouped according to the out-sample performance. The information required to select the best trading rules does not exist until after the experiment had finished, and so a trader would not be in a position to only trade the best performing rules.

It is interesting to acknowledge that one of the authors works for a firm whose core business is vending market information and hi-tech market analysis and trading tools.

Oussaidène et. al. [OCPT97] also found that simpler rules operated in a more reliable manner than more complex rules - this is an interesting result as it is consistent with Refenes'[RZCB95] results presented in section 2.1.10, and with results of experiments presented in this thesis.

2.3.9 Portfolio selection

Loraschi et al. [LTTV96] have applied genetic algorithms to portfolio selection: the problem is to select the weighting of assets in a portfolio such that the risk exposure of the portfolio is minimised for any given level of expected return. As the risk associated with a portfolio increases, the required return increases, carving out a non-decreasing curve in risk/return space called the Efficient Frontier. Much of portfolio theory uses the notion of variance of returns as a measure of risk, and while this has given some very elegant theoretical results, it has not always worked so well in practice. Instead, to attempt to capture the idea of risk as the chance of incurring a loss, this study examines the distribution of losses.

This problem then can be expressed as the two-objective optimisation problem:

$$\min\{\mathrm{Risk}(\mathbf{p})\}$$

$$\max\{\mathrm{Return}(\mathbf{p})\}$$

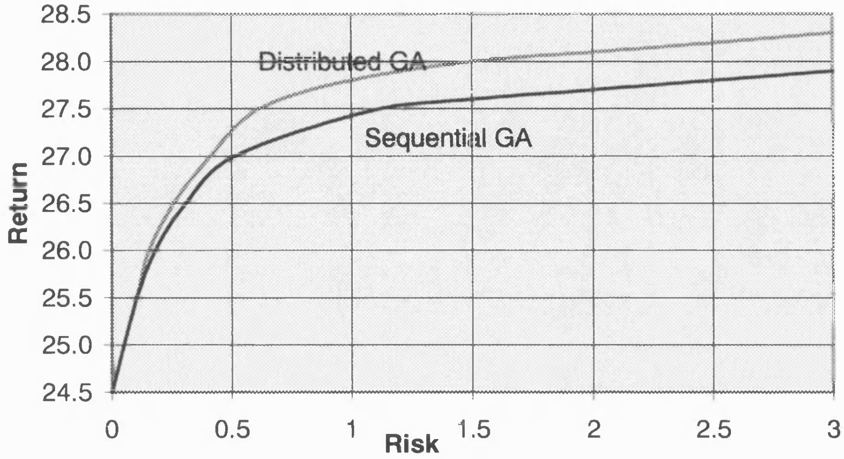
where p is a portfolio of assets.

Genetic algorithms are often effective in these selection problems, which is one reason why this experiment is interesting, given that there are 10^{39} ways of selecting (say) 20 different assets from 100 candidates. The authors use a number of coarse-grained *islands*, each of which has its own unique sub-population on a separate virtual machine. Small random exchanges of individuals occur between islands once during each evolutionary cycle. Apart from the exchange of individuals, each island operates as a separate GA in its own right. It is due to this segregation of sub-populations, each exploring a different region of the searchspace that the GA can be slowed in its race for population convergence and in the process discover better solutions. This distributed GA is compared to a single, unsegregated population of equal size and the results contrasted[LTTV96]. The results are shown in Figure 2.8. The representation that is used is not discussed, although it is probably simply a list of variable length, of assets.

The distributed GA finds a consistently higher return portfolios for equivalent risk than the sequential reference GA, although part of this result is due to the extra computational power that can be applied to coarse grained genetic algorithms, and indeed this is one conclusion that the authors draw.

Figure 2.8: Efficient Frontier for 53 Asset Portfolio:

Distributed and Sequential GAs



However, it is not possible to assess from this information what risk/return would be offered by a passive tracking portfolio. This is the benchmark that the GA has to beat.

2.3.10 Credit Evaluation

Credit scoring is a difficult problem to automate as it is characterised by large, inconsistent, incomplete, expensive and changing data sets. In addition, it is clear that human intuition is often challenged or even surpassed by modelling techniques. Walker et. al. [WaHG95] describe an intelligent application OMEGA for evaluating applicants. OMEGA is a powerful tool that can carry out database processing, carry out model induction from data, do the training and validation and even produce reports for financial management.

Credit scoring is the process of attempting to determine from information such as salary, marital status and age whether the applicant is a "good" or "bad" risk. (A good risk is one where there is little chance of default on repayment). Even very simple rules tended on average to produce more reliable results than credit managers - they "are less subjective, not easily distracted by charm or beauty … and will not become ill or agitated when important decisions have to be made"[WaHG95].

Walker et. al. use a genetic programming approach to rule induction, akin to that described in section 2.2.8. In addition, this system also has the capability to use custom operators. The model induction takes place over a database that has 30000 entries and up to 60 variables per entry. The credit manager using the system can select which variables from the system are likely to be the most relevant, and which variable is to be predicted. The user then specifies what constitutes an acceptable model in terms of robustness and reliability. The system then offers various linear analysis capabilities, or the data can be fed direct to the modelling engine, where the GA attempts to evolve sets of rules that have the explanatory requirements the credit manager requires. This is done using cross-validation techniques, where the error is monitored on data sets disjoint from the training set, in an attempt to prevent ill-conditioned solutions. The system also automatically switches to alternative evolutionary strategies such as simulated annealing when genetic diversity deteriorates.

The results of the system's use at a bank is reported, but the GA system only outperforms the loan application rating benchmark by 2.31% (65.15%, up from 63.68%). However this seemingly meagre increase actually represents a significant saving in costs and consequent profit gain.

The author also carries out a comparative study of credit scoring techniques on another (standard) data set that has been used before for comparative analysis. This consists of 1000 loan applications (700 good, 300 bad), each with 20 predictors such as age, current account status and location. Noting that the default accuracy for this task is 70% (700 good cases of 1000), the results of the various applications are presented in Table 2.2[WaHG95].

Table 2.2: Comparative performance of credit scoring applications

Algorithm	**Method**	**Accuracy**
OMEGA	Genetic algorithm	77.4%
CN2	Rule Induction	72.0%
NEWID	Rule Induction	65.1%
C4.5	Rule Induction	72.7%
AD	Neural Network	72.7%
MAD	Neural Network	70.9%
COUNTER PROPAGATION	Neural Network	68.7%

2.4 Summary

In this chapter, an overview of the operation of neural networks was presented. This consisted of an overview of the operation and training of neural networks. The papers concerned with applications of neural networks in the financial domain examined in depth were:

- a taxonomy of networks for time-series forecasting (specifically currencies);
- market classification for trading equities;
- an investigation into the use of neural networks for forecasting differential returns between equities and cash for tacit asset allocation purposes.

It is important to acknowledge that there is an enormous amount of neural network research, literature and case studies, and that it is impossible to give more than a flavour of the work that has been done in any sensible amount of space.

Genetic algorithms are not quite in the same situation - they do not have a rigorous theoretical background as yet, and so much of the research that takes place is somewhat applied. The section on genetic algorithms began with an overview of how they operate, and then looks at some financial or business applications to observe how these technologies are being developed and deployed. The applications were:

- Trading rule induction with genetic programming;
- Portfolio optimisation;
- Credit evaluation rule induction from databases.

Chapter 3:

Neural Networks for Residual Value Forecasting

This chapter investigates data-intensive techniques for forecasting the residual value of used vehicles after a 3 or 4 year hire period. A database is available of monthly samples of used vehicle prices for all current models in a range of ages and conditions. Linear regression models and neural network time-series models are developed to address this problem and are compared.

3.1.1 How the Fleet Hire Business Operates

The fleet hire industry exists because there is a market for medium term (3-4 years) rented vehicles and it is this industry that accounts for more than half of the annual new UK vehicle sales[CAP97]. In addition, individuals are now starting to demand similar services from the fleet hire companies, but the majority of hire contracts are signed with companies who wish to offer car ownership as part of their employees' remuneration.

To supply a vehicle to a customer, the fleet hire firm must first buy it from the vehicle manufacturer. The hire firm then loans the vehicle to the customer for an agreed period, in return for a fixed monthly rental. At the end of the contract period, the vehicle is either purchased by the customer, or returned to the fleet hire firm where it is then sold, often direct to garages or through auction. Most vehicles will depreciate over the hire period, and so the hire firm has to recoup this loss and any other cost by charging rental. The bulk of the hire charges that the customers pay goes directly into financing the depreciation of the fleet.

The reason that this industry can exist is largely due to the economies of scale associated with the fleet hire firm's operations, and hence be in a position to offer customers better products at lower prices than the customers could achieve themselves. A large contract hire company should have a significant bargaining tool when negotiating the vehicle prices with manufacturers due to their volume of business. If a manufacturer refuses to offer highly competitive prices to the hire firm, then the hire firm will simply be in a position to offer better value with a rival manufacturer's vehicles.

3.1.2 Pricing Mechanisms

From discussions with analysts at Lex Vehicle Leasing, it appears that there are two main ways in which the level of hire charges can be set for a vehicle. The first is a forecast led approach, where the hire charges are the sum of the expected depreciation over the hire period and other costs such as maintenance agreements, commissions and profit margin. To estimate the depreciation over a 3 or 4 year period, it is clearly necessary to forecast the mean *residual value* of the vehicle at the end of the hire period. The publication CAP Monitor[CAP97] is probably the most respected and widely used source of such forecasts, and almost forms an industry yardstick for the expected future value of vehicles. The problem with in-house forecasting is that as CAP Monitor is widely used, there is a certain similarity between each company's forecasts. If an individual company thinks that a vehicle will depreciate less than its competitors forecast, then this means its hire charges will be lower and the firm will consequently write lots of business on that vehicle as customers shop around. However, when the vehicle is sold at the end of the hire period, there is a chance that the rental has been set too low and that the vehicle's depreciation over the hire period will not have been recouped. Conversely, if the in-house residual value forecast is too low, then the hire charges will be too great and customers will simply go elsewhere.

The second approach to setting hire charges is a market-led approach, where the company attempts to out-value the competition in some manner. This can either be through reducing prices, offering superior levels of service and support, or indeed any means of making the company and its products more attractive to the customer. This approach is most easily used by very large hire firms, as they should be able to negotiate very good rates from manufacturers, streamline their business and exploit economies of scale. With a market led approach, it is important to assess what the market rates actually are. Sometimes this is done through 'normal' channels, such as agreements between companies, while other fleet hire firms go to great lengths to find out empirically (i.e. by posing as customers) what rates their competitors are charging, and how easy it is to persuade the competitors' sales-force into giving discounts to clinch the deal. There is a serious drawback with a market led pricing scheme: if a vehicle is popular, then firms will tend to reduce their rates to take advantage of the demand for that vehicle. The rental on the popular vehicle will be low, so by implication, the vehicle's residual value at the end of the hire period must be high for the hire company to recoup the costs of the vehicle's depreciation. However, if the hire charges for a popular vehicle are driven too low across the industry, then the hire firms will write lots of hire contracts for that vehicle, as it will then also be perceived by the customers as better value than vehicles from rival manufacturers. This poses no immediate problems, but when the contracts expire 3 years later, the market is suddenly flooded with 3 year old once popular ex-fleet hire vehicles, and as supply out-strips demand, the bottom falls out of the resale market. This increases further the losses incurred by the contract hire companies.

Many companies operate a blend of these two pricing strategies, with each component acting as a reality check on the other. If the hire charges derived from these two approaches differ by a wide margin, then this is an indication that the hire charges for the vehicle in question should be examined. Errors do arise periodically, and potentially they can be very expensive, either in terms of lost business or a failure to recoup asset depreciation. At the very least, an intelligent forecasting system could act as a valuable safety mechanism.

3.2 Nature of the Problem

It is straightforward to express the exact nature of this problem: to use intelligent systems to assist with the forecasting of vehicle residual values. The industrial partner who commissioned this work, Lex Vehicle Leasing, was keen to use neural networks to sift through huge amounts of data and discover knowledge that would assist with running their business. At the time, residual values were set by what the Lex's management deemed to be "a rather unsatisfactory blend of discussion, intuition and a possibly excessive reliance on CAP Monitor". One of the principal aims of this project was to introduce some consistency into the residual value forecasting process, and try to make it more systematic and scientific. In addition, at the time the work was commissioned, the UK was beginning to climb out of a recession and Lex wanted assistance with assessing the likely future behaviour of the market. Neural network modelling has been attempted, along with a comparison of this approach with some other modelling and forecasting techniques. Prior to any feasibility study being carried out, Lex had an agenda that consisted of a number of goals:

1. To try to forecast the movement of the vehicle market as a whole.
2. To try to forecast how the vehicles of a specific manufacturer would depreciate relative to the whole.
3. To try to forecast residual values for individual vehicles.

Once the software was completed, the intention was for the company to distribute it freely to competitors. There were two main motives behind this:

1. To try to introduce some stability into an otherwise fairly volatile marketplace
2. To try to influence Lex's competitors to take a view of residuals that were closer to Lex's than they otherwise would be. This would then enable Lex's economies of scale to result in highly competitive pricing that smaller companies would find difficult to match.

3.3 Data

Fleet contract hire is primarily concerned with cars, although it is possible to hire trucks, vans, forklifts and other industrial durables from many contract hire firms. This study has been conducted exclusively with car residual value data, as this is by far the largest section of the market, the data is likely to be more stable than in other sectors as the sample sizes are much larger and this is where the bulk of Lex's contract hire business is. This data was disseminated throughout the company in book form, but in the previous three-and-a-half years, it had been available on disk as well. This is the data-set that this investigation was conducted with.

Lex had been using CAP Blackbook[CAP97] in its pricing operations, which give the prices for most makes and models of car, in a range of conditions, for every plate for which the vehicle was available. An extract has been reproduced below to illustrate the data's organisation. The descriptions that follow below are paraphrasings of CAP's own.

Figure 3.1: CAP Blackbook Format

	Ins Grp	Date Intro	Cost New	***CAP Retail***	**Year & Letter**	Miles 000's	***CAP Clean***	Aver -age	Below Av.
Cavalier 2.0i									
2.0iCDX 16v Hatchback 5dr	13	Apr94	16230	***10450***	**1994L**	10	***8925***	8150	7575
	8917			***9550***	**M+625**	30	***8025***	7550	6825
				8775		50	***7250***	6650	5925
			17000	***11850***	**1995 M**	5	***10300***	9900	-
		Current	17175	***11095***		15	***9575***	9100	-

Vehicle Description: Shows the model designation, engine size, number of doors and any additional information to help precisely specify the vehicle such as trim variant.

Insurance Group: An estimate of the vehicle's insurance group. Below is a CAP identification number, unique to this vehicle.

Date of Introduction and Cost New: Date of introduction or change in the vehicle specification. Approximate cost new in January.

CAP Retail: Reasonable trade retail figure allowing for a sensible profit margin derived from CAP research.

Year & Letter: The first year of registration in the UK and the correction for vehicles first registered in the subsequent year: they are slightly newer and hence worth more.

Mileage Variation: A flexible format that allows CAP to only devote space to the most common vehicle mileages. As vehicles age, The mean vehicle mileage will increase, and hence the most useful mileage intervals to give residual values for will change.

CAP Trade Prices: CAP Clean is a trade price for a vehicle ready for immediate sale, original throughout, with unmarked bodywork, clean, undamaged interiors and is mechanically faultless. Vehicles in excess of 2 years old with slight paint defects can be considered clean if the rest of the vehicle is up to standard. Average vehicles may require a little mechanical or cosmetic work, but could be brought up to CAP clean condition for reasonable cost. Below average vehicles are unsaleable without mechanical or cosmetic work. They often look scruffy.

The resulting data set is made of thousands of vehicles (there are in excess of 10000 unique vehicle id numbers), each in a range of variants, and each variant in a range of ages and conditions, but there are at most only 44 monthly samples of how the value of this vehicle has changed over the last three-and-a-half years.

3.4 Data Preprocessing

This data-set is in excess of 50MB in size, which at the time was too large for it to be easily manipulated on the available hardware. In order to make it more manageable while the system was being developed, a cut down version of the data-set was constructed that would:

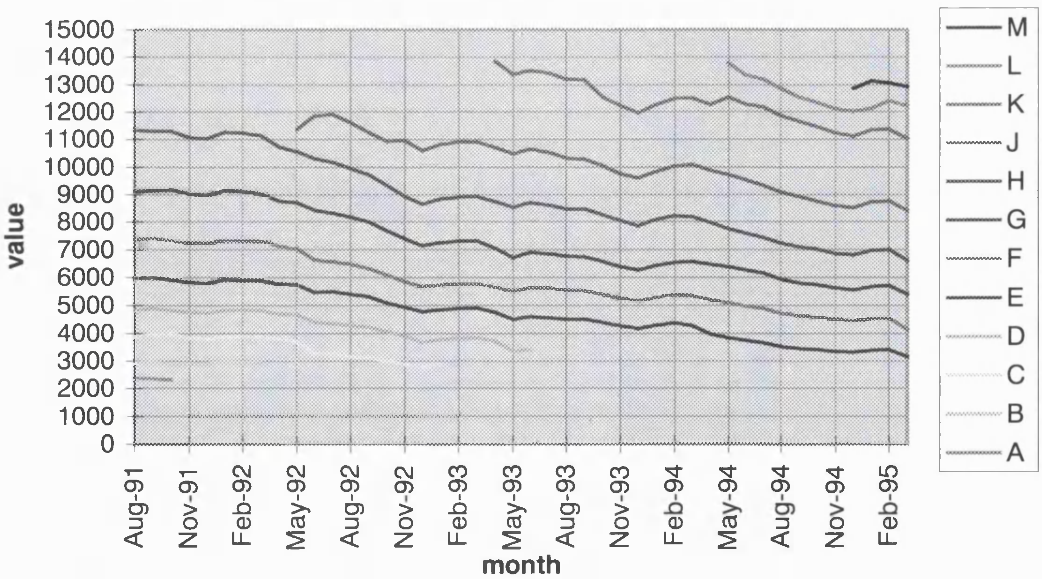
1. Display similar characteristics to the path of depreciation of a single vehicle.
2. Demonstrate the feasibility or otherwise of the chosen approaches.
3. Be of a sufficiently small size such that it could easily be manipulated on the available computers.
4. Rapidly demonstrate to management the value of data-intensive analysis, in an industry where such approaches are uncommon.

3.4.1 Constructing the Synthetic Series

To construct this data-set, it was decided simply to take the mean value of all the vehicles mentioned in CAP Blackbook each month. This would clearly be perturbed to a small degree by the issuance of new and large ranges of budget or luxury vehicles, but the effects of this would be dwarfed by the rest of the data. In order to retain some information as to how vehicles depreciate, as opposed to how the aggregate value of the vehicle market progresses, a number of distinct series were constructed from the raw data, each documenting the time-evolution of the market value of vehicles with a particular registration letter. This retains the smallest information set that is needed to trace the ageing of a specific, though admittedly large, group of vehicles. These separate series can easily be constructed by parsing each vehicle datum for its registration information, and then adding it to the appropriate series.

Each plate registration is available new in two calendar years - from August to December, and from January to August in the following year. The notion of second letter was introduced to help cope with the fact that towards August, a vehicle bearing the current plate can be nearly a year old if it was purchased soon after the introduction of the new plate, whereas a vehicle purchased in June will be one month old. For this reason, the calendar year is divided into two sections: August to December, and January to August. This complicates matters slightly, as the data for the second letter entry in CAP Blackbook disks is not complete and consistent - it is not available for many vehicles. In addition, the sample size for the 2nd letter entries is much smaller than for the complete data set. For this reason, and because this market proxy is simply for developing the prototype forecasting system, the second letter entry has simply been ignored.

Figure 3.2: Residual Value Depreciation (Synthetic Series)



As described above, CAP Blackbook gives the current trade resale prices of every vehicle in a range of mileages and conditions. To unify and simplify the residual value forecasting process, the company used the lowest mileage entry under the heading “CAP Clean”. CAP Clean is an assessment that the car is in a state of near flawless presentation, is mechanically sound and the vehicle is ready for immediate resale. Once the second letter entries have been removed, the decay of value of CAP clean vehicles is shown in Figure 3.2:

At this stage it was judged that it was not a terribly sensible proposition to attempt to forecast individual vehicle's residual values in a manner similar to the way in which one might attempt to forecast the depreciation of a specific manufacturer or vehicles of a specific registration. There are three main reason for this:

1. Time/space constraints on inducing 10000 bespoke models for residual value forecasting.
2. 10000 separate forecasts are not usable unless integrated seamlessly into other automated systems.
3. The data will be less stable as fewer samples will exist.

A method for attempting the forecasting of individual vehicles is proposed in the concluding sections of this chapter.

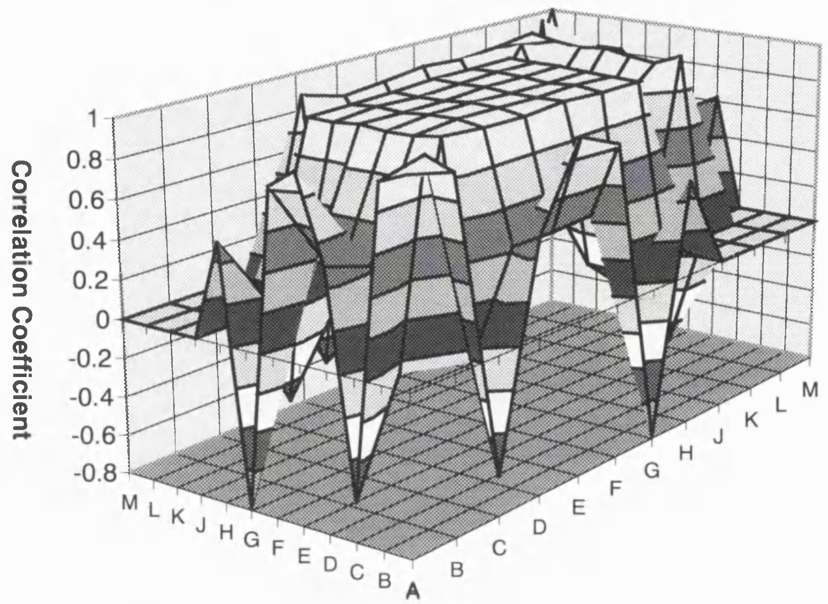
3.4.2 Correlation Coefficients

It is clear from figure 3.1 that all the different plate registrations progress in step with each other. This is particularly clearly exposed if a chart is plotted of the correlation coefficients between the series for the different plate registrations. The correlation coefficient between a pair of series x,y is defined as[Hays81]:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{n\sigma_x\sigma_y} \qquad \text{Equation 3.1}$$

where $-1 \leq \rho \leq 1$ and μ_i and σ_i are the mean and standard deviation of series i respectively. ρ near 1 indicates that the series move in step, $\rho = 0$ indicates that there is no correspondence between the series at all, and ρ near -1 indicates that the two series are perfectly out of step - when one moves up, the other moves down. Some of these coefficients will be undefined as their time-series do not overlap - the series for plates A,B and C do not exist at the same time as the series for plates K, L and M so there is no meaningful value to calculate.

Figure 3.3: Correlation Coefficients



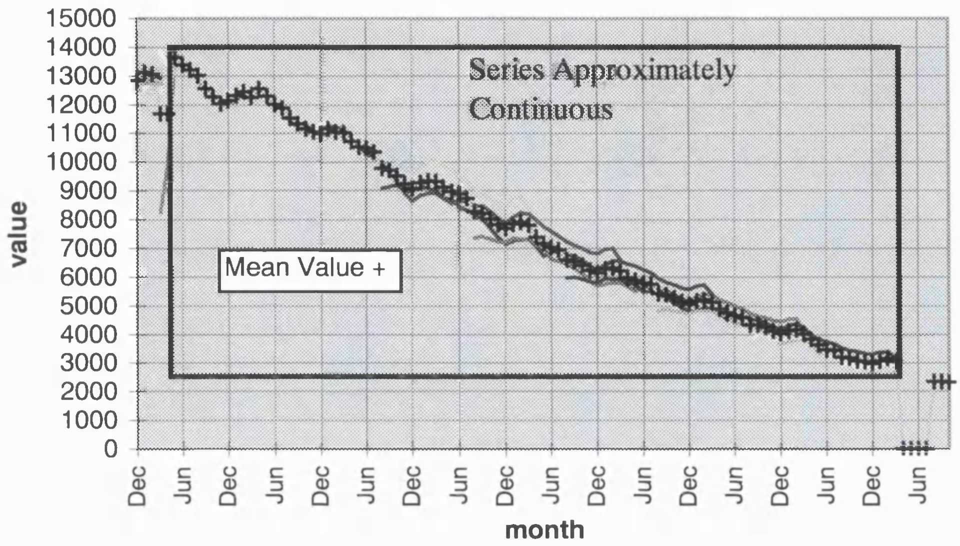
It is clear that there is a significant plateau in the graph - this is where the correlation coefficients are all near 1 for plates E,F,G,H and J. Around the periphery of the plateau are some undefined and negative values. Little value should be attributed to these downward spikes as they are simply an artefact of very small data points. They are the product of taking correlation coefficients of other series with series A, which has only 3 defined values. Negative correlation coefficients also appear for series L which is in the process of stabilising. This figure then demonstrates that all vehicle registrations respond to external economic factors in very similar ways, as very little divergence exists between the depreciation of vehicles with different plates.

3.4.3 Combining the Synthetic Series

One problem that has already been noted is that there are relatively few samples of the behaviour of the depreciation of individual vehicles. This is a problem if it is planned to use either time-series forecasting techniques or intelligent systems that induce their own models: there will probably be too little data for a reliable model to be found.

To maximise the value of this data set so that time-series forecasting techniques can be used, the positioning along the x axis of the individual (un-normalised) decay curves can be changed. The only constraint is that at any position along the x axis, all the vehicles must be the same age. In this way, a new synthetic series is built that documents the progression of vehicles from new to old. Where multiple series exist at the same x coordinate, the average value is taken of those that are defined.

Figure 3.4: The Serialised Synthetic Series

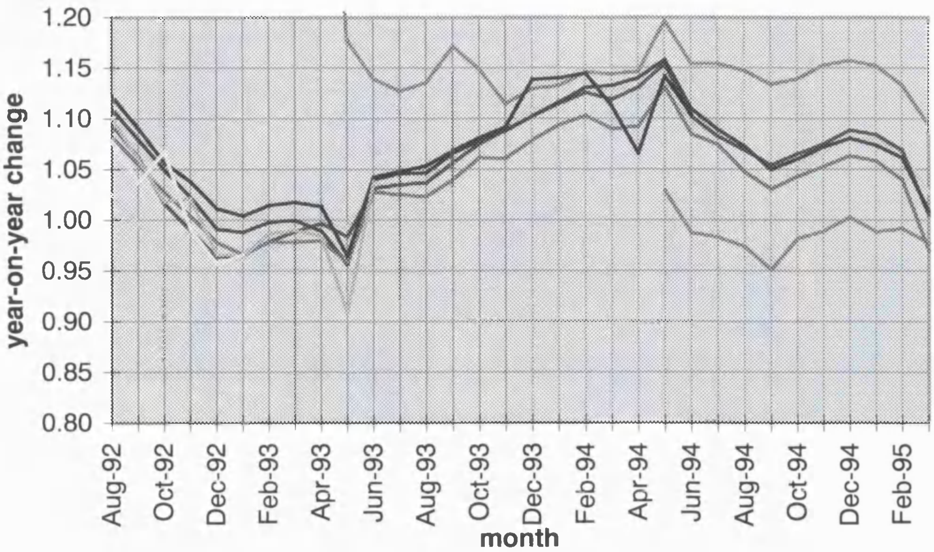


At the beginning and end of these series, the data is erratic. When a new plate registration is introduced, it takes several months for second-hand examples of those vehicles to arrive on the market in sufficient numbers for the mean value of the plate to be well defined. Similarly, when a vehicle is very old, it does not add value to CAP's customers to present this data, and so if this information is collected, is not published. Consequently, again the sample size diminishes and the value ceases to have real meaning. For this reason, the beginning and ends of this synthetic series have been cropped.

3.4.4 Year-on-year changes

The year-on-year change of a quantity is defined as the fractional (or percentage) change in a quantity from the same period one year ago. One of the main benefits of examining this quantity is that it immediately disposes of precisely seasonal variation.

Figure 3.5: Year-on-year changes in Residual Values

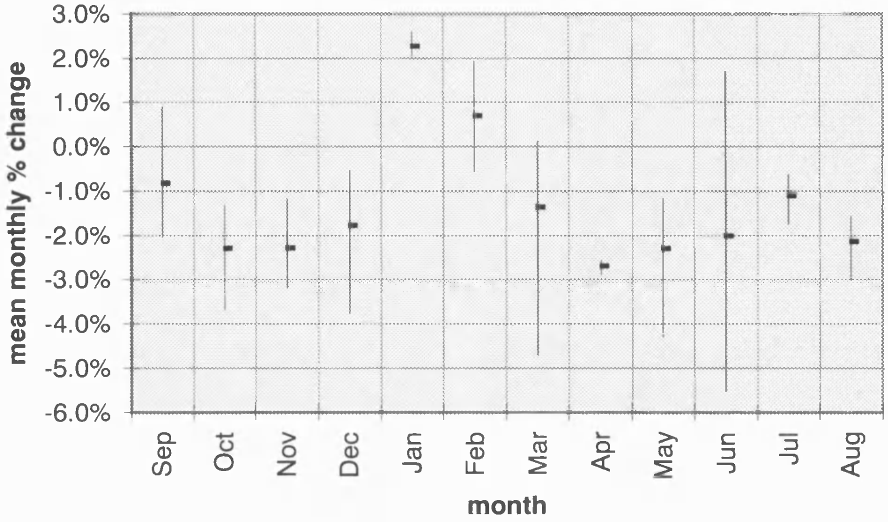


If the year on year changes of the synthetic series are examined as shown in Figure 3.5, it can clearly be seen that these ratios are undergoing an evolution of their own, and that this progression appears to be loosely cyclical in nature. If the year on year changes really do follow a cycle, then only about one-and-a-half cycles are represented, and this is too little to confidently assert that such a cycle exists at all. In addition, the shape of the cycle is not stable and the series exhibits a significant level of noise. As a result, the observation can be made that this resulting series is not one that could easily or reliably be forecast with the data that is currently available. For this reason, although forecasting year-on-year changes would be valuable if it were reliable, it will not be attempted, as it is both unlikely to yield satisfactory results, and it will be impossible to test the resulting model thoroughly with the available data.

3.4.5 Seasonality

The mean seasonal variation in the synthetic series can be examined to assess the intra-year dynamics. If they appear reliable, then they can be removed before the forecasting is done, and then added back in to the results of the forecast. The value of doing this is that none of the forecasting techniques' power then needs to go into modelling such variation. Indeed, from a casual examination of the synthetic series, it appears that a combination of linear and seasonal modelling may forecast the series as well as a more complex approach.

Figure 3.6: Seasonality of Residual Values: Mean monthly % changes



It is straightforward to calculate the mean monthly percentage change in residuals from the data. This is shown in Figure 3.6. The vertical bars are the ranges of the monthly changes, and the 'tick' is the mean monthly change. CAP Blackbook[CAP97] states that:

> "from a low point in December the market climbs towards the Spring, levels out during the Summer holiday months, falls during August, strengthens slightly in September and then declines from October to December."

It can be seen from Figure 3.6 that there is some value to this view of seasonality, but that there are significant deviations from CAP Blackbook's view. The error bar for June is approximately the height of the entire chart at a 7% range, and this severely limits both the value and the validity of removing seasonality from the series.

It is important to realise that the seasonal behaviour of residual values takes place in a larger economic picture that is not seasonal in nature. At the time the project was commissioned, the UK was starting to come out of a recession, and the company wanted assistance with forecasting in the uncertain times ahead. So it is likely that a strong non-seasonal component will exist in the price series, and this will further reduce the seasonal signal-to-noise ratio.

3.5 Modelling Benchmarks

Before neural networks are used for modelling the time-series, a benchmark is needed to compare the results against. Without this, no principled method exists of deciding whether or not the neural network approach is working well. The two benchmarks that will be used to assess the effectiveness of neural networks on this problem are linear regression and an exponential decay curve.

3.5.1 Linear Regression

Linear regression is the process of fitting a straight line to a number of points, such that the line passes as close as possible to every point. More formally, it is the forecasting of dependent variable y, from a linear combination of independent variables $\mathbf{x}$[Hays81]:

$$y(\mathbf{x}) = f_0 + f_1 x_1 + \ldots + f_n x_n \qquad \text{Equation 3.2}$$

The factors f_i can be found such that some error metric is minimised. For linear regression, this error metric is often chosen to be the root mean square error, i.e. it is the square root of the average of the squares of the errors that is minimised. For data sets that have specific properties or for applications that are sensitive to particular types of errors, alternative error measures can be chosen. Using the *absolute error*, for example, will place less emphasis on *outliers*, points far from the regression line[Mast93]. This may be more appropriate than the mean square error if the data set has a significant number of samples that are anomalous to the bulk of the data.

The optimal values for the factors f_i can be found analytically[Hays81]. n samples of m independent variables can be arranged in a n*(m+1) matrix, A, where each row is a sample and each column contains the values for each independent variable. One column is set entirely to 1 (which is why an (m+1) column matrix is needed). The dependent variable is expressed as an n row column vector, Y. These two matrices then are connected through a column vector of to-be-estimated coefficients, B:

$$AB = Y \qquad \text{Equation 3.3}$$

If there are more samples than independent variables, then in the likely event that Y does not lie in the sub-space spanned by A, B can be constructed:

$$B = (A'A)^{-1} A' Y \qquad \text{Equation 3.4}$$

The derivation of this is not given here, but will be in almost any intermediate level statistics text. This equation will minimise the mean square error in the predicted values of Y. This solution for B need not be unique, and this leads on to a potentially serious problem with linear regression - that of points being *near-singularity*.

3.5.2 Linear Regression Bogeyman

Consider a regression experiment $Y = aX_1 + bX_2 + c$ with the following values:

Table 3.1: Regression thought experiment

Sample	X_1	X_2	Y
1	2	1	3
2	4	2	6
3	6	3	9
4	8	4	12

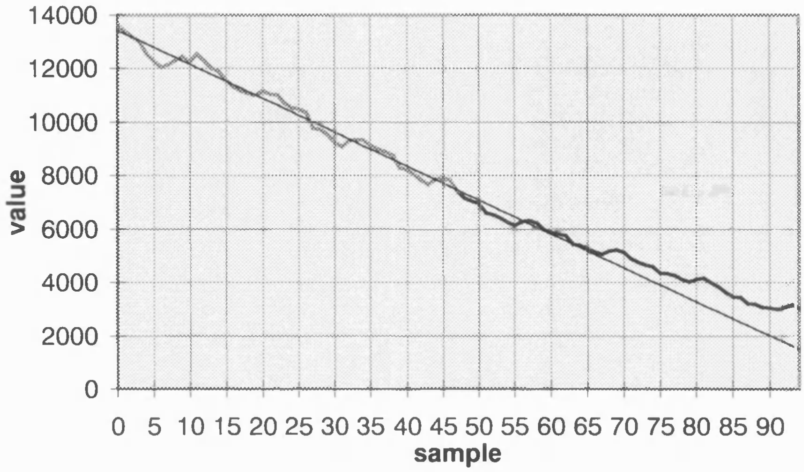
It is obvious that (a=1; b=1; c=0) is a solution, but so is (a=2; b=-1; c= 0), and even (a= 1,000,000,000; b=-1,999,999,997; c=0). The problem lies in the fact that there is an exact linear relationship between X_1 and X_2, and so the system is said to be *singular*. In this situation, multiple regression should not be attempted as it can clearly lead to very ill-conditioned solutions. The real problem comes when the data are *marginally singular* - the data will pass any test of absolute singularity, but the program will attempt to minimise the mean square error as far as is possible. This will often result in huge weights as in the third example above, good (low) training errors and appalling out-sample performance. However, in a well designed experiment, the data will not be near singular, and considerations of singularity are irrelevant. However, it is good to be aware of the potential pitfalls associated with any particular technique!

3.5.3 Use of Data

In order to be able to compare the intelligent systems in a useful way with the chosen linear methods, the same approach to data usage will be taken as with the intelligent systems later. Informally, this consists of forming a model from one sub-set of the available data and then testing the resulting model on a disjoint data-set.

In this case, 'forming the models' is the finding of the parameters for the linear regression line. The results of this are shown in Figure 3.7.

Figure 3.7: The Linear Regression Benchmark



Model statistics are presented in Table 3.2.

Table 3.2: Linear Regression Benchmarks

	Mean square error	**Mean forecast value error**
Training set	44229	£210
Test set	228200	£478

Table 3.2 could indicate that a level of over fitting in the model exists as the training score is significantly better than the testing score. The first section of data corresponds strongly to the data points, and it is only after a considerable period of synthetic time (72 samples is equivalent to 6 years) that the data starts to deviate from the model. It is unsurprising that the model begins to collapse after some time, as this would lead to absurd forecasts of extremely aged vehicles having negative residual values.

3.5.4 Exponential Decay Curves

The linear regression in the section above forms the first benchmark. This may not be the best benchmark that can be used, as one might not expect vehicle prices to drop linearly. Given that a vehicle depreciates rapidly as soon as it leaves the forecourt, and that at the end of its life its depreciation rate is low, it may be more appropriate to model the depreciation of vehicles with a decay curve.

The approach used to find an exponential decay regression line is very similar to that used for linear regression. The regression equation is now:

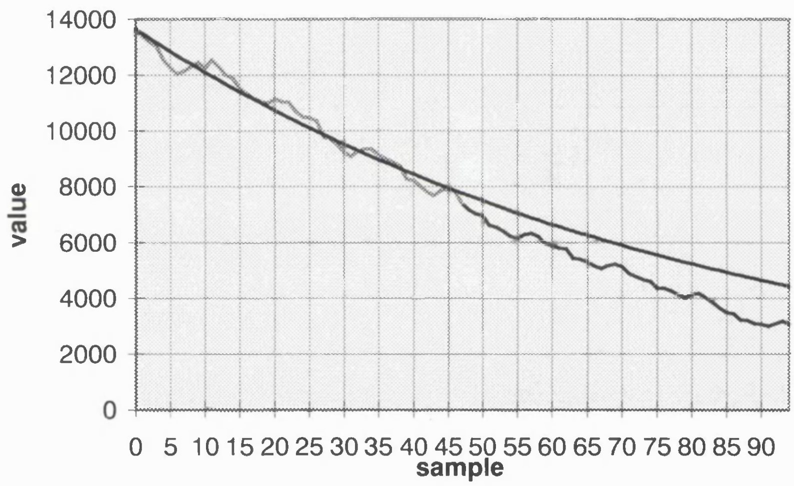
$$y(x) = Ae^{Bx} \qquad \text{Equation 3.5}$$

If logs are taken of the data values, then this transforms the problem into a linear regression. The values for the parameters are then:

A = 13650; B= -0.012

The corresponding exponential regression fit is shown in Figure 3.8:

Figure 3.8: Exponential Regression Benchmark



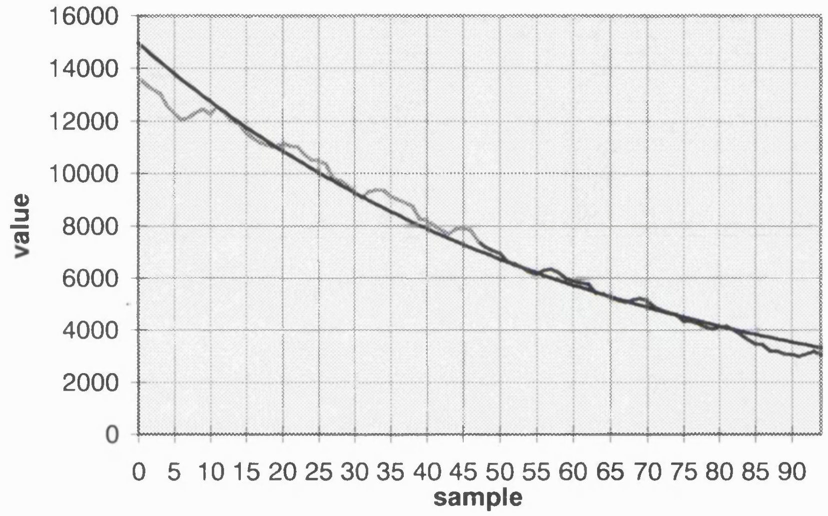
The behavioural properties for this means of modelling are shown in Table 3.3.

Table 3.3: Exponential Regression Benchmark

	Mean square error	**Mean forecast value error**
Training set	108921	£330
Test set	351352	£593

From a visual inspection of the data points and the decay curve, it is clear that the regression parameters obtained by this method are reasonable, but that the regression slides out of alignment in the testing set. This is partially due to the fact that the first year's data does not lie on the decay curve that is used to model the subsequent depreciation. In turn, this largely due to the fact that vehicle residuals drop rapidly as soon as the vehicle leaves the forecourt, but then the vehicle remains 'nearly new' for a year or so. Consequently, it might not be expected that the vehicle's value decays in the same manner as in subsequent years - and indeed it does not. If this anomalous section of data is removed and the decay curve is re-fit to the *entire* data set, it can be seen from Figure 3.9 that there is intrinsic value to modelling residual values with a decay curve:

Figure 3.9: Decay curve of debatable value



Although none of this data is out of sample, it may still be instructive to look at how well it performs. See Table 3.4.

Table 3.4: Exponential Regression Benchmark 2

	Mean square error	**Mean forecast value error**
Training set	135058	£368
Test set	21912	£148

This is experimentally unacceptable, as not only does this model use all the available data, so no data is left for an independent testing, but data is selected arbitrarily so that the chosen model will work well. However, we are not in the situation that the artificial system is, as we have some prior knowledge of the behaviour of the data and some understanding of the mechanism that generates the data, and while it is debatable science to use out-sample data to improve the model, it is not difficult to imagine such tweaks being used in a business context. The result of this is likely to be an excessive confidence in the stability of the system. This will not be done here as what is wanted are benchmarks that can be compared directly with the learning systems.

3.5.5 Comment on Regression Benchmarks

From this initial exploration of the data set, it appears that these linear models do a reasonable job of modelling the decay of residual values of vehicles, and it may be the case that neural networks will not be the most appropriate solution available. In addition, these simpler modelling techniques have other benefits:

1. The number of free parameters in these simple linear models are very low. In this instance, there is less danger of over-fitting and consequently the forecasts could be expected to generalise better into the future than an approach that has many free parameters. If data is limited (it is) and attempts are made to use a technique with many degrees of freedom, then it will not be possible to either build or test the model as thoroughly. In general, an inverse, monotonic decreasing relationship exists between the complexity of the model and the thoroughness of any applied test procedure for any fixed amount of data, although the precise shape of this trade-off will depend on the specifics of the problem.

2. A simple linear model will also be much more easily understood and interpreted than a non-linear solution that has many free parameters. This can be of considerable importance when the technology is unproven and a case has to be made to management that the system should be deployed. If the system is a black box, then it will only be "trusted" once a (successful) track-record has been established. Obtaining a track-record requires the system to be in operation for some time, and this leaves the system in a chicken-and-egg situation. If the system is easily understood then it can be easier for management to have confidence in it. However, it is not impossible for management to seek a complex solution to their problem because their perception of the problem is one that requires a complex solution. This is not always the case.

3.6 Neural Network Experiments

Once the benchmarks are defined, experiments can be carried out with neural networks. So that the results of the neural network experiments can be compared directly to the benchmarks, the same division of data into training and testing sets will be followed: the system will be trained on the first half of the available data and tested on the second half as before. Two experiments have been conducted: one is a pure time-series approach, with only the raw series used as inputs, while the other is given extra information and transformations of the price series in an attempt to enhance the network's performance.

In both cases, the network topology is the same - it is a straightforward feed-forward network with 5 inputs, 3 hidden units, and a single output. The activation function used is the *logistic* function[Mast93]:

$$f(x) = \frac{1}{1+e^{-x}} \qquad \text{Equation 3.6}$$

Each price sample is scaled so that the range of prices fed to the network spans the range 0.1 → 0.9, and hence makes the most use of the range of the input layers' activation range while not severely distorting the most extreme values. The non-linear processing of the input data is left to the hidden layer and the output neuron.

The speed of convergence is not an issue with such small networks and so little data. The objective here is to find networks that can genuinely do what is required of them.

3.6.1 Training

The simulated annealing and conjugate gradient descent techniques[Mast93] mentioned here have been covered in section 2.1.6. In this section however, these techniques are used together to speed network learning[Mast93]. In both the experiments presented here the network was trained with the following algorithm:

1. Randomise weights.
2. Simulated annealing is used to find a reasonable set of starting weights.
3. The conjugate gradient method is used to minimise mean squared output error.

4. Then go to step 2 until termination criteria have been met.

A natural question would be "Why return to step 2 after training with conjugate gradient descent? Training should be complete by now, shouldn't it?" The answer is that because the system started with a random set of weights it will probably be drawn into a local minima, so simulated annealing can be used to hop into another area of the space where the network can be trained further. If a lower point in the error space can be found through the use of simulated annealing, then conjugate gradient descent can be used to efficiently take the network to the base of that local minima. This cycle continues until a pre-specified number of cycles have elapsed, a network is found that has a suitably low forecast error, or the system is unable to find better points in the space with simulated annealing.

In addition, the network's errors on the training and test sets are continuously monitored. It is expected that the training error will drop as the network is trained so that it progressively accounts for more and more of the training data. If the network were trained indefinitely, it would ultimately have just learnt the training set "by rote" and consequently have lost the ability to generalise. One of the termination criteria applied here, suggested by [Smit93], is that the error on the testing set begins to increase. This is a little on the naïve side and can represent little more than direct training on the validation set[King96]. However, it can indicate approximately where the training should be arrested, and that any further training simply captures the idiosyncrasies of the training data. Consequently, the error on the test set begins to increase as the model is then progressively built from noise or other non-transferable structure inherent in the training set. This is not an infallible approach but gives a basis for reasonable performance expectations, for what is a proof of concept system.

3.6.2 Modelling Experiment 1: Time-series Forecasting

The data presented to the network are a series of samples of the recent behaviour of the price series. The inputs are the current price, and the price 1, 2, 3, 6 and 12 months previously. The network forecasts next months residual value, and then by using this forecasted value as an input, forecasts further into the future can be created.

Figure 3.10 shows the decay path produced by the network, and Table 3.5 shows the model's error statistics.

3.6.3 Results

Figure 3.10: Out-sample Residual Value Forecasts (time-series experiment)

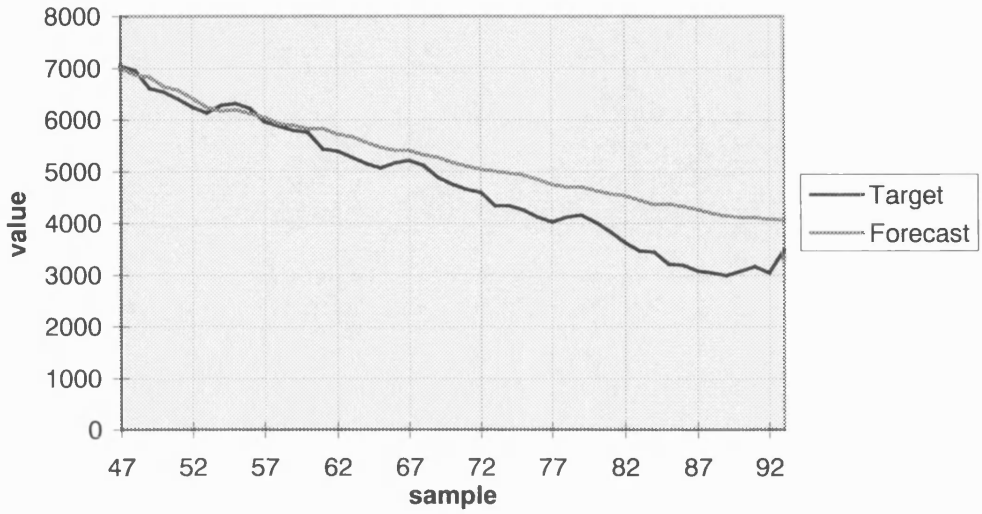


Table 3.5: Time Series Forecasting Performance

Error Measure	Error
Mean Absolute Error	£508.22
RMS Error	£627.20

3.6.4 Comment

There are a number of interesting points to note about this experiment:

- The network consistently forecasts a value that is too high. One of the limitations of neural networks is that it is difficult to determine exactly why a model behaves as it does. In this case for instance, it can be conjectured that the training set has mis-represented the problem in some way.

- The errors are comparable to those of the regression experiments. Indeed, the forecast series is nearly straight with an element of a decay curve, but little seasonal component - which might imply that the network is approximating a blend of the linear and exponential regression benchmarks.

3.6.5 Modelling Experiment 2: First Difference Forecasting

The network in this experiment is trained with the same algorithm as before: repeated applications of simulated annealing and conjugate gradient descent.

In this experiment, the network is given transformations of the input series to assist with its construction of an internal model. The new inputs to the network are:

1. Current price:

2. Sine & cosine of (2π*month number/12): This is to allow the network to take some account of the seasonality that is present in the decay curves. It was noted earlier in the chapter that seasonality is present, but not sufficiently reliable for it to be extracted with simple linear techniques. The network may be able to do a better job. Presenting the time of year in this way is continuous and preserves adjacency between December and January.

3. First difference of the price with the price six months ago: This gives the network more precise information than in the previous experiment about how residual values are changing in the short term, as changes in the price series are mapped onto much larger ranges of the neuron's range.

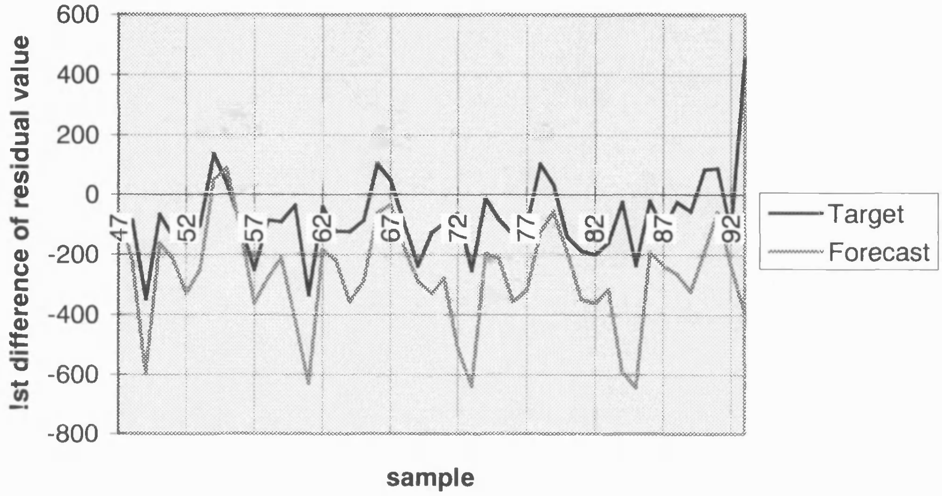
4. First difference of the price twelve months ago: Similarly, this gives more precise information about the slightly longer term rate of depreciation.

5. Target activation: First difference of current price to next months price: As the network is attempting to forecast the change in price rather than the absolute price level, the network will be much more sensitive to training errors.

3.6.6 Results

Figure 3.11 shows a chart of the target series compared to the network's actual forecasts.

Figure 3.11: Forecast of first difference of monthly residual values



The forecasted depreciation can then be reconstructed from this first difference series. This is shown in Figure 3.12:

Figure 3.12: Implied depreciation path

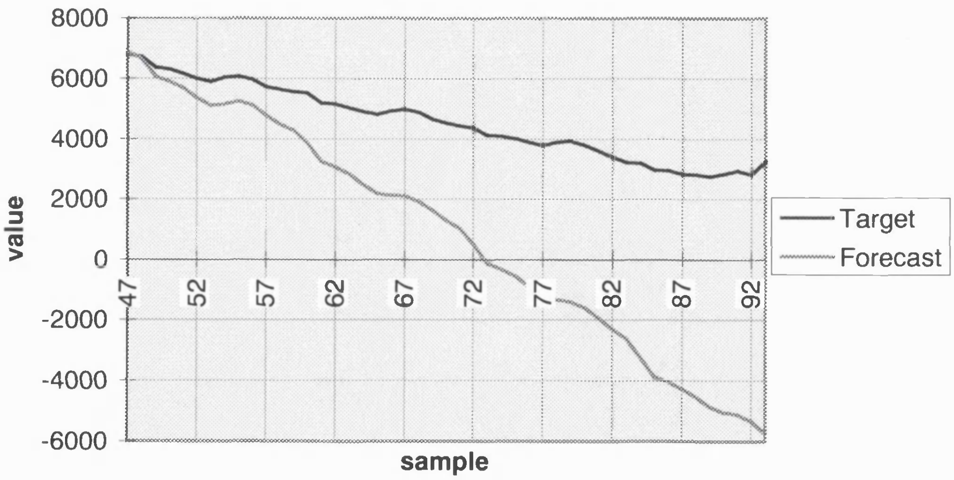


Table 3.6: Augmented Model Performance

Error Measure	Error
Mean Absolute Error	£3639.46
RMS Error	£4471.71

3.6.7 Comment

- A systematic error exists in the networks forecasts of the first difference series. This is probably due to a structural change taking place between the training and testing sets: in general, the depreciation rate of vehicles in the training set is greater than that for the testing set. A consequence of this failure of the training set to adequately represent the problem domain is that the then network over-estimates the rate of depreciation. Alternatives to partitioning the data in this way would be jack-knifing or a probabilistic partitioning of the entire data set into training and testing sets, but both of these raise problems:

  i) In its most extreme form jack-knifing[LeBa91] is the process of training multiple networks on all data except one point (the individual point unique to each network), and then testing the network on that point. For n data points, there are n ways of selecting the odd-one-out, so n networks are trained - one for each of the n training set/test set configurations. This then allows the use of the maximal amount of data in the training phase. Jack-knifing would probably give the best results, but would require many times more time to execute, and is unlikely to transform the system from a non-viable one to one that worked well.

  ii) Probabilistically partitioning the data into training and testing sets runs the risk of building training and test sets that are very similar. Given that a time-series forecasting approach requires the individual samples to be "windows" of the series that span periods of time, one of two situations will arise: either the training and test sets will "overlap" to some extent, or very few samples will be available. The latter is unacceptable as little data is available to start with, while if the data sets overlap, the test set could probably be nearly entirely reconstructed from the training set. This will then not constitute independent testing, and simply be a measure of whether the network has learnt the training set or not.

- Small systematic errors in the first difference forecasts become cumulative when the implied depreciation forecast series is reconstructed from the first difference forecasts. As a result the RMS errors between the target and the implied depreciation series rapidly grow to levels where the forecasts are not useful.
- The network has done a reasonable job of modelling the seasonality of the vehicle market. This is a non-trivial exercise but unfortunately the systematic errors in the forecasts make this model unacceptable.

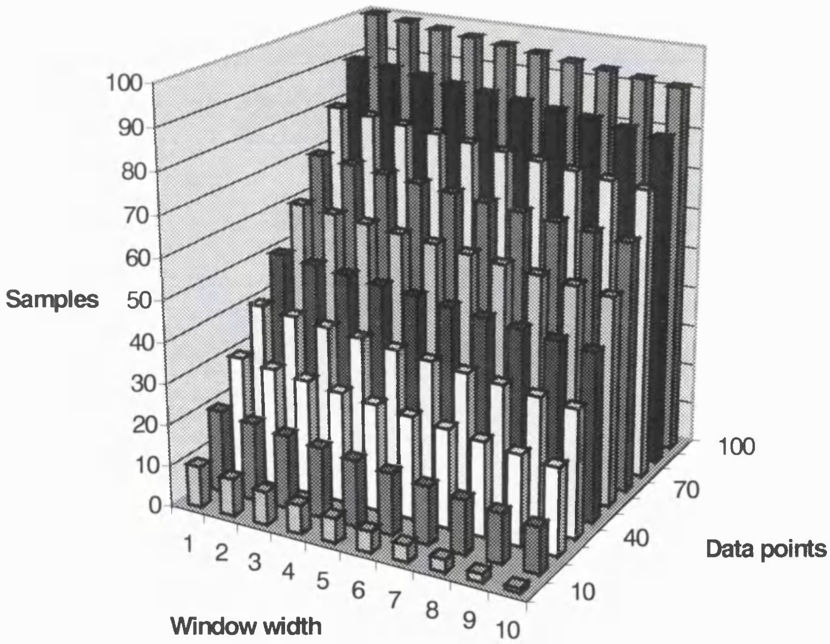
3.7 Discussion

The goal of this project was to make residual value forecasts of sufficient accuracy for the system to be of use in the setting of hire charges, and if possible to make residual value forecasts to within £50 of the eventual value of the vehicle.

As these experiments have shown, time-series analysis is probably not the best way to use neural networks for this forecasting problem. It is debatable whether the information required to make forecasts to within £50 of the eventual residual value is actually in the data that was used. While this error target has not been met, simple systems have been described that generate reasonable forecasts from the available data. While an RMS forecast error of approximately 7% gives an error of nearly £500 in the residual value forecast, this error could be viewed as being spread over 36 payments during the 3 year hire period. To use Figure 3.2, over the 3 years from February 92 to February 95, an average E-plate vehicle would have depreciated from £11200 to £7000. This depreciation rate is approximately £117 per month. The actual hire charges are usually of the order of 50% greater than the cost of the anticipated depreciation due to maintenance agreements, profit margins etc., so the monthly payments on this vehicle would be approximately £175. A 7% RMS error in the residual forecast result in errors in the depreciation rate of £8.17, which result in a final error in the monthly hire charge that is both under 5%, and £10.

A forecast of next month's residual value is not of mission critical importance, and the use of longer forecast horizons would make the longer term forecasts much more accurate. However, as the width of the required samples increases, correspondingly fewer samples from the data are available. For a window size w, and number of data points n, there will be n-w+1 samples that can be submitted to the network. If lots of data is available then the window size w can be large before it has much impact on the number of training examples. This relationship is shown in Figure 3.13. However, data is limited and hence this is an important factor. In order to use longer forecast horizons the networks would have to be trained and tested with very few examples. At this stage in the project the task is to assess the viability of time-series forecasting and having many samples of network behaviour enables a more informed assessment to be made. This type of experiment allows informed expectations to be made of the likely performance of the system.

Figure 3.13: No. of samples for given no. of data points and window width



The underlying motivation for this project was to assist with the setting of hire charges. It is important to recognise that residual value forecasting is not really what has been done. The focus of this experiment has drifted slightly from forecasting vehicle residual values to the time-series forecasting of the synthetic depreciation series.

The synthetic series was constructed for proof-of-concept purposes, to allow the development of the software and methodology, and to assess whether a time-series forecasting approach will work. It is clear from the regression experiments (Section 3.5, Tables 3.2, 3.3) that the bulk of the modelling that is required for the synthetic series can be done with basic linear models. Furthermore, from the amount of data that is available, only simple models could be induced by an autonomous model building system and still be tested thoroughly.

The decay of residual values is intimately linked to the prevailing economic circumstances at the time. This has been ignored in this study, and all forecasting has been attempted solely from the time-series. The use of macro-economic data would be valuable if past forecasts of the economics were available in sufficient numbers so that a network could be trained to associate forecasts with depreciation. However, economic forecasts are often unreliable, and it is not impossible that this project would effectively be attempting to make a connection between a pair of unrelated data sets, namely the economic forecasts and the actual residual depreciation that took place. This can be done on in-sample data, but if the two data sets really are unconnected then the out-sample results will be meaningless.

Useful results and conclusion have come from this project, but it is clear that an investigation of greater depth is required to make the best use of the available data, the technologies and explore the possibility of using additional information.

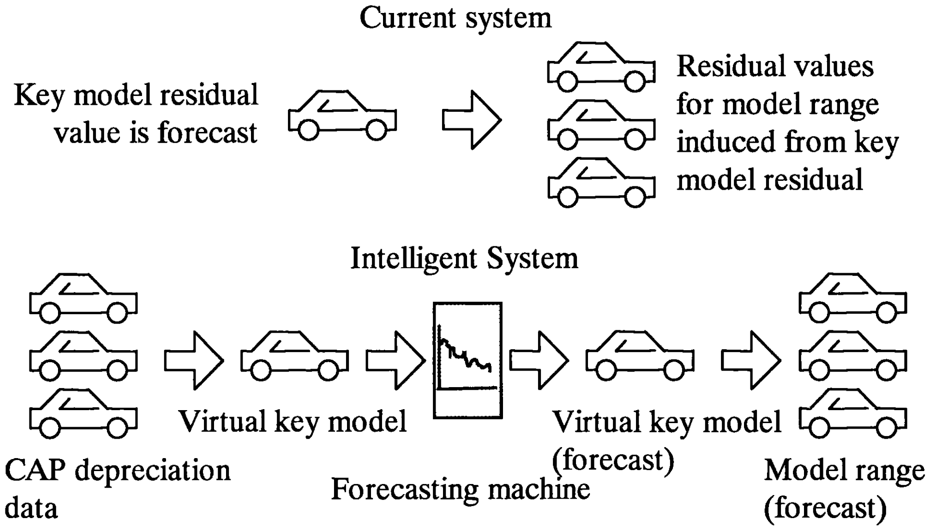
3.8 Further Work

Given that the regression forecasts are good enough for them to be of value, a methodology for efficiently forecasting individual residual values will now be desribed.

3.8.1 Virtual Key Models

It is potentially possible to produce forecasts for individual vehicles. The current process at the company for setting residual values hinges on the concept of a *key model*, that is, the most important vehicle in a model range. See Figure 3.14. The residual value is forecasted for this vehicle, and then forecasts for the other models in the range are derived from this forecast by an expert deciding whether the other models will depreciate faster or slower than the key model. The process of forecasting the residual values for an entire range of vehicles is now a more consistent and straightforward process than if the residual values were forecast for each vehicle individually.

Figure 3.14: Virtual Key Model System



The key model concept can be borrowed and applied to the intelligent system. Through modifying the filtering process that can build time-series from the raw CAP Blackbook database, the *virtual key models* can be built in the same way. This would take a straightforward modification to the filter that extracts data from the CAP Blackbook data set. Instead of filtering out vehicles of a certain registration or manufacturer, the data for vehicles of a certain model range can be filtered out. This approach is likely to be slightly more stable than simply trying to extract data for the key model, as the model range will be a group of all the vehicles whose depreciation behaviour is relevant. Moreover, this sample will be an order of magnitude larger than the information set for the key model on its own. However, one problem that is unavoidable with a system like this is that for new vehicles, where no depreciation data is available, there is no autonomous method for the system to reconstruct the residual value forecast for the virtual key model forecast.

Once the residual value for the virtual key model has been forecast, the implied residual values of the vehicles can be reconstructed from that model range, by comparing the actual depreciation of the virtual key model with each of the individual models in that range.

3.9 Summary

The salient points of this chapter are as follows:

- The modelling problem was to forecast residual values for ex-fleet hire vehicles.
- This was attempted with a database of the values of both new and used vehicles, in a range of ages, conditions and specifications.
- Linear and exponential regression benchmarks were compared to non-linear, neural network models for this forecasting problem.
- The linear models worked as well as the neural network time-series forecasting approach.
- All models produce errors that are too large for the system to meet its original brief, but the system could definitely be used as a valuable sanity check against the residual values stored in the database that are used for pricing.
- It is vital to ensure that the training set adequately represents the problem domain.
- For some problems, non-linear techniques need not be better than linear methods.

Chapter 4:

Genetic Algorithms for Trade Filtering

This experiment is to attempt to capture and extend expert knowledge about financial trading in a rule-based system. A history of trades was supplied by an industrial partner, together with a description of the market at the time of each trade's entry and the outcome of that trade. This unusual data set was used in conjunction with a genetic algorithm rule induction engine to both create a partial model of the expert's knowledge and discover new rules inferable from the data.

4.1 Background

This chapter documents a collaborative project between Sabre Fund Management and The University College London. Like many fund management companies, Sabre has considerable assets under management, but the company has a very low head-count.

Sabre management was concerned that much of their trading performance was due to a single trader, and the system that he had developed over many years of trading experience. Concern lay mainly with the observation that the operation of Sabre Fund Management as a whole might be severely compromised if some accident befell this trader.

Sabre and UCL undertook an exploratory project to discover if useful sections of this trader's system and trading performance could be replicated by an intelligent system. It is often difficult for experts to divulge their knowledge in a way that makes computerisation simple. Experts are either unwilling or unable to express their knowledge completely and consistently (even if their knowledge is complete and consistent), or spend the considerable amounts of time necessary in knowledge elicitation exercises[GoKh95].

Computers are rightly not often regarded as having "common sense", and it is often hard to transfer this into the machine[GoTr95]. This has particular relevance here as information "obvious" to the expert will often remain unexpressed in a knowledge elicitation exercise. Moreover, these missing sections of common sense knowledge could be vital to hold the whole framework together. This is particularly significant with expert systems, as if the system is confronted with information that is partly outside the domain of expertise, the reliability of the system collapses as the machine has no informed basis with which to interpret the new information. This is known as the *brittleness problem*[NeSi76,Zade84].

4.2 Nature of the Problem

Over a period in excess of twenty years, a trader at Sabre Fund Management had developed a method of charting based on his own views, opinions and experience. Charting is a branch of technical analysis, whose practitioners believe that profits can be made from the analysis of the movement of market price. It is a method of trading that enjoys a questionable reputation between various members of the finance and investment community[1]. It is treated with contempt by most economists and students of financial theory who assert that no value should be attributed to analysis of market histories("it is patently false"[Malk90]), while it is sometimes met with respect from other technical analysts who look on economists and financial theorists as being divorced from reality ("They[financial theorists] are simply wrong"[Schw92]). However, a level of in-fighting also exists between technical analysts about the validity and value of various techniques that have been labelled technical analysis. The validity (or otherwise) of charting will be covered in chapter 7 where the debate about the behaviour of markets will be dealt with in greater depth. At this stage, it is sufficient to note that the Federal Reserve Bank of New York has published papers that demonstrate that in special situations, in specific markets, charting can produce excess

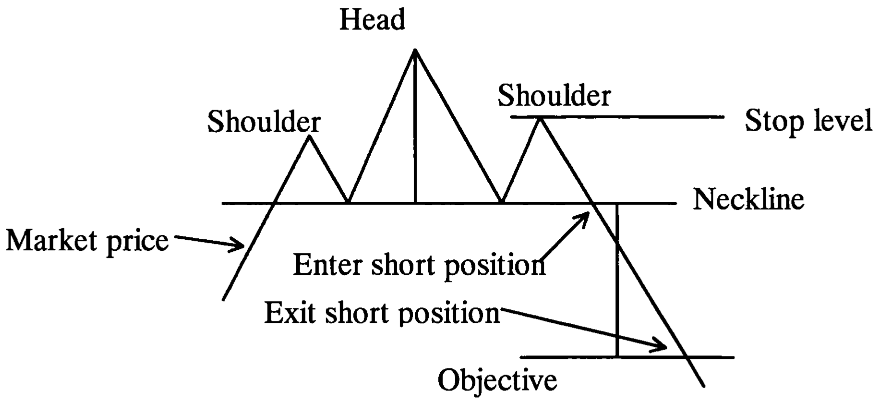
[1] While the existence of different views of market behaviour is to be expected, the passion with which these views are held can be surprising. As 'Adam Smith' writes [Smit76] "The random-walk fellows seem particularly out to get the Chartists. As I said, one random-walk professor choked on his dessert at my house at the very suggestion that charts could be taken seriously. We now have a rule that all random-walk professors must finish their desserts before the subject of charts is brought up."

profits that are inconsistent and inexplicable with conventional financial theory[OsPh95].

Charting is the process of examining the histories of the movement of market prices to attempt to find repeating patterns that have predictive value in the market. It is clear that if *reliable* patterns can be found, the trader should be able to make profits from anticipating future price movements. However, no charting trading strategy will work 100% of the time and one practitioner's opinion of the likely future behaviour of a market can often be very different to another chartists' interpretation of the same chart. In practice, the success rate for chart trades is often close to 50% (i.e. near random) and there is often little consensus between the chartists themselves about the likely future behaviour of the market. This has led to problems with the rigorous evaluation of technical strategies.

Probably the most widely known chart pattern is the "Head and Shoulders", and it is this formation that the Federal Reserve Bank cite as sometimes having predictive power in certain markets. It will be instructive to examine a chart formation in order to understand what it is that the Sabre trader has developed, and what the content of the data relates to. The Head and Shoulders construction is shown in Figure 4.1.

Figure 4.1: The Head and Shoulders Chart Formation



The Head and Shoulders formation is characterised by the market making three tops, with the central top higher than either the preceding or following tops. The neckline in this example is shown to be horizontal, but this need not be the case - what is important is that the price makes a top, retreats, makes a higher top, retraces back to the neckline and then makes a lower top than the 'head' and then returns to the neckline. At this stage, a short position is entered[2]. The position is then held until the objective is reached, which is the same distance from entry point as the neckline is from the head, or the price rises above the most recent shoulder, in which case the position is closed at a loss and the formation is said to have 'failed'. The whole pattern can be inverted, so that a series of lows is made and the trader must go long to profit from the anticipated price rise.

The reason that this pattern could have predictive power about markets is possibly because it works, possibly because lots of dealers know about it, but more likely to be because lots of people know that people know about it. Practitioners think that the Head and Shoulders formation is an important consideration, and therefore it is important irrespective of whether charting actually worked to start with. To enter into a Head and Shoulders formation as shown above, the trader must sell the market to make profits from the ensuing anticipated downwards price movement. However, the act of selling pushes the price down further. In general, when supply exceeds demand, the price drops and vice-versa. So the greater the number of traders who seek to profit from the Head and Shoulders formation, the better the formation will work[3]. And with 97% of the (1 trillion dollar) daily trade in currencies being speculative[OsPh95], there are probably sufficient numbers of traders looking at speculative situations like Heads and Shoulders to make them work.

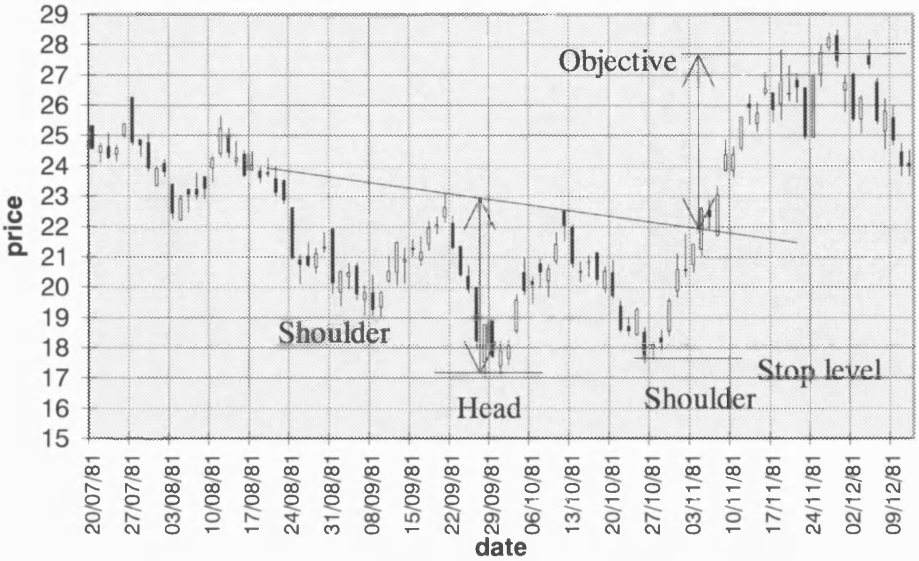
The movement of market prices is clearly much more noisy than in example 4.1, and so for completeness, a formation that could be called a Head and Shoulders is shown here in the US Bond Futures Market. However, there is significant scope for interpretation of charts, and some chartists may disagree that the formation shown actually is a Head and Shoulders. This is one factor that has prevented the serious study of charting.

---

[2] A short position is one that profits from a price fall. It is done by borrowing stock and then selling it, with the intent of buying it back at a lower price, and returning it to the owner. If the price drops, the trader makes a net gain, if the price rises, the trader loses. In the futures markets, as the trade is simply a contract to buy or sell in the future, nothing needs to be borrowed.

[3] This is clearly not the whole story as there are such influences as traders can have a range of position sizes, opinions will vary continuously about whether the formation will 'work' and the market is constantly being affected by exogenous factors such as the outbreak of news.

Figure 4.2: US Bond Government Bond Futures' Market



The charting system that the Sabre trader had developed is much more complex than this, but contains many of the same basic features as the Head and Shoulders pattern. He had identified and recorded many extra features that included the prevailing market conditions, an assessment of the strength of the long and medium term trends, the pattern type (for example, Head and Shoulders), a range of sub-patterns and the elapsed time between critical events and points. In addition to the assessment of the market conditions, the Sabre trader also recorded information as to whether the market conditions represented a good example of the formation of the pattern, and whether the trade was ultimately profitable or not.

It is this final datum that enables the investigation of charting - it will be possible to discover which, if any chart formations are reliably profitable, which are unreliable, and which reliably incur losses. With an autonomous learning approach to this problem, it may be possible for the system to invent new charting trades of its own.

The data available for the project is very unusual - the trader has been very systematic and thorough in maintaining records for a period in excess of twenty years during the development of his system. In addition, this is information that few dealers would be either prepared to divulge or even be in a position to reveal. Owing to the difficulty of extracting expert knowledge, and the availability of this very useful, interesting and unusual data set, it was decided to try an autonomous learning systems approach in order to capture the expert's knowledge.

4.3 Data

The data from Sabre is in the form of a record of 3130 trades conducted a period of 22 years, and a number of observations about the prevailing market conditions and the current pattern that the trader is attempting to exploit. An illustrative excerpt from the data is shown below in table 4.1. Each row is a separate trade.

Table 4.1: Sabre Trading Data

Type	CoCo	Pat	Sub type	pat-dur	Conc dur	Context	pt cnc by	pt cnc ing	dy 2 cep	trend med	trend lng	good ex	success
Bear	{FTO	DESR	1	15	15	DES	TTR	BT	10	N	A	3	F
Bear	{FTO	TTR	1	9	23	DES	-0-	BT	2	N	N	4	F
Bear	{DJT	AR	2	9	28	DES	-0-	TR	3	N	N	4	F
Bull	{DJT	ASCR	2	13	-0-	-0-	-0-	TR	10	N	N	4	F
Bear	yCS	DESC	1	9	9	DES	-0-	BT	1	F	F	4	S

The fields recorded for each trade are:

Type: [Bull/Bear] - The overall prevailing market direction.

Contract Code: Sabre trades in all equity sectors, commodities, currencies, government bonds and futures. Each market has a unique identifier.

Pattern: The chartist has identified 10 different chart patterns that he considers to be of value.

Sub-type: Each pattern can occur in one of 5 sub-types.

Pattern Duration(weeks): The length of the pattern.

Concluding Pattern Duration(days): The length of the final phase of the pattern.

Context: An assessment of the market character.

Pattern Concluded by: The pattern following the main pattern.

Concluding Pattern Name: The pattern of the final phase of the complete chart formation.

Days to Confirmed Entry Point: Time elapsed between critical points on pattern and the trade entry point.

Medium Trend: Assessment of whether the trade is with or against the medium term trend.

Long Trend: Assessment of whether the trade is with or against the long term trend.

Good Example: Assessment of whether the market closely corresponds to the chart pattern.

Success: Was the trade a success or failure?

Some fields such as month, year and entry/exit price have been removed for reasons of space and clarity.

It is interesting to note that the semantics of the labels here are practically irrelevant - a genetic algorithm will be used to try to find regularities in trading record, and so what the actual patterns are is not important. Provided that the chartist interprets the codes in a consistent manner, it does not matter what the patterns that they refer to actually are.

4.3.1 Data Exploration

The aim is to identify any predictability that may exist in the data. To do this the empirical probability of success will be compared to the conditional probability of success given that a variable is in a particular state. P_v is the probability that a case selected from the sub-set is a success, whereas P_a is the probability that a case selected from the entire history of n cases is a success. The null hypothesis:

$$H_0: P_v = P_a$$

is tested against the alternative hypothesis:

$$H_0: P_v \neq P_a$$

The test statistic is[Hays81]:

$$Z = \frac{P_v - P_a}{\sigma_{P_v}} = \frac{P_v - P_a}{\sqrt{\frac{P_a(1-P_a)}{n}}}$$

If the sample size, n, is below 30, then tables must be used to correct for the limited number of degrees of freedom. If a confidence interval of 95% is assumed then the null hypothesis can be rejected if Z exceeds $Z_{\alpha/2}$ where $\alpha = 0.05$. This value is 1.96. If the Z-score exceeds 2.56 then the implied confidence in the hypothesis is 99%[Hays81]. The value of the Z-score can be used to calculate the significance of the result.

To use this test statistic, the empirical probability of success is required to compare sub-sets of the data to. This is shown in Table 4.2.

Table 4.2: Complete Data Set

Data	Samples	Successes	%Success	Z-Score
All	3130	1249	39.9	N/A

The empirical probability of success is 0.399 from the entire data set.

Each of the variables (listed in section 4.3) were examined with this test, and isolated types of Contract Codes, Contexts and Concluding Pattern Names are significant at either the 95% or even the 99% level. These results have not been presented because, for instance, a success rate and significant Z-score for a pattern of type "ABR" conveys no useful information, but where the results are interpretable the tables are given below in Tables 4.3-4.6. However, at this stage it should be noted that most of the variables have at least one state where the test statistic is significant at the 95% level.

Table 4.3: Pattern Type

Pattern Type	Samples	Successes	%Success	Z-Score
Bull	1975	774	39.2	-0.65
Bear	1155	475	41.1	0.85

Table 4.4: Medium Term Trend

Medium Term Trend	Samples	Successes	%Success	Z-Score
Against	417	161	38.6	-0.54
Neutral	1653	649	39.3	-0.53
For	1060	439	41.1	1.00

Table 4.5: Long Term Trend

Long Term Trend	Samples	Successes	%Success	Z-Score
Against	481	179	37.2	-1.20
Neutral	1637	637	38.9	-0.82
For	1012	433	42.8	1.87

Table 4.6: Good Example

Good Example	Samples	Successes	%Success	Z-Score
1 (worst)	89	30	33.7	-1.19
2	573	213	37.2	-1.33
3	1410	501	35.5	-3.35
4 (best)	1058	505	47.7	5.20

4.3.2 Comment

There are a number of observations that can be made about these Z-scores. Negative Z-scores indicate that the conditional probability of success is lower than the empirical probability of success, and hence these are situations that should be avoided.

Pattern Type: There is nothing statistically significant in these results.

Medium Term Trend: Although the Z-scores are higher, they are of insufficient magnitude to be significant at the 95% level.

Long Term Trend: Going with the trend enhances the probability of a successful trade to a degree significant at the 90% level.

Good Example: The Z-scores here are extreme - if the current market is a good example of the chart pattern, this will improve the probability of success beyond the 99% confidence interval. Fairly good examples have large negative Z-scores, which implies that unless the market is a very good example, the trade will probably fail.

To a casual glance even the best success rate of 47.7% (Table 4.6) would not seem to be enough to make profits, and would be out-performed by simply tossing a coin. However, because the objectives and stop levels are not the same size, the trader can take several losses for each winning trade. Sabre sought a mean win: mean loss ratio of 4:1, so only one out of every 5 trades needs to work for the trader to break even before costs.

4.3.3 Data Preprocessing

The original data set consisted of 3130 entries. The data was separated probabilistically into the sets shown in Table 4.7:

Table 4.7: Data Partitioning

Data Set	Size
In-sample	1375
Validation	700
Out-sample	1055

Partitioning the data in this way ensures that there are no time dependencies within any of the data sets.

4.4 Genetic Algorithm Rule Induction

It is extremely unlikely that it is possible to build a complete model of market behaviour from the 1375 training examples that are available. Instead of this, the system can search for pockets of predictability. Instead of attempting to build a complete model, the system can simply remain out of the market until it has mechanics that bear significant relation to our partial model. Positions are entered when the current market behaviour is recognised as having some deterministic component.

One method of finding these pockets of predictability is through the use of genetic algorithm rule induction. This has already been covered in Chapter 2, so only an overview of the system will be given here.

4.4.1 Representation

In order to evolve rules describing successful charting trades, it is clearly necessary to devise an encoding that will allow the GA to explore the space of chart trades. In the previous sections, the effects of individual conditions upon the probability of success were examined: for example a trade is much more likely to work if the trade is a very good example. To enable the GA to explore *multi-variate* conditions, a framework must be provided that will allow the GA to formulate such queries. If it is acceptable to restrict the GA to a fixed number of variables or conditions on any given execution, then the action of the genetic operators is as already described in Chapter 2, and the encoding scheme is straightforward:

Figure 4.3: GA Encoding Scheme

Variable1 id	Variable2 id	...	Variable n id
Variable1 value	Variable2 value	...	Variable n value

The number of parameters for each rule is specified in advance for each run of the GA. Each gene has a pair of values:

1. Variable ID: This is a symbolic representation of a particular observable. Possible values would decode to Type, Pattern, Pattern Duration, Good Example, Days to Confirmed Entry Point etc.
2. Variable value: This is the condition that the variable above must have for the rule to fire. Some variables, such as Days to Confirmed Entry Point are real-valued, and so these cannot be coded in the same way as an observable that has symbolic values like "Bull" or "Bear". Observables that are real-valued are clustered into 5 bins using the k-means algorithm[Hart75].

This encoding scheme has restricted the GA to search for rules that have a prespecified number of conditions. This is not a problem as the GA can simply be executed several times, each time searching for rules with a different number of parameters.

4.4.2 Rule Evaluation

Once the GA has assembled a rule, it needs to be evaluated in order for the selection process in the evolutionary cycle to operate. This is achieved by examining the data for trades where all the rules' contingent clauses would be met by the description of a past trade. When a match is found, the success rate of the GA's rule is updated from the success or failure of the trade in the data that matched the rule. If no trades can be found where all the conditions are met, the rule's fitness is undefined and the rule is eliminated. Otherwise, the fitness of the rule is given by the Z-score, which is a combination of the number of activations and the improved probability of a successful trade outcome. In this way, the GA searches for rules that both fire frequently and have a high conditional probability of success.

4.4.3 Genetic Algorithm Execution

The GA used to evolve charting rules had the following properties:

1. A "simple" GA was used with tournament selection, to search for rules with a randomly chosen number of conditions.
2. A population of 50 was used for 20 generations.
3. Elitism was used to ensure that the fittest individual always survives through to the next round.
4. A level of mutation (probability per cycle per individual of 0.1) was used to assist the GA in exploring the space.
5. After the 20 passes through the evolutionary cycle, the fittest individual is output. The GA had usually converged(save for mutation) or was nearing convergence by this stage.
6. The GA would then re-execute itself. The net result of this was that a number of rules would be evolved up with varying numbers of clauses in the rules.
7. Rules were then tested on the validation set, and any that failed to fire were removed. The remaining rules were then tested on the out-sample data.

4.4.4 Results

The rules that have been found by the GA on the in-sample and validation sets are tested on out-sample data that is disjoint from data used in the construction of the rules. Again, a table of rules of the form "Pattern: TTR; Subtype: 4; Context: DES" etc., presents little useful information. The most useful way of presenting the results is to show the distributions of Z-scores for rules with a specific number of parameters. (Figure 4.4). Rules that failed to fire out-of-sample have been ignored.

Figure 4.4: Distribution of Z-scores for rules on out-sample data

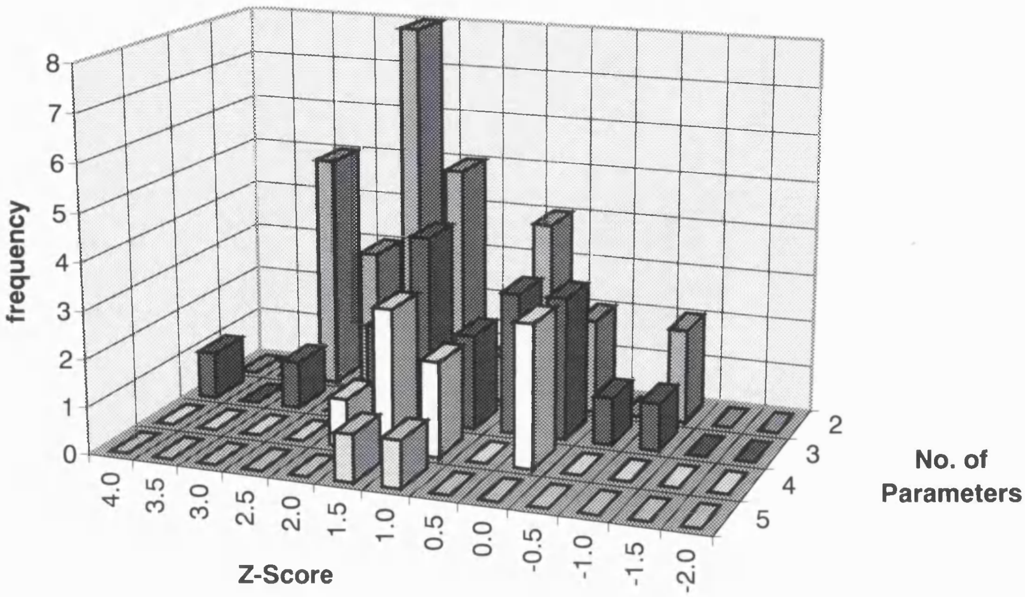


Figure 4.4 and Table 4.8 clearly show that larger numbers of higher Z-scores are obtained with the rules with fewest variables. The mean Z-score for 5 parameter rules is greater than the mean Z-scores for the 3 and 4 parameter rules, but the sample size is only 2 and the significance of this result is minimal.

Table 4.8: Mean out-sample Z-scores

No. of Parameters	No. of Rules	Mean Z-Score
2	30	1.43
3	18	0.83
4	9	0.71
5	2	1.01

This mean Z-score for 2 parameter rules is quite high, but not high enough to reject the null hypothesis at the 95% confidence level. One of the best intelligible rules found by the system was "Long Term Trend = FOR; Good Example = 4", which has a certain intuitive appeal.

4.5 Discussion

It is interesting to note that it is the simple patterns that appear to work better than more complex rules with many conditions. There may not be enough data to justify building complex rules with more than 2 parameters, and finding such rules might be just progressively fitting noise in the market. The single largest Z-score recorded in the entire investigation was for the condition “Good Example = 4”, which is clearly a 1 parameter rule. This result that simple rules are more robust is consistent with the findings of Oussaidène et. al.[OCPT97].

The fact that the out-sample results have positively skewed Z-scores implies that charting is not completely random and hence could have value. This is not a proof that charting works, just a demonstration that a chartist’s interpretation can be systematic, which in turn demonstrates that charts do present information that is useful. However, this does not constitute a proof of the value of charting because:

1. It is impossible to determine from the information available whether the trader has out-performed or under-performed the market. Charting will only be considered of value if it permits the traders to out-perform the market *on a risk adjusted basis.* Otherwise, the trader could simply invest his capital in an efficient index tracking fund[4]. However, debate exists about whether risk adjusting returns (a theoretical finance activity) is sensible for a trading scheme that should not work if theoretical finance is correct. This is covered in greater depth in Chapter 6.
2. It is impossible to say what extra information the trader uses in his everyday trading activity. Schwager[Schw84] asserts that methodologies such as charting, technical analysis and fundamental analysis do not have intrinsic value in themselves, but that they provide the trader with a framework with which to think about the market. If the information content comes from elsewhere, then it is not impossible that charting is of no value other than as being a mental framework. It is clear from these results however, that there is some intrinsic information content to the chart formations. If this was not the case the GA would not have been able to find as many rules as it did that had out-sample Z-scores significant at the 95% level, but it is not clear as to whether this would allow a trader to outperform the market.

[4] Many financial theorists claim that investing in portfolios that track stock market indices is, on average, the best an investor can do.

One criticism of this work would be that this is only a single sample path, and it could all be coincidence from a random data set. This is unlikely as it can be seen that the *distributions* are starting to appear in figure 4.4, and 17% of the rules are significant at the 95% level. If this was completely random, and given that the distributions are starting to appear, it would be expected that approximately 5% of the rules would exist in this section of the Z-score distribution. It is interesting to note that the distribution is loosely normal, and so there is probably some random element at work.

Stationarity could be an issue - the time dependencies were removed from the trading history by probabilisticly dividing the records into the in-sample, validation and out-sample sets, and it is possible that the nature of markets have changed over this period in an intrinsic way. For example, the rates and means of information dissemination have changed and new financial products have become available. Moreover, the chartist himself may have changed some opinions about charting over the 22 years that he has been developing the system he currently operates. It is also not inconceivable that the position sizes he enters now are different from those he used to place when he first began developing his system. The larger the position sizes, the greater their effect on the market and so slightly different approaches need to be taken towards trading.

The system found some patterns that provoked some interesting reactions from the trader. These reactions were usually one of two types:

1. "Hmmm, I never thought of that, that's very good..."

2. "This appears to work well, but the pattern doesn't mean anything."

It is interesting that some of the trader's 'home-grown' patterns had very high, but negative Z-scores. This means that the trader had thought that he had found a pattern that worked, but in fact, it had a reliably lower success rate than simply behaving randomly.

4.6 Summary

The main points of this chapter were as follows:

- The problem was to capture expert knowledge about trading in financial markets without the need for knowledge elicitation exercises.
- The expert knowledge under study in this chapter was a charting methodology - a way of trading based on the analysis of the movement of the market price.
- The data available was a history of past trades - the market conditions that led to the trade being entered and the trade outcome.
- This data was analysed to see if any of the patterns worked at statistically significant levels.
- A genetic algorithm rule induction engine was used to build multi-variate rules.
- Simple rules fire more often and appear to work more reliably than more complex rules.
- Information is definitely contained within charts, but it not clear that their analysis would enable the trader to out-perform the market.
- Good and useful results were obtained by the system.

Chapter 5:

The Continually Adaptive Trading Engine[1]

This chapter describes an experiment in building a system that can remain attuned to changing market conditions. The data available were closing prices for the stocks in the FTSE 100 index from 7th January 1983 to 13th December 1995. Genetic algorithm operators are used on a flexible trading system framework to rapidly build trading systems that are effective in the current markets.

5.1 Background

Financial markets are in a constant state of flux. The relationships between macro-economic and market variables and relationships between market variables themselves are constantly changing. Therefore, any empirical model constructed by deriving relationships among observed variables will only be valid for a limited time period as the values for the parameters slide out of relevance. This is a criticism of most 'static' mechanical trading systems which do not possess the machinery to cope with such dynamically changing financial market behaviour.

Recently there has been an explosion of interest in applying 'adaptive' intelligent techniques for financial market trading. Neural networks[BeJa90], genetic algorithms[Davi91], fuzzy logic [Zade84] and intelligent hybrid technique[GoKh95] based systems have recently been applied to equities, bonds and foreign exchange trading [Debo94, GoFe94, GoTr95].

---

[1] This chapter is based on J.W. Viner and S. Goonatilake, "*A Study of a Genetic Algorithm Based Continually Adaptive Trading Engine*", Proceedings of Intelligent Technologies in the Human Related Sciences, 1996.

In the development of such adaptive trading systems there are usually four main steps that are carried out:

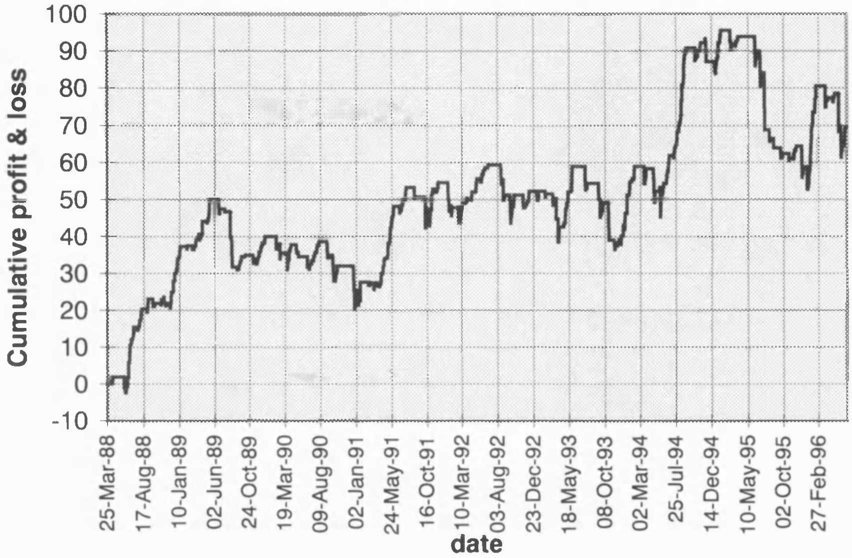
1. Training the algorithm using in-sample data (learning the relationships)
2. Validating the models
3. Testing the model using out-of-sample data
4. Trading the models

In the training phase these methods are presented with training pairs consisting of independent variables (e.g. transformations of prices) and target variables (e.g. price in 10 days up/down), and at each cycle the algorithm attempts to minimise the classification error. This process is repeated until a low level of prediction (classification) error is obtained. In the case of a genetic algorithm, the GA cycle is applied until a low level of error is obtained or until the population converges to a 'good' (set of) solution(s). The models are then usually validated on another data set (validation set) to assess the robustness of the models. Finally, the models which are considered to be robust are applied to genuine out-of-sample data.

However, adaptive systems based on this type, similar to purely 'static' technical trading systems, can have highly variable performance, from highly profitable in some years to sustained loss taking in others, in markets of a significantly different character. This is shown in figure 5.1, where a fixed moving average strategy is traded for a period of 8 years on the US Bond. It is clear that sometimes the system fails to produce positive returns for periods of nearly 2 years, but for periods where the system is in tune with the market, the rate of profit making can be alarming.

This observation is partially responsible for undertaking the work presented in this chapter which adopts a different approach to genetic algorithm based financial trading. The genetic algorithm is not "trained" in a conventional sense, to make the population converge or to produce a set of rules that have a low classification error, which are then applied to out-of-sample data. Instead, the genetic algorithm is continually producing new trading models (new rules) which are "pretend" traded on very recent market data. The rules which prove to be profitable in the current market climate, are then applied for conducting (simulated) trades.

Figure 5.1: Cumulative P&L curve for typical static trading system



This chapter is concerned with the concept and development of a continually adaptive trading engine that is applied to equities. One of the main motives for this work is to develop an evolutionary framework that can adapt and hence survive changes in dynamic financial markets. The genetic algorithm based framework presented here provides an automated way to conduct a search of trading strategies using existing technical indicators (e.g. moving average crossover and oscillator strategies). As the genetically derived trading models (rules) are very transparent, they have significant advantages over neural network based trading models which are inherently 'black-box' trading systems. This difficulty with assessing exactly what it is that the black box does has led to constant criticism of systems built on such techniques[Eco92].

5.2 Data

The data available is the closing prices of stocks in the FTSE 100 index between 7th January 1983 and 13th December 1995. Some stocks are not available for all of this period as the they did not actually exist in 1983. Similarly, some stocks have been ejected from the index as their market capitalisation no longer places them in the index.

Data for those companies which have dropped out of the index at some point between 7/1/83 and 13/12/95 was not used in these experiments. It should be noted that this introduces *survivorship bias.*

The data for each company is partitioned into two sets:

1. An initialisation phase that runs from when data for the company first became available to 18th June 1991.
2. A testing phase which runs from 19th June 1991 to the end of the data. This is a period of 1000 trading days.

Table 5.1: FTSE data format

Date	Price
12/5/95	919.00
15/5/95	910.00
16/5/95	899.00
17/5/95	906.00

The data is in the format [date][closing stock price], and an example is shown in Table 5.1. It is important to realise that in a very real sense, the closing price does not actually exist - it is not possible to enter at the close, as by definition the market will have shut. In addition, in some markets it is not the final price of the day. It is a synthetic value derived from the prices of the last few trades. However, it is best proxy to the market price that is available without resorting to intraday data, and it will rarely deviate dramatically from the price a trader would have got by entering a trade "at-the-market"[2] 60 seconds before the market closes. In normal market conditions, 60 seconds is sufficient for the order to be filled in liquid markets, and stocks in the FTSE will usually be liquid enough for the trading simulation to be representative. Clearly there are exceptions to this - the stock market crash of 19th of October 1987 is an

[2] A trader can enter different types of market orders -he can specify a bid or offer at a particular price, without being sure that the order will be filled, or entered "at-the-market", at the current market price.

example, but these high deviation events are rare and anomalous to everyday market activity.[3]

Unlike futures, these series do not need to be adjusted in any way to make them continuous.

5.3 Technical Trading Strategies

A framework has been developed as part of this work for the description of technical trading systems, which can then be searched with a genetic algorithm. At present, the framework is simple but could be extended to encompass a wider range of possibilities that are based around other indicators, data streams, and more advanced technical relations. At this stage of development, the system is only designed to time entry and exit points for a single, pre-specified security.

The basic idea is that a trading strategy is composed of two principal elements: a module that generates buy signals, and a module that generates sell signals. In conjunction, these two modules constitute a strategy with which the system can enter and exit the market. The Continually Adaptive Trading Engine is then built from a group of these individual trading strategies, each trading the market separately.

Each of these buy or sell modules is simple, with at most two parameters, and is based on the behaviour of a single technical indicator. Below is a summary of the technical indicators supported, a description of the market dynamics that trigger buy or sell signals, and the principal pitfalls associated with each of these indicators[4][Kauf87]:

5.3.1 Technical Indicators

Moving averages: A moving average is the average of the last n periods prices, whether monthly, weekly daily or hourly or even shorter time-frames. If a short moving average is above a longer moving average, this is because the price has had a

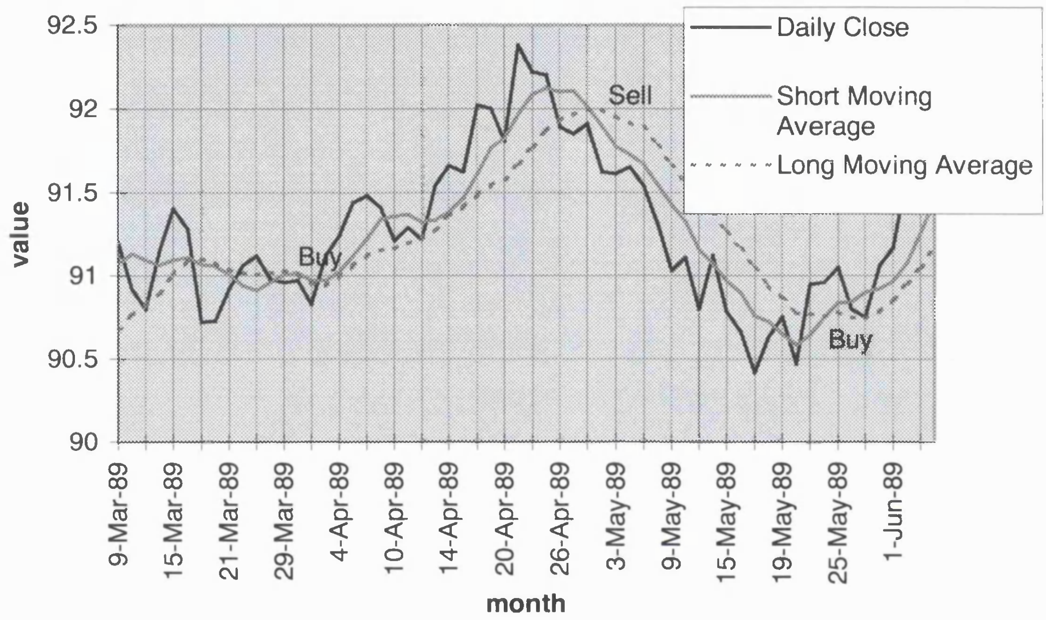
---

[3] Market practitioners have been aware of these "fat tails" in the distribution of daily price movements since at least as far back as 1959[Osbo59]. There has been significant debate ever since over what impact this observation has for the validity of many aspects of financial theory.

[4] A financial economist would say that the pitfall of all of these indicators is that they present no information, as no value can be derived from the analysis of price histories. This is a view that is coming under threat as flaws are found in the current econometric paradigm. This debate will be examined in Chapter 7.

recent upward trend [Kauf87]. This would generate a buy signal. A sell signal occurs when the short moving average crosses down through the longer moving average. An example of this is given from the German Bund 10 year Government Bond Futures market, although it is important to remember that this is an example that has been chosen to show this trading strategy working well. It is not hard to find examples where the strategy repeatedly incurs great losses.

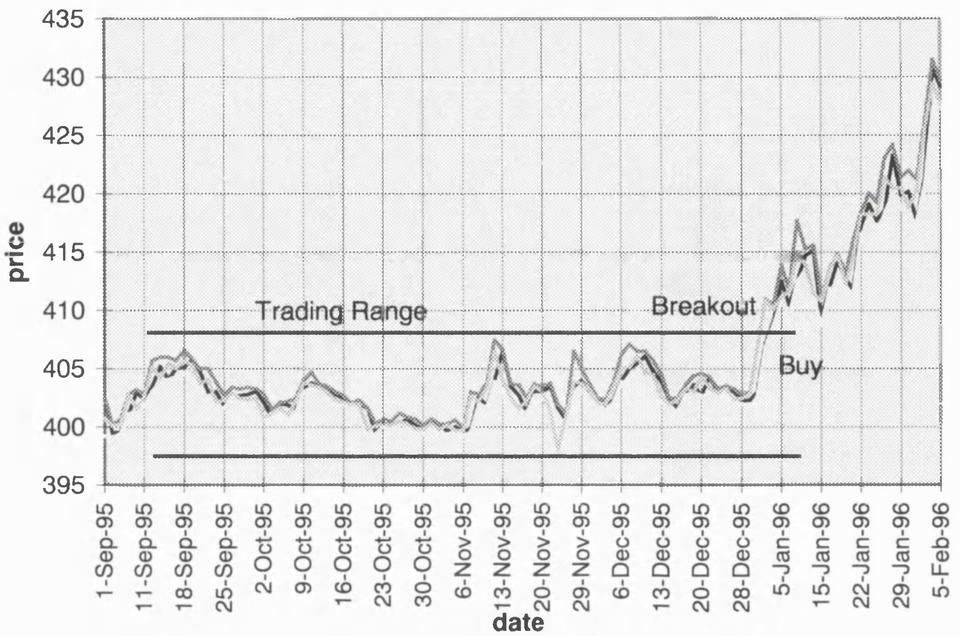
Figure 5.2: Moving averages - German Bund 10 Year Government Bond Future



This trading strategy has two parameters, namely the lengths of the two moving averages. Short moving averages are far more susceptible to random price fluctuations than longer moving averages. However, longer moving averages are in danger of being too long and being too slow to react to changes in the market for the signals they generate to be useful.

Trading Range Breakout: This is an indication of a stock becoming re-valued due to a palpable change in the value of the securities [Kauf87]. It is often meant to indicate that the market will resume its trend after a period where the market was not trending. (Figure 5.3) The main problem is with false breakouts, where the price increases to a new level but then retracts to within the original range.

Figure 5.3: Trading Range and Breakout in Gold



The simplest way of automating the trading range breakout is to use a single parameter, the length in days of the price range that must be broached by the market price. Some advocate a second parameter that gives the minimum required 'leap' from the existing range, in an attempt to protect the indicator from false breakouts[Kauf87].

In practice, the boundaries of the range are often blurred, and *squeezes* take place. Consider the situation where the price is near the top of the range - lots of traders will be short (attempt to profit from falling markets) as they expect the price to remain in the range. Other dealers can sometimes bid up the price to such a degree that all the short traders lose money as the price rises above what they sold at. If they then want to cut their losses, then they must buy the market that they sold previously. At this point, those that bid up the price in the first place hold onto their securities and restrict supply, and as the shorts try to reverse out of their trades the price rises further than it normally would. The original bidders then sell their securities at a profit.

Relative Strength Indicator: In an oscillatory market, and in the absence of new information, one would expect a stock price to remain in a range centred on the approximate value of the stock. This is also shown in Figure 5.3. As the price drops an upturn in price is expected as the stock becomes undervalued. Similarly, as a stock becomes overvalued, a downturn in the price action becomes expected. This is a range based indicator and so has one parameter. The principal pitfall is similar to that of trading range breakout - if the price is high within the recent trading range, it may be about to breakout, rather than retrace.

The relative strength indicator is calculated as follows[Kauf87]:

$$RSI = u / (u + d) \qquad \text{Equation 5.1}$$

where u is the number of days from the sample period where the market rose and d is the number of days where the market declined. These quantities are calculated for a pre-specified number of days.

5.3.2 Stop Losses

Stop losses are also part of the sell strategy. That is, if the price of the security drops below a certain fraction of the paid price, then the position is closed to minimise losses. This only requires a single parameter to specify the maximum price drop tolerated. The drawback with stops is that often a price drops slightly before it rises. Setting a stop too close means losing winners that started weakly, while setting it further away requires the suffering of larger losses in the event of a genuine price drop. The use of stops was seen as so important that every system has a stop in addition to a strategy for selling. At present they are simple and are set in place when the position is opened, although more advanced schemes are possible with intelligent stops that move with market conditions.

5.3.3 Comment

A moving average based indicator will generate signals that are in conflict with the relative strength indicator. The key to resolving these disputes is to trade the right indicators in the right markets. Table 5.2 summarises the markets that each indicator is most suited to:

Table 5.2: Appropriate market conditions for technical indicators

Market Conditions	Appropriate Indicator
Trending	Moving average
Oscillatory	Trading Range Breakout
Oscillatory	Relative Strength Indicator

At present the system is not intended to work on futures and options for a number of reasons, but primarily due to the fact that it does not have adequate risk management strategies[Kauf95b]. As futures trading can be highly leveraged, a trading system without adequate risk management can potentially lead to very large losses. In addition, the current framework does not permit the system to go short. Traders can easily go short in the futures markets, and entering the futures market with a system that cannot go short could be an unnecessary handicap. This is not a problem in the equities markets as traders cannot easily short the market.

A key advantage of this system which is built around simple technical trading indicators is that the analysis and validation is much easier than with black-box techniques such as neural networks, where it is usually not possible to decode the system's reasoning. The lack of explanatory power has been a constant criticism of neural networks and its application in financial trading[Eco92]. In contrast, a genetic algorithm based scheme which produces trading models in a rule format is easily understood by traders and can be judgementally revised if necessary.

5.4 Evolutionary Framework

Instead of trying to develop effective trading systems by hand, a genetic search of possible trading systems built on these indicators would not only be a very elegant method of evaluating a large number of different strategies, but would also unlock the possibility of continually evolving systems that change and mutate to match current market conditions. The system has then been developed to sustain a large population of systems that are constantly breeding with each other, evolving and mutating, that will always be searching for trading systems that are better attuned to the current market. Then only the systems that have been performing best in the recent past are used for trading. This is done by only acting on signals from individual systems whose recent performance exceeds some threshold. The recent mean return is required to be greater than the sum of the trading costs and a moderate mean return per trade.

One way to visualise this is as a having 'moving window rules buffer' where rules are constantly moving in and out of the cache with new rules being created from the genetic algorithms operators (crossover and mutation). The system will not necessarily predict a market character in the future, but instead when a particular market character is established it will attempt to produce a good trading strategy to exploit it. Due to the concurrent nature of the development of multiple solutions by a GA, this approach is similar in many respects to holding a portfolio of trading rules, which in turn represents a diversification of trading strategies. The Markovitzian analogy [Mark52] has been taken further than this, as it is required that the individual solutions are not highly correlated, and so the risk associated with the portfolio is reduced.

5.4.1 Strategy Representation

A key aim is to keep a large number of systems as genetic "stock" at hand, so that as the market changes, the system will either already have a suitable set of systems in hand, or very quickly be able to assemble such systems that will be most effective in the short and medium term future. If effective trading systems exist within this framework, it is likely that they will be found (or close approximations to them).

Table 5.3: Anatomy of a strategy

Buy Strategy			**Sell Strategy**			**Stop Loss**
Strategy Type	Parameter p1	Parameter p2	Strategy Type	Parameter p1	Parameter p2	% drop
Moving Average	12 day	25 day	Range Breakout	15 days	-	4

Each system is as described before in section 5.3, with a strategy for buying and a strategy for selling, each with a few attendant parameters, and a stop loss. Each strategy type could be one of moving average, trading range breakout or relative strength. If the chosen strategy does not have a second parameter then this field is ignored. It is the eventual aim to extend this framework to allow a much more sophisticated mode of operation, but the present framework allows a moderate level of complexity, and is sufficient to demonstrate the concept.

To illustrate the present scheme, the encoding of a plausible strategy is shown in Table 5.3, where the system buys if the 12 day moving average is higher than the 25 day moving average. The position is closed if the price doesn't make a breakout in 15 days, or if the price drops by 4%.

5.4.2 Genetic Trading

It is straightforward to test a trading system based on the above framework. Daily price data is fed into the machine which makes buy/sell decisions on information that would have been available at the time. The only equivalent of training in this approach is that of an initialisation period where the algorithm is run through a given set of data in order to produce good 'genetic fragments' (parts of trading strategies), that may later be used to produce good trading rules. The system keeps a record of its recent behaviour as the daily data is submitted, although sometimes it may not be possible to perform an exact evaluation of its effectiveness. For instance, if a trading rule is yet to close a position, or has never opened one at all, then any evaluation of its effectiveness must be incomplete. To protect rules that may be worthwhile but are too new to have developed a track record, the current realisable return from an open position is used as a proxy to the system performance. To prevent rules that are new and completely untested from squandering capital on unproven strategies, the recent performance of a strategy must exceed some threshold. This threshold is a combination of the trading costs and a moderate mean return per trade.

This system is not based on the conventional GA model of generational replacement and fitness-based probabilistic selection. The population is kept constant but is replaced in an iterative manner. This is so that the GA can continue to run while the individual evaluations are being performed, which takes place on a rolling basis. It is important to recognise that the evaluations are only ever partially defined. Unusually for a GA, the objective is not to get the population to converge. In some respects this is similar to Harvey's SAGA[Harv92] which is a GA also designed not to converge. Instead, the objective is a diverse set of good rules and as many good fragments as possible so that in an unforeseen shift of the market, the system can (if need be) rapidly assemble completely new rules. However, if there is no selection pressure, the mean fitness of the population will not increase. Selection pressure is then exerted by selecting either a weak or duplicate rules from the existing rule set for replacement by the new individuals that are created by crossover and mutation. Duplicate rules are removed before weak rules. A weak rule may still add value but a duplicate rule will not increase the information contained in the rules. Random rules are generated occasionally to maintain a fundamental genetic diversity.

5.5 System Testing and Performance Evaluation

Historical data is used for the stocks in the FTSE100 from the 11th February 1983 to the 13th June 1995. In order to make the experiment a genuine test as possible, the systems are initialised on all the available data up to 11th June 1991, and then the performance of the resulting population is tested on individual stocks selected at random over the interval 11th June 1991 to 13th June 1995. 10 runs (10 different populations) were executed to mimic how this system would compile a portfolio, as each population can only trade a single price series. The outputs for each of these 10 separate runs was then combined on a trade-by-trade basis and the analysis then is performed on this series of buy/sell signals that spans 10 stocks.

It is important to realise that the length of time that the system can be run for depends upon the amount of data that is available. Each day's data is only used once for active trading and the evaluation takes place over the trading performance over time of the strategy.

5.5.1 Performance Measures

The records of individual system performance were originally kept on a trade-by-trade basis, but a consequence of this was that it would take an unreasonable amount of time for the influence of a successful long term trading system to diminish in inappropriate market conditions. If this system only traded once every other year, these long trades would be kept in the record of recent performance for years. To side-step this, only the trading activity of the last 180 days counts towards the evaluation of an individual. If no positions were closed in this interval, then a proxy to the system evaluation is taken in the form of the current return if the current position was closed. If no positions at all have been opened in the last 180 days then the evaluation is undefined.

5.5.2 Results

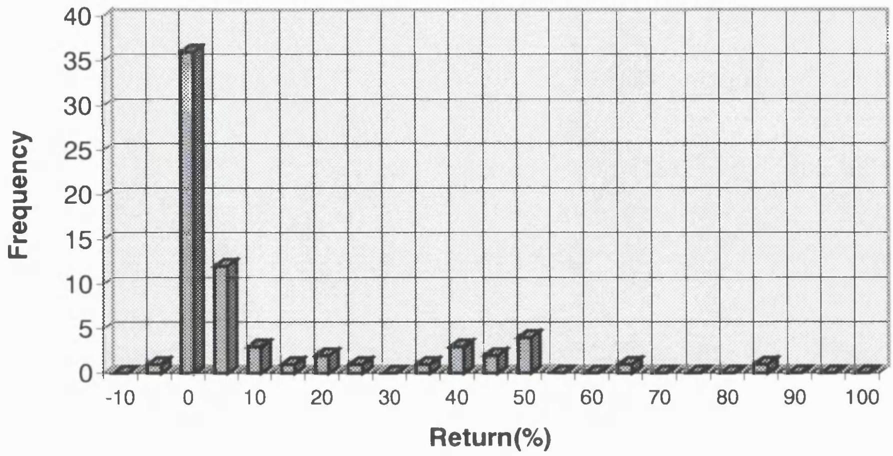
The 10 stocks selected at random by the system were Asda, BP, British Gas, the FTSE index, Cable & Wireless, GEC, GKN, HSBC, LandSec and Natwest. The performance of the system has been summarised in Table 5.4.

Table 5.4: Performance measures for a portfolio of 10 stocks

trading period (days)	1000
trading costs	2.0%
total trades	68
profitable trades	31
losing trades	37
% profitable	45.6
av. Return (profitable trades)	19.7%
av. Return (losing trades)	-4.0%
max. profit per trade	80.4%
max. loss per trade	-5.0%
mean return per trade	8.7%
variability of return	3.7%
mean trade length (days)	34
mean trades per year	16.1

It is not possible to construct a cumulative profit and loss curve that is meaningful without taking on a series of fund management issues. This will be discussed in greater depth in section 5.6. However, it is possible to construct graphs of distributions of returns, as these are very instructive as to the likely behaviour of the cumulative profit and loss curves, once the fund management issues are resolved.

Figure 5.4: Distribution of Returns



5.6 Discussion

The system appears to be in the market if the outlook is promising enough for a position not to be too expensive. This inference can be drawn from the fact that the system is usually in the market, but often takes small losses. One result of this degree of market exposure is that the system can take advantage of range breakouts and trends from the start, and that the UK equity markets tended to go up over the test period. If the market declines, the system escapes and waits until the market prognosis improves before re-entering.

A result of this behaviour is that the return distribution is very asymmetric - any individual profit is likely to be much larger than any individual loss (Figure 5.4), but the probability of making a loss is greater than the probability of entering a winning trade(Table 5.4). This is not seen as too much of a problem as the mean profits are much larger than the mean losses, and a single profit will pay for several losses. However, there are problems with this type of trading behaviour, where losses are more likely: the possibility of sustaining a string of losses is much greater, and these draw-downs are important for a number of reasons:

1. If the system loses too much in a single stretch, then it will simply be withdrawn from the market, even if it is profitable in the longer-term.

2. Draw-downs can be psychologically difficult, traumatic even, for the human who is running the system.

As has already been noted, it is not straightforward to compile a cumulative profit and loss curve for this system. It was anticipated that it might be possible to use Monte-Carlo simulations to reconstruct the dynamics of the P&L from the probability of successful trade outcome, mean win size and mean loss size. This has turned out not to be the case. From Table 5.4, it is clear that the mean profit is much larger than the mean loss. The information that this does not convey are the dynamics of the profit and loss curve - in reality, there will be periods of sustained draw-down, slack markets and explosive periods where huge gains and losses are made very quickly in volatile markets. The statistical overview of the system does not convey any information about this aspect of the behaviour of the system.

The reason that it is difficult to compile a cumulative profit and loss curve for this system is that at any one time, there are many sub-systems that *may* require financing. When the system enters a position, the securities must be paid for, either outright or in terms of using a credit line. Potentially, the system can place many concurrent trades - in this example the system is trading 10 markets, each with 10 sub-strategies. Consequently, it is theoretically possible for the system to have 100 positions spread across 10 markets at the same time. However, if each sub-strategy is simply assigned 1% of the total capital, much of the capital will remain unused and the profits made will be very small as a fraction of the risk capital. There are a number of possibilities for resolving this problem:

1. **Adaptive Position Sizing:** The essence of this idea is that each sub-strategy always opens the largest position that it can - so the first trade entered will be as large as possible. If another sub-strategy in the same market attempts to open another position, then ownership of the position is split equally between those systems that are currently in the market. However, this will increase the operating costs of the system dramatically.

2. **Use the overnight loan market:** It is rare that all 10 sub-strategies are in the market at the same time - indeed for some of the markets traded in this experiment, the system only entered one trade in the 1000 days of trading, and some of these

were only held for a single day before being stopped out. If each of the 100 sub-strategies is allocated 1% of the total capital, then much of this will remain unused. Noting this behavioural property of the system, something else can be done with the capital: place it in the overnight loan market, where it can be lent to a third party for a return slightly below than LIBOR[5]. The positions entered by the system are now highly deleveraged, and in the experiment here would produce a return of approximately 6% over the test period. The net result of this is that the system would make an approximate return of LIBOR+100, or 1% over LIBOR, which could be deemed satisfactory, if unstartling.

3. **Limit the number of positions:** This is the simplest option, which is simply to have a maximum number of positions (such as 3) for each market, and individual strategies can use one of these credit lines if it is unused. It is likely that these credit lines could be adaptively assigned to those markets that have the largest numbers of profitable, concurrent trades.

The most effective solution is likely to be a combination of all these: whenever capital is unused, it can be placed in the overnight market[6], and as it is extremely unlikely that all 100 individual sub-systems will need to be funded at the same time, the level of deleveraging can be reduced to increase the returns. If the maximum number of credit lines is set at (for instance) 25 instead of 100, then any profit or loss will be 4 times greater than if 100 credit lines are available from a single capital base. In this way, the total return over LIBOR can be increased.

It is interesting that the system trades frequently on some markets and extremely infrequently in others. However, the system's briefest foray into the market was a 1 day trade on the FTSE index, which was stopped out immediately. This is consistent with a view from standard economic theory that the information set of the FTSE index is too large for any simple processing of the price series to yield any additional information about the correct price of the index. Consequently, it is very difficult for

---

[5] LIBOR is the London Inter-Bank Offer Rate: The return offered for a 6 month loan.

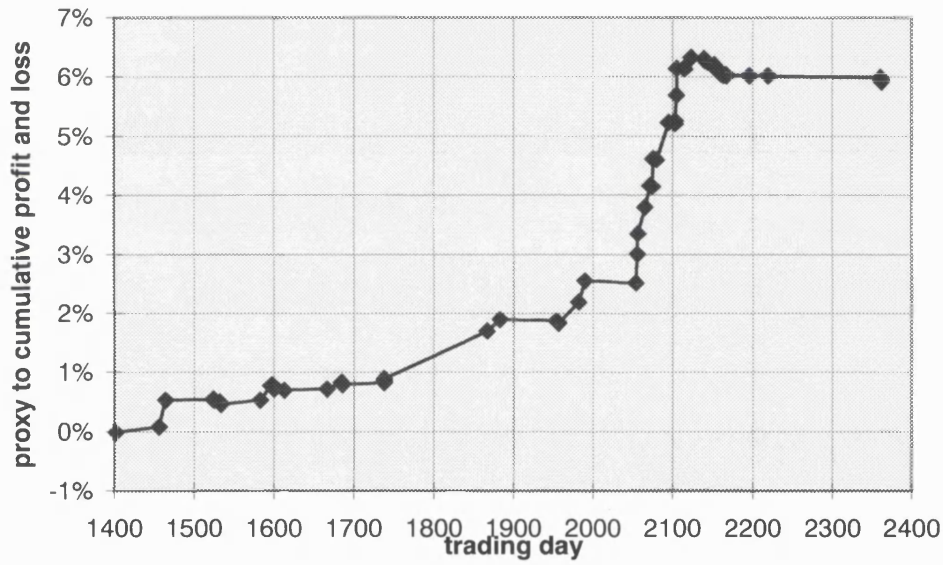
[6] In the travel industry, the profit margins on the holidays themselves are so slim that no profit is effectively made on the reselling of holidays. The profits come from the fact that customers pay for the holidays in advance, the travel company can (legitimately) delay payment until the following day, and hence profit can be made by lending in the overnight markets.

the system to trade the index, as the sub-systems must demonstrate that it can turn in a profit before the cash management system permits it to enter any trades.

Conversely, the system enters many trades on the companies that the index is calculated from. This is consistent with the theory that the companies' information set is sufficiently smaller that it is feasible for the system to enter speculative trades. It is important to recognise that there is no stock selection: each group of 10 strategies is assigned to a randomly chosen equity and the system attempts to make profits from it.

It is possible to compile a proxy to the profit and loss curve can be compiled by simply summing the returns when positions are closed. This is shown in Figure 5.5:

Figure 5.5: Proxy to cumulative profit and loss curve



This is a valuable exercise, because this can give information as to whether the system meets the original brief of performing consistently in the market over a protracted period. It is clear that the draw-downs that were feared from the examination of the ratio of winning trades have not materialised, but related effects are present in the P&L curve. The system fails to make profits in final 250 trading days. This is because the system is not able to form the profitable models of market behaviour it requires to trade over this period. This can be inferred as no positions are closed *at all* in this final period - positions are not even being stopped out. This means that none of the strategies the GA was exploring were working well enough for them to be financed, in all but one of the markets (LandSec). As a result of this observation, it cannot be legitimately claimed that the system *always* evolves along with the market, and that there are some market characteristics where the system cannot operate profitably. However, in these circumstances, the system does not lose large amounts of money, and most of the capital could be lent to third parties through the overnight loan markets.

There are several potential reasons why the machine has been unable to find profitable trading rules for the period between trading day 2123 and the end of the simulation.

1. The system does not have the technical indicators it requires to take the views of the market that it needs to trade profitably. This could be a factor as there are indicators that present very different information from the indicators that are supported. The momentum indicator is probably the most useful of the indicators that are not included. This measures the rate of movement of market price, which is information that the system cannot currently obtain.

2. The rule framework does not allow the system to express the trading rules that are effective during this period. This is closely linked to the previous point - the system simply cannot carry out the necessary processing on the available information for the system to make profitable rules.

3. The GA was unable to find the profitable rules that did exist in the rule space at the time. If the GA was unable to find the rules required, then it is likely that they occupy a relatively small fraction of the space. These solutions are then likely to be ill-conditioned or over-fit as they rely on precise parameter settings, and are consequently unlikely to work well for very long.

4. The rules that would be profitable in near future are not those that would have worked well in the recent past. Therefore the system would not allow them to trade if they were found in the first place. This point is closely related to the point above - if profitable strategies are found, then they are unstable and probably over fit.

5. As a result of these criticisms, it cannot truly be said that the system has met its brief: to always remain attuned to market conditions. It does work well most of the time, and the draw-downs are low, but there are periods of market behaviour where the system simply does not operate effectively. In the system's defence

however, little capital is ever lost, and in these periods of non-activity, the system could still be making slightly under LIBOR on the risk capital in the overnight markets.

5.7 Summary

The main details of this chapter are as follows:

- The problem was to attempt to construct a system that can survive changes in the market place.
- The data available was the daily closing prices of stocks in the FTSE 100 index, for a period between 7th January 1983 to 13th December 1995.
- The system uses crossover and mutation style operations on representations of trading strategies in order to remain attuned to market conditions.
- Although fund management issues remain unresolved, a proxy to the equity curve exhibits a high profit to draw-down ratio.
- The overnight loan market could be used to make returns on unused capital.
- The system does not attempt to trade the FTSE index. This is consistent with the financial theory that the information set of an index is to large for the index to be out-performed.
- The system works fairly well, and does not make substantial losses.

Chapter 6:

The Bull-Bear Trading Engine

This chapter investigates automated trading system induction. The data available is up to 20 years of daily open, high, low and closes, with some auxiliary data streams, for the European, Japanese and American Government Bond Futures markets.

6.1 Background

Governments nearly always spend more than they collect through various fund raising mechanisms such as taxes. This shortfall has to be made up somehow - and one method of raising capital now, is to issue a bond. A bond is like an IOU - it is a piece of paper that states that the bond-writer owes the bond-holder money. The investor buys the bond, thus transferring funds from his account to the government's. This entitles the bond-holder to a rate of interest on his *principal*, in return for the use of his money. After an agreed period, the bond matures and the principal is repaid.

Bonds exist in a wide range of maturities, from 3 months to 30 years, and indeed some bonds never mature. In general, the longer the maturity, the greater the rate of interest that will be offered - the investor usually has to be compensated for the extra time that their money is tied up in the bond. Similarly, the worse the credit rating of the bond-writer, the greater the required rate of interest. Large corporations also issue bonds, and interestingly, sometimes have better credit ratings than the governments of the countries where these firms are based. However, for a domestic market, the risk of a government defaulting on a bond is practically nil, because even if the government has no money to repay the principal, it can print money. Governments can be bad risks for foreign investors, as foreign investors will be subject to changes in the exchange rates. Any returns made on the bond could be dwarfed by losses incurred from a drop in the value of the currency of the bond-writing country.

Bonds are often referred to as fixed-income investments. This is because most bonds offer a series of fixed coupon payments. As a result of this, the price of the bond will vary as the prevailing interest rates change:

If the 1 year rate is 10%, a 1 year bond whose coupons are worth £10 will cost £100. If the 1 year rate then rises to 11%, the then price of the bond must drop to £90.91, as 11% of £90.91 is £10, the fixed coupon payment. Similarly, if the rates drop, then the prices of bonds rise.

Unsurprisingly, as interest rates vary unpredictably, there is a market for bond futures. Futures are contracts to exchange a specified commodity at a specific price at a specified date in the future. As the prices of bonds are connected to the interest rates, government bond futures are an important derivative product that allows investors and businesses to hedge their financial health against adverse changes in interest rates. Also, there is a healthy speculative activity that seeks to make profits from anticipating the future movement of the market. Bond futures are sometimes called interest rate futures.

Futures trading is done on margin. This means that to trade a future (not the bond itself) on a US Treasury bond worth $100,000 at maturity, only requires the trader to put forward some capital as a mark of good faith, that should the trader need to deliver a $100,000 Treasury bond, he has the financial reserves to do so. This margin requirement is usually between 0.25% and 2% of the price-at-maturity of the bond.

A consequence of this is that futures trading is very highly leveraged. With $450, a dealer can trade one futures contract on a thirteen week Eurodollar bond worth $100,000. At $25 per basis point (hundredth of a percentage point), the rate of such a bond only needs to move by 0.18% for the trader to double his money, or go bankrupt. For this reason, futures trading is *de-leveraged* to reduce the risk of ruin. Of course, this also reduces profit margins, but making less profit is preferable to almost certain bankruptcy. Typical de-leveraging would be to use 5-10% of the risk capital for margin requirements. A risk-averse institution might only use 1%.

6.2 Data

The data used in this project are daily data for a range of European, American and Japanese Government Bond Futures Markets of varying maturity. More specific information about each of these markets is shown in Table 6.1. There are two points to note about this table:

1. The minimum margin requirement increases with market maturity. The longer maturity markets are characterised by being more volatile than the shorter term instruments, and so the broking firm must insulate itself against the increased possibility of the client suddenly having a large out-of-the-money position.

2. The tick values are approximately the same. However, the tick sizes are different in different markets. There are 100 ticks to the point in the Japanese and continental European instruments, while these instruments in the UK and US have 32 ticks.

Table 6.1: Market information(Lehman Brothers Inc)

Market	Maturity	Country	Margin	Tick Value	Commission
Eurodollar	3 month	US	$450	$25.00	$9.68
Euromark	3 month	Germany	DM 800	DM 25.00	DM 11.28
Pibor	3 month	France	FFr 12,000	FFr125.00	FFr 50.00
US T-bills	3 month	US	$304	$25.00	$9.60
US T-note	2 year	US	$1,080	$15.625	$9.60
US T-note	5 year	US	$1,215	$15.625	$9.60
Bund	10 year	Germany	DM 3,500	DM 25.00	DM 11.28
Gilt	10 year	UK	£1,000	£15.625	£5.00
JGB	10 year	Japan	¥1,600,000	¥10000	¥8191.50
Matif	10 year	France	FFr 15,000	FFr 50.00	FFr 50.00
US T-note	10 year	US	$1,800	$31.25	$9.60
US T-bond	30 year	US	$2,700	$31.25	$9.60

For each day that the markets are open, the following information is available about that day's trading:

1. **Open:** The price of the first trade of the day.

2. **High:** The highest price achieved during the day
3. **Low:** The lowest price reached during the day.
4. **Close:** A proxy to the final price of the day.
5. **Open Interest:** This is the number of contracts outstanding - a high open interest indicates that large numbers of dealers have positions.
6. **Implied Volatility:** If dealers think that the future markets will be more volatile then this increases the price of options[1]. The option pricing equations depend on the market volatility over the life of the option, and these equations can then be rearranged to discover what anticipated volatility is implied by the option's price.
7. **Volume:** The number of transactions that have taken place that day. This number is a multiple of two because both the buyer and the seller report the trade.

Open interest, implied volatility and volume figures are not compiled until after the market closes, and so these figures cannot be published until the following trading day. As a result, the system can have access to today's open, high, low and close, and the previous days implied volatility, open interest and volume. These values can be fed into the machine and the trading signals generated for that day's trading. Some information vendors can supply these figures on an intraday basis, but this information was not available.[2]

There is some debate about whether one is allowed to use the closing price (and the day's high and low) in model induction and trading simulations, as it is only defined once the market has shut. As with many of the debates that surround work of this nature, the argument is between the economists and the technical analysts, and the problem with resolving these debates is that these two groups have very different

[1] Options are derivative products: they give the owner of the option the right, but not the obligation, to either buy or sell the underlying security at a pre-specified price in the future. A premium is paid to the writer of the option as they are obliged to supply the security at the agreed rate if the option is exercised. Options are more expensive if the markets are expected to be volatile, as there is a greater chance that the price will reach levels where the option can be exercised.

[2] Traders sometimes use unusual methods for obtaining intra-day information about the activity taking place in the trading pit. I have witnessed a request for the phone to be pointed at the pit so that the trader on the other end of the phone can attempt to infer the mood of the market from the volume, frequency and tone of the pit trader's shouts.

starting assumptions. In fact, there is so little intellectual common ground that it is often difficult to have any sort of meaningful debate.

The view taken here is that an order entered at-the-market in the last 90 seconds of trading will be near to the closing price. It will not be exact, but it will be a reasonable proxy to the closing price. If the economists are correct about the movement of the market being random, then ironically this negates the criticism that the system cannot enter a position at the closing price. If the market price is a random walk then in the final 90 seconds of trading that will occur after the trade was entered, the market will, on average, move with the trade as frequently as it moves against the trade. In addition, if the market is efficient then both the prices at the time of entry and the market close will be correct. For these two prices to be far apart would take the out-break of unexpected news in that final 90 seconds. This, by definition, is unexpected, and will increase the net present value of the trade by as much and as often as the net present value of the trade will be reduced. The net effect of this on the profitability of the system will be zero.

In addition, trading after the close of business is available in many of these Government Bond futures markets - after a delay of 15 minutes, the market reopens, the day's open, high, low and close are known, but the volumes are much lower and so the bid-ask spreads tend to be wider. 15 minutes is more than enough time for the system to react and for the signal to be conveyed to the broker.

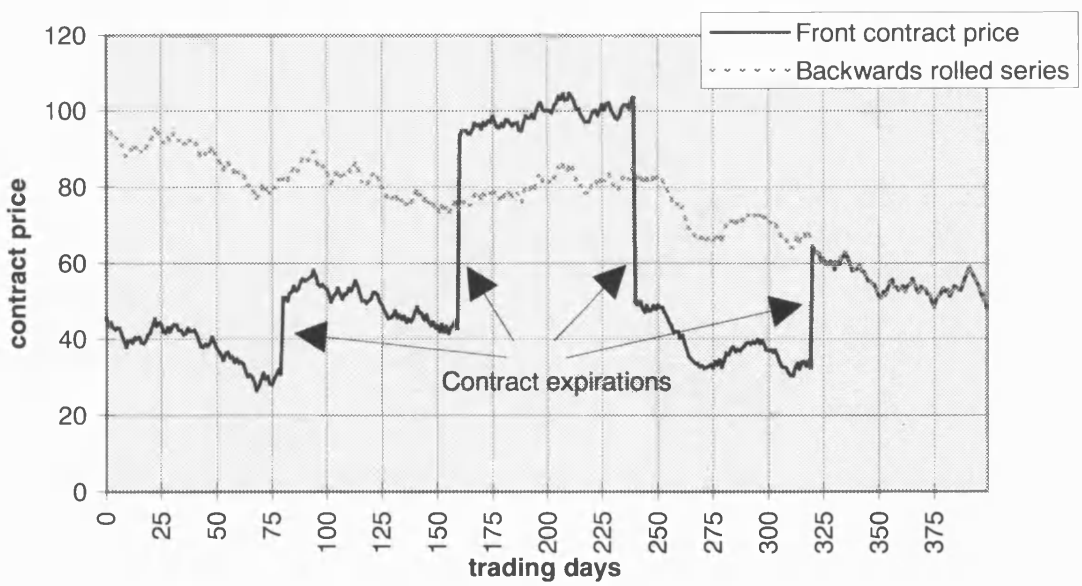
6.2.1 Data Preprocessing

Futures price data is not continuous. A futures contract is a contract to either buy or deliver the underlying security, such as gold, oil, pork bellies, or government bonds, at an agreed date. The delivery price of the underlying security changes as the future is traded, until the contract expires. At the close of the market on expiration day, those dealers who still have positions then have contract to either deliver or take delivery of the underlying commodity at the market price at expiration. Most traders will simply *reverse* out of their positions by selling(buying) the contracts they had previously bought(sold), to leave no net position.

Many different futures contracts will be traded in the commodity at the same time with different expiration dates. Futures usually expire every 3 months, in March, June, September and December. For instance, in May, June Wheat, September Wheat and December Wheat futures will all be trading simultaneously at different prices. The next contract to expire is termed the front contract, and the following ones are the second, third, fourth etc. There can be as many as 40 of these following contracts, but the further it is removed from the front contract, the less liquid the contract's trading will be as most trading is done in the front contract. For this reason, speculators usually trade in the front contracts, but if they wish to hold their position over an expiration, then the position must be moved into the new front contract. This is termed *rolling the position*. Rolling is done by selling the existing position and buying the same sized position in the second contract. If these two trades take place simultaneously then the net present value of the trade will be unaffected, even if the prices are very different.

For this reason, all these futures series have been "backwards-rolled" to turn the futures price into a single continuous series. This is shown in Figure 6.1. The daily price changes have been preserved in the synthetic series, but the 3 month fragments have been connected up so that the prices are the same on rollover day. In the trading simulation, the profit of a long trade can now be calculated by simply subtracting the buying price from the selling price, even if the trade would have spanned multiple contracts.

Figure 6.1: Backwards rolling of futures prices



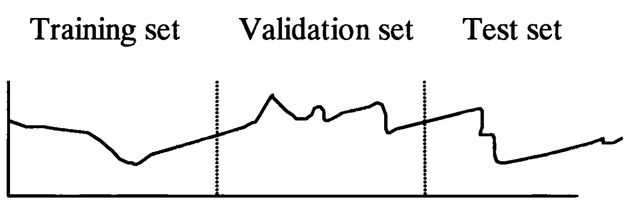
There are a few points worthy of note about adjusting futures prices in this way:

1. If positions are rolled into the next contract, the trading costs associated with rolling, such as bid-ask spread, commissions and slippage must be taken into account in the trading simulation.
2. Ratios or logarithms of the prices cannot be taken anymore as the absolute values of the past prices are no longer valid. In fact, for some very long series such as sugar, the price actually goes negative in the distant past, due to the cumulative effect of removing the discontinuities in the prices.
3. When preprocessing the data, the data is not in fact rolled on expiration day, because in the final days before the contract expires the contract becomes progressively more illiquid. Dealers will be trading in the new contract so that there is no danger of being caught with a net position when the contract expires. For this reason, the data is rolled a week or so before expiration day, on the first trading day of the expiration month, to mimic how positions are rolled prior to expiration.

6.2.2 Data Partitioning

The data is partitioned into three approximately equally sized sets, for training (finding rules), validation (checking rules) and testing (assessing the system's likely future performance). The test set is out-of-sample - no information is available from this period when the rules are being induced. In this way, assessing the behaviour of the system on out-sample data can be an accurate test of how the system would perform in the future.

Figure 6.2: Data Partitioning Schematic



It is important that when the system finds rules on one market and is tested on another, that the start date for the testing data is after the end date for the training and validation data. For this reason, the boundaries between the sets have specified dates so that no "data snooping" can take place. The important dates are shown in Table 6.2.

The first market to be examined in the course of the investigation was the Eurodollar 3 month. Data for this market was (easily) available from 9th December 1981 to the present day, and it was partitioned into three approximately equally sized sets.

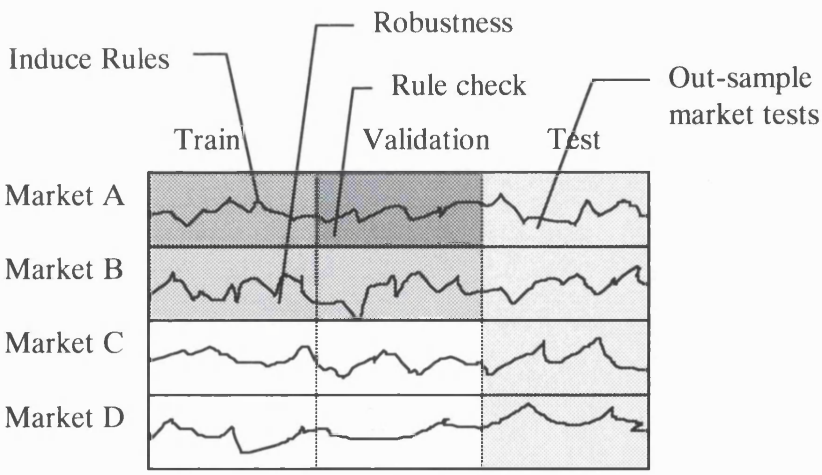
Table 6.2: Partitioning Dates

Data	Start Date	End Date
Training set	9-Dec-81	9-Sep-86
Validation set	10-Sep-86	7-Jun-91
Test set	10-Jun-91	3-Jan-96

6.2.3 Robustness Testing

Once the system has found a set of rules, they are used on a validation set to assess whether the rules are simply fitting noise or whether they have caught something fundamental to the way markets behave. Those rules that survive this test are then subjected to a larger, second validation set that is compiled from a different market. This is shown in Figure 6.3: Market A is the original source of the rules, Market B is the second validation market, and Markets C and D are only used for out-sample testing.

Figure 6.3: Robustness Tests



The training and validation sets are contiguous sections of market history. This second validation set runs from the start of the training set to the end of the (first) validation set. Rules must perform acceptably over all sections of market behaviour. The remaining rules are then a combination of extremely over-fit rules and rules that *do* have information content.

6.3 Genetic Algorithm Rule Induction

This section describes the components and the operation of the genetic algorithm rule induction system. After examining the work with Sabre Fund Management and the Continually Adaptive Trading Engine, it was decided that the process of finding trading rules with genetic algorithms is one of the approaches most likely to work satisfactorily. This can be thought of in terms of building implicit partial models - or *finding pockets of predictability*.

The benefit of this type of approach is that it makes very few assumptions about the nature of financial markets. In fact, the default assumption is that the *markets are random*, and nothing can be done to consistently generate profits. The genetic algorithm then begins to search for holes in this assumption, and tries to locate tradable pockets of determinism. One of the greatest strengths of this approach is that no attempt is made to build a complete model of the market. Those times when the action of the market is simply beyond our comprehension or modelling capability, are simply ignored.

It is worth looking at the entire GA system in overview before examining the components in great detail, in order to understand how the components fit together. There are a number of main components of the system: the data, the genetic algorithm, the rule framework and the trading system. The GA operates on the framework and the data to find trading rules that will then contain our implicit model of the market. The GA is executed repeatedly to build up a portfolio of diverse rules that are derived from different peaks in search-space. Six different types of rules are induced - rules for buying and selling, in upwards, downwards or sideways markets. This simple market characterisation is carried out to limit the range of market conditions that the sub-model need address. It is not the goal to find the highest point in the space - a mesh is wanted of overlapping sub-models that each work well within their subset of market behaviour. For this reason, rules are also induced on the validation set and tested on the training set to find as varied a set of rules as possible. The rules are subjected to tests to ensure that they are reasonable, and once a 'final' set of rules has been compiled, they are used in conjunction with the trading system for out-sample testing.

6.3.1 Long and Short Rules

The rules that the GA is being used to find are all in the form:

if <*clause1*> **AND** <*clause2*> **then** enter position

This can only recommend that a trade be entered, whereas this system is designed for deployment in the futures markets, where one can *go short*. This is sometimes called "selling the market" where the trader will profit from a falling market. In order to generate rules that can recommend both long and short trades, two distinct sets of rules are induced: a set for going long and a set for going short.

6.3.2 Market Classification with the Adaptive Average

The type of rules needed to generate sell signals after a large price increase will be very different to rules that simply go with a downward trend. If the genetic algorithm tries to find rules for both these market behaviours at the same time, then it will find itself attempting to climb two separate peaks in the fitness landscape. To bring some cohesion to the nature of the fitness landscape, each rule is designed to generate either buy or sell signals in one of a limited class of longer-term market behaviours. This simple characterisation was done with an *adaptive average*. Multi-modal optimisation techniques exist[Deb93], but it was judged that as the work was somewhat exploratory, keeping the system simple to start with was a good idea.

The adaptive average is like a moving average - it attempts to remove some of the noise from the movement of the market price by taking the average price over the most recent n days. It was noted in Section 5.3.1 that long moving averages are sometimes too slow to react, but short moving averages can be too susceptible to random fluctuations. In order to combat this, the adaptive average extends in times of volatility to smooth the signal, while in calm markets, it contracts so that it can react quickly to changes in the market price.

The adaptive average was used to help assess what the nature of the current markets are - it loosely categorises the market behaviour into upwards, sideways and downwards markets. This categorisation is done so as to divide the in-sample data as equally as possible into the three categories. This is achieved by looking at the gradient of the adaptive. If it lies within a certain range of values then the market is deemed to be sideways, if above, then trending up, and if below, then trending down.

Rules are then specifically induced for going either long or short, in one of upwards, sideways or downwards markets.

6.3.3 Technical Indicators

The GA must be supplied with a view of the market if it is to make observations as to what are the precursors to an ensuing, tradable market behaviour. This is done by giving the GA a range of indicators that it can interrogate for current and past values of the market behaviour. For certain indicators, it can also look at moving averages of past values to smooth the data. The indicators supported are:

1. **Bollinger Bands:** The upper and lower Bollinger bands are calculated by taking a moving average of the price series and adding or subtracting the standard deviation of the price series to give a pair of bands that surround the moving average.

2. **Chande Momentum:** a combination of momentum and relative strength(described below): It is defined as (up days - down days)/(up days + down days).

3. **%d stochastic:** a three day moving average of %k (described below).

4. **Implied Volatility:** an auxiliary series that gives the anticipated market volatility as implied from option pricing.

5. **%k stochastic:** the position of the current price expressed as a fraction of the previous n day's range: (current price - range low)/(range high - range low).

6. **Momentum:** the current price minus the price a specified number of days ago. This gives a "velocity" of price movement.

7. **Moving Averages:** the arithmetic average of the last n days open, high, low or close.

8. **Open Interest:** an auxiliary series that gives the number of contracts outstanding - a high open interest indicates that large numbers of dealers have positions.

9. **Price:** The open, high, low and close prices of the security are available.

10. **Relative Strength:** the fraction of the last n days that the price has risen. If this reaches 100% then the price has risen for every day of the sample.

11. **Ratio of Body to Daily Range:** a measure of how much direction the market has - the body is the difference between the open and close, while the daily range is the difference between the high and low of the day.

12. **Daily Range:** the high-low range of the day's trading.

13. **New Range:** a boolean indicator which is true when the closing price breaks higher than the highest or lower than the lowest price of the last n days.

14. **Volume:** another auxiliary series that is the number of transactions that day. Note that the volume must be a multiple of two, as each trade is registered by both market participants, and that yesterday's volume is the most recently available figure as it is not published until several hours after the market has shut.

15. **Standard Deviation:** The standard deviation of the change in closing prices for a specified period.

6.3.4 Operators

Once the system can take measurements of the state of the market with the indicators described in the previous section, comparative operators are needed to form the clauses of the rules. These are shown in Table 6.3

Table 6.3: Rule Operators

Operator	**Meaning**
=	equals
>, <	greater/less than
!	penetrates
v, ^	penetrates to below/above
}, {	moving in same/different direction
], [	moving faster/slower
a, d	accelerating faster/slower

Rules are of the form;

if (*correct market*) **AND** (*clause*$_1$) **AND** (*clause*$_2$) **then** <go long/short>

and they activate when both of their clauses are true and the adaptive average has characterised the market as being appropriate for the rule. Each clause is a relationship between the dynamics of a pair of indicators. For example:

(open interest is accelerating faster than the volatility did three days ago)

6.3.5 Rule Evaluation

It is straightforward to discover whether a rule is effective or not. The evaluation procedure is to step through the market history on a daily basis, entering positions every time the rule's conditions are met. After a position is entered, it is held until it is either stopped out or a certain period of time elapses. Upon trade closure, a note is made of whether a profit was made or not, and the evaluation process returns to stepping through the market history. After one complete pass through the data, it is known how many trades were entered, and what fraction of these were profitable. This data is then used for calculating the fitness of the rule. For reasons that are discussed in Section 6.4, sizes of the profits and losses are not used in assessing a rule's fitness.

This "certain period of time" that the trade is held for is specific to each rule. The GA is executed repeatedly, and with each execution, a time-frame is chosen at random for the maximum holding time of the trade. This is done for the following reasons:

1. The GA can build up a portfolio of rules that come from completely different parts of the search-space.
2. This allows the GA to implicitly select the best time-periods to assess rules over, which in turn allows the GA to find the best rules.
3. It means that distributions can be constructed of the time periods of most successful rules.

Although even the longest trade holding periods are still fairly short at 10 days, it is found that the longer term rules tend to be slightly more reliable in the later testing procedures. Extending the maximum trade time beyond 10 days has a punitive effect on the speed that the GA operates at, while at the same time the extra benefit of extending the time-frame decreases.

The rules are originally found using a simulation of trading that does not include slippage, commissions or bid-ask spread. The reason for this is to try to make it as easy as possible for the GA to iteratively improve rules. Once rules have been found that meet some standard without commissions or slippage, they are tested again, this time including the costs of trading to assess their impact on the rule's performance. If this impact is excessive, then the rule is culled.

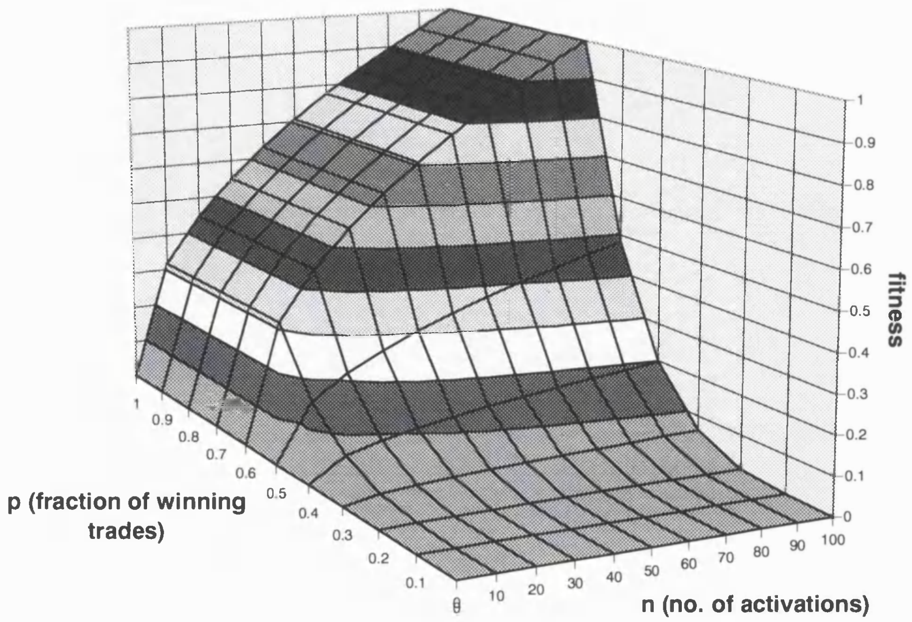
6.4 Fitness Evaluation

The fitness function is as follows:

$$f(p,n) = \min(1, 7.71*p^4) * \sqrt{n} / 10$$

where p is the fraction of winning trades to all the trades the rule enters (percentage of winning trades), and n is the number of times the rule fires. It is easiest to explain the rationale behind this definition if the fitness landscape is plotted. See Figure 6.4.

Figure 6.4: Fitness Landscape



In the early stages of the GA when mean fitnesses are low, selection pressure is applied preferentially to p, the fraction of winning trades. This is reflected in the landscape in that the steepest slope in the foreground of the figure is to be found in heading directly up the graph. In later stages of the run when solutions have been found that offer a sufficient percentage of winning trades (beyond 60%), the only way the GA can improve on the rules is to find rules that fire more often without compromising the rules' performance. It was judged that a rule that was more than 60% effective was likely to be over-fit and so be unlikely to operate well in out-of-sample situations. Through placing a limit on the percentage of winning trades, the GA can be prevented from deliberately noise-fitting.

The returns that each rule makes are ignored when the rules' fitness is being calculated. The reason for this is that even though high deviation events occur in the market more frequently than predicted by financial theory, it is extremely difficult to profit from them in a systematic manner:

1. If the return-per-trade is a factor in assessing the fitness of a trading rule, then the GA will extremely rapidly home in on the high deviation events and create over-fit trading rules that coincidentally pick out one or two of these occasions. This will lead to rules with no information content. The triggers for these events are often news-based, and so they are intrinsically unpredictable.
2. Some of these high deviation events, such as stock market crashes are un-simulatable, as there is no way that a accurate profit and loss calculation can be done. The assumption that an order will be filled near the market price will no longer hold in these situations.
3. Events such as Black Wednesday, when the Chancellor of the Exchequer withdrew from the Exchange Rate Mechanism and set new interest rates several times in one day cannot be sensibly modelled with daily data.
4. These events are anomalous to every-day market behaviour.

6.5 Genetic Algorithm Execution

The GA engine used to evolve the symbolic rules was similar to that used in Chapter Four. After some preliminary experiments had been carried out, a GA was used that had the following properties:

1. The GA was executed repeatedly, each time with some randomly chosen parameters - the time-horizon for trade evaluation and the exact nature of the rules sought during this individual run of the GA (long or short rules, in one of upwards,

downwards or sideways markets). In addition, rules were induced either from the "training" or "test" set and then tested on the other.

2. A simple GA was used with ranked tournament selection, with a randomly chosen starting population.
3. A population of 20 was used for 10 generations. This is very low number of iterations and due to the fact that the population would converge extremely rapidly.
4. A high level of mutation (0.1 probability per cycle per individual) was used.
5. After 10 generations, the population had invariably neared convergence. To let the GA act for longer on the rules would simply result in over-fit rules that would fail to perform out-of-sample.
6. All rules were then tested on disjoint data. Those that failed to operate on both the training and test sets were eliminated.
7. The GA was then re-executed, with a new random set of parameters for selecting rule type, trade direction (long or short), time-horizon, and data source. Repeatedly selecting parameters randomly in this way is a simple method for letting the GA implicitly select the most reliable environment for inducing rules.

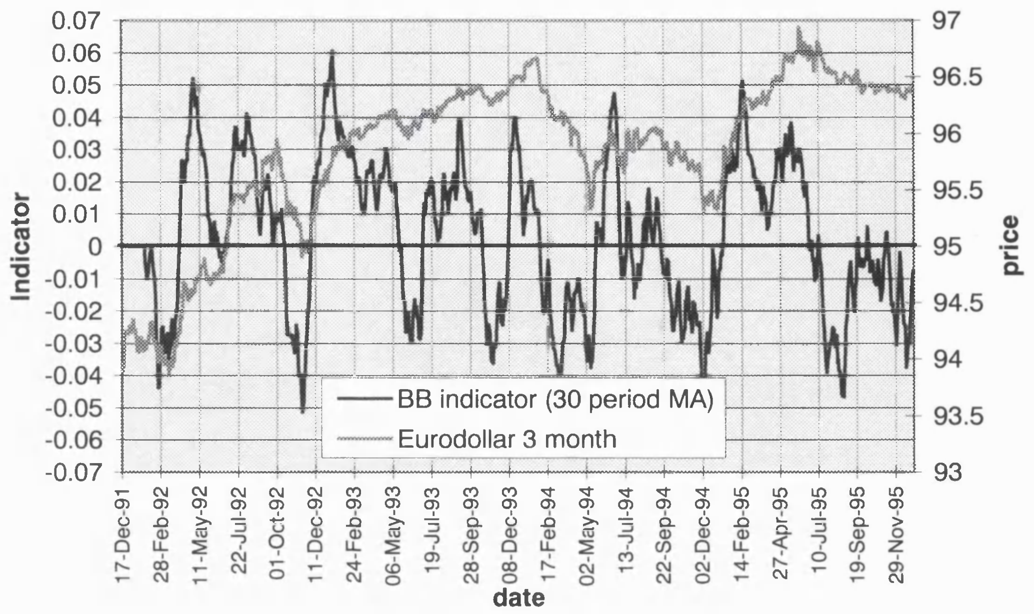
6.6 Trading the System

The system has a portfolio of rules for buying and selling in a range of longer term market characters. To trade from these rules, the individual behaviours of the rules need to be combined in a meaningful way so that a single trading signal is generated. This is done by looking at whether, on balance, the rules are bullish or bearish. An indicator can be created:

BB(t) = (no. of long rules active at time t / total number of long rules) -
(no. of short rules active at time t / total number of short rules)

The BB indicator is charted along with the Eurodollar price history that gave rise to in Figure 6.5.

Figure 6.5: 30 day BB indicator on Eurodollar 3 month (Eurodollar rules)



If the BB indicator is positive, the rules are bullish and indicating that the market is likely to rise. The system buys and holds the position until sufficient selling rules are active for the BB to turn negative. The system liquidates the existing position and enters a short trade. This is termed stopping and reversing. A clear consequence of this is that the system will always be in the market, and a consequence of this is that the marked to market[3] cumulative profit and loss curve will have the same volatility as the market.

A long-short consensus indicator of this type will be fairly noisy, and so moving averages are used to smooth the signal. It is important to deflect criticism that a well chosen moving average will do the trading in the first place. To counter this argument, several different moving averages are used and the results are combined. The periods of the moving averages are chosen at 15, 30 and 50 days. It is important to note that these figures have been chosen arbitrarily, and have not been the basis of any optimisation. Several separate moving averages are used to try to capture the information content of the rules and to minimise the impact/effect of the actual values chosen.

[3] Marking to market is the process of calculating the net-present-value of an instrument using the closing price as a proxy for the instrument's value.

Each of these moving average periods is traded separately: if the BB indicator changes rapidly from Bearish to Bullish for instance, the 15 period average will reverse first, then the 30 period average and finally the 50 day average will reverse. The system must have either a long or short position in each of these time-frames. A consequence of this is that there are effectively four different positions that the system can have at any one time, depending on the (anticipated) market behaviour (Table 6.4):

Table 6.4: Portfolio orientation

Indicator	Commitment	Net Position
Long on 3 time-frames	Full: Bullish	3 long contracts
Long on 2 time-frames	Partial: Bullish	1 long contract
Long on 1 time-frame	Partial: Bearish	1 short contract
Long on 0 time-frames	Full: Bearish	3 short contracts

No stop losses are used.

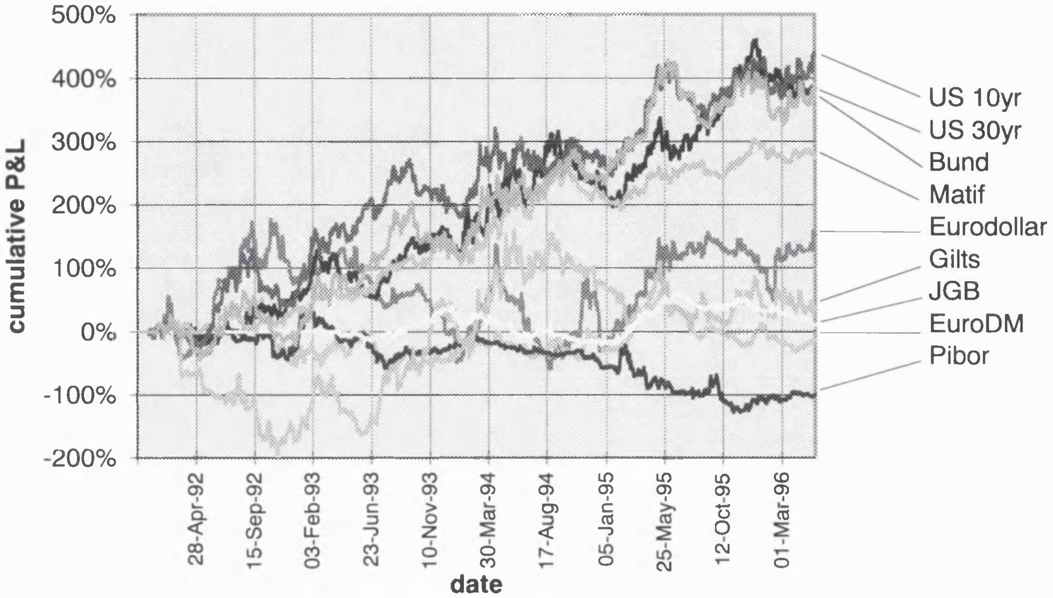
6.7 Results

The system was set on a range of European, Japanese and American Government Bond futures markets. Each of these will have specific tick values and margin requirements (see Table 6.1). All the cumulative profit and loss curves (P&L's) presented here have been converted to the return on risk capital. As noted in Section 6.1, futures trading requires deleveraging and so all markets are traded using 10% of the risk capital for meeting margin requirements.

The growth of risk capital for each market is charted in Figure 6.6.

Figure 6.6: Cumulative Profit and Loss on European, Japanese and US markets:

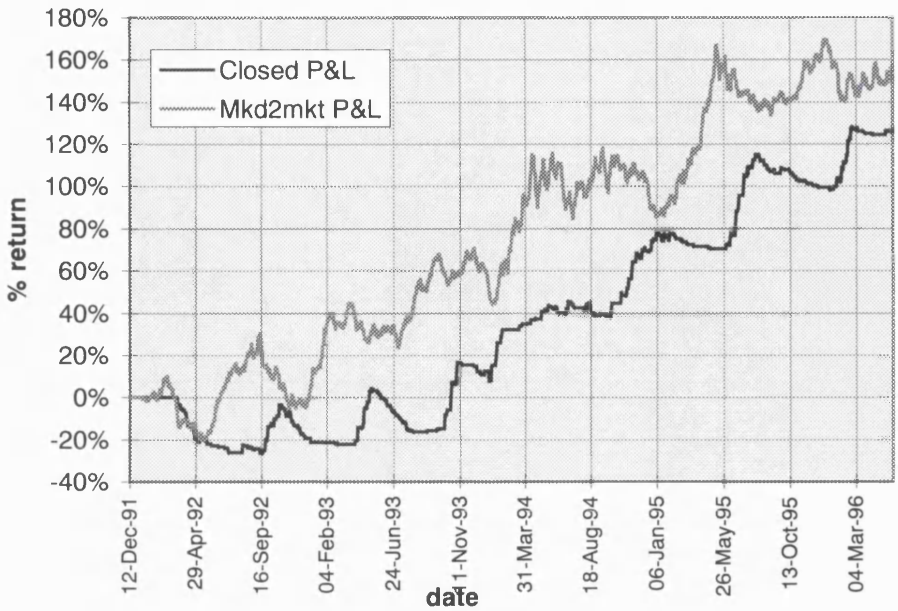
Eurodollar rules 3 month rules



It is clear that the trader would go bust trading either Gilts or Pibor on their own, as the returns for these series drop to below -100% (i.e. all risk capital has been lost). However, these individual P&L curves *can* be combined on an equal cash basis, to show the behaviour of a global portfolio, because risk capital can be transferred from a market that is being traded successfully to one that has received a margin call. The resulting equity curve is shown in Figure 6.7.

The closed and marked-to-market cumulative profit and loss curves have both been presented, quite simply because they frequently separate and have different characteristics. The closed P&L is only updated on trade closure, while the marked to market P&L takes into account the value or cost of any trades that are currently open. The marked to market P&L is a proxy to the financial state of the trader if all positions were liquidated. For instance, on 4th February 1993, the closed P&L shows that the dealer has lost 20% of his risk capital, while if all his positions were closed, he would actually have made 40%.

Figure 6.7: Multi-market Cumulative Profit and Loss: Eurodollar 3 month rules



6.8 Maturity

Figure 6.6 hints that there may be some dependence on maturity - simply from eye-balling the graphs it can be seen that the order of returns, from best to worst runs US 10 year, US 30 year, Bund(10 year), Matif(10 year), Eurodollar(3 month), Gilts(10 year), JGB(10 year), EuroDeutscheMark (3 month), Pibor (3 month). Interestingly, for any specific maturity, this is approximately the order that would be obtained by ranking the markets by liquidity (high to low).

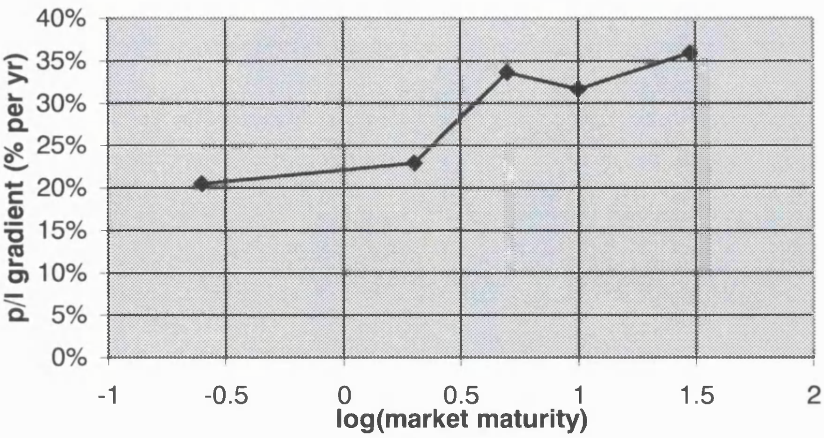
This observation of the results of the previous experiment will be clouded by various factors that are specific to the countries involved - for instance, it is not obvious that returns can be made in the French 10 year market from non-linear observation and analysis of a country-less 3 month instrument. To test this in a more thorough manner, this test is repeated but over a homogeneous set of markets. The US Treasury offers a range of certificates of varying maturities (3 month, 1 year, 2 year, 5 year, 10 year and 30 year). Any country specific factors will now not influence the resultant equity curves.

The previous experiment was conducted with rules from a 3 month instrument that is not fundamentally based in the US. Eurodollar issues are bonds that are denominated in dollars but that originate in other countries (originally Europe - hence the name). So a new set of rules must be induced. The US Bond was chosen in order to isolate this effect if it is present at all, and to determine whether this is a peculiarity of the short maturity markets or not. Other than the choice of source data (the US Bond), the experimental procedure was identical to that of the previous experiments, and is described in Section 6.5. The cumulative profit and loss curves for each of the US Treasury markets is presented in Figure 6.8.

Figure 6.8: Performance comparison across US Treasury Maturities: US Bond rules

A least-squares-fit regression line with the intercept constrained to zero is a reasonable means of characterising the rate of equity growth for each of the maturities. Other methods, such as endpoint detrending are not so appropriate in this context as they place disproportionate importance on the final return made at the end of the test period. A regression line will model closer the intra-test behaviour of the equity curves. A graph can then be plotted of the regression gradient (mean rate of equity accrual) against the log of the market maturity. See Figure 6.9.

Figure 6.9: P/L gradient vs. Market Maturity: US Treasury Futures

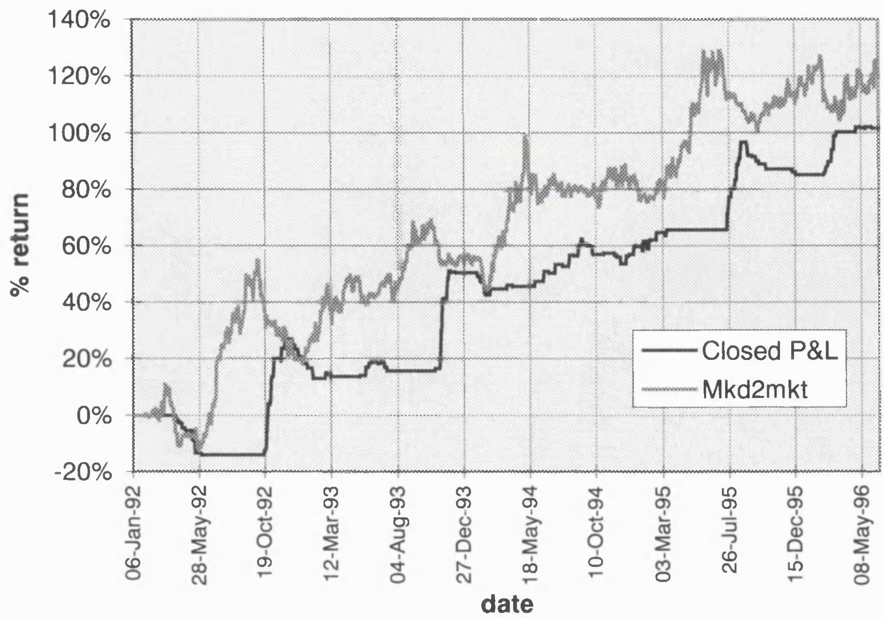


There appears to be some dependence here, but the sample size is low with only 5 data points. One solution to this problem would be to run the experiment on other maturities. Unfortunately, the data for no other markets was available.

In order to try to assess the chances of this result occurring by chance, Monte-Carlo simulations were used to find what fraction of random results produce a result with the same characteristics. Although insufficient data exists for a model of any complexity to be hypothesised from Figure 6.9, a regression line can be drawn: the gradient of this is 0.793. If p/l gradients for each of the maturities are selected at random from the range $0\% \rightarrow 40\%$, then a distribution of gradients can be constructed, and from this a proxy to the confidence level of the actual result determined. The Monte Carlo simulation was executed 16384 times, a distribution of gradients constructed, and from this a confidence level of 86% can be established for the regression gradient of Figure 6.9 of 0.793.

For comparison with the first experiment on a range of European, Japanese and American markets, the P&Ls from trading the Bond rules on the US Treasury markets shown in Figure 6.8 can be combined on an equal cash basis as before to yield an overall equity curve. See Figure 6.10.

Figure 6.10: Cumulative Profit and Loss for US Treasuries: US Bond rules



6.9 Commodities

The system was originally designed to operate on Government bond futures, but it has also been applied to the commodity futures markets of copper and gold. It is not sensible to attempt to trade the commodities markets with models derived from the Government Bond futures markets, so new sets of rules were found from the copper and gold market histories. Indeed, one of the motives for looking at the system's performance on gold is that the gold market behaves quite unlike any other market, whether one cares to look at bond futures, equities, currencies or money markets. For instance, in times of crisis, there is a very real tendency for there to be a "flight to quality" by investment institutions: no matter what happens to the financial markets, a block of gold will always be valuable simply because it is a precious metal, and unlike many other commodities, it will not degrade over time. Gold is seen as a symbolic safe haven of capital. The other major use investment institutions have for gold is as a hedge against inflation. Providing the world's gold mines do not flood the gold market, it is likely to retain its value in real terms quite simply because it is technology-free, non-degradable, rare, tangible asset.

The same pre-processing steps were carried out on the data: the futures prices were backwards rolled in the same way and the data was partitioned into three approximately equal blocks. Comparable numbers of rules were induced as in the Government bond futures experiments. The system was deleveraged to 10% of the risk capital as before.

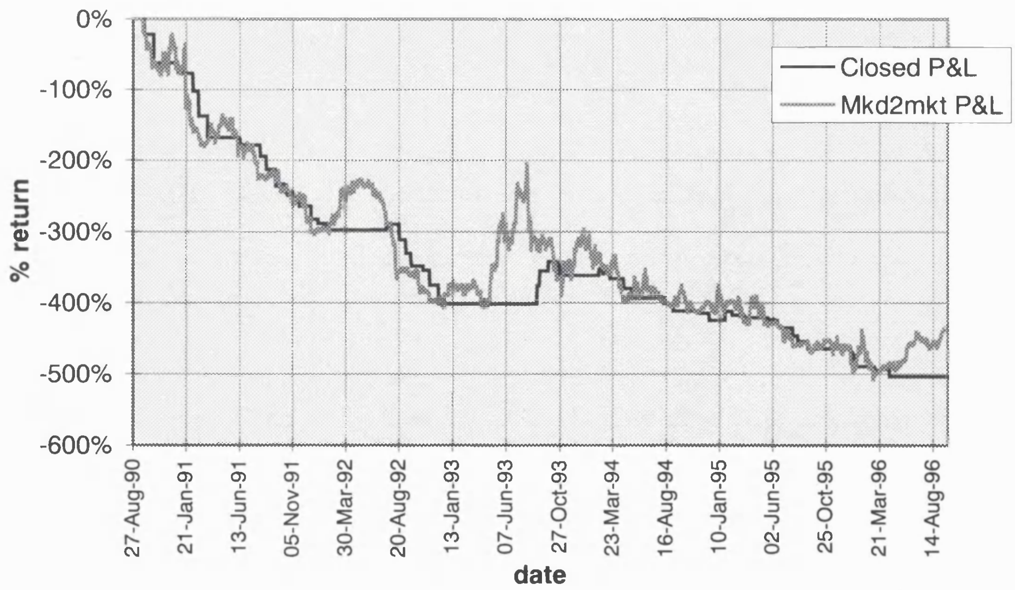
6.9.1 Copper

The GA was unable to find any rules that performed satisfactorily on the test data. Consequently, as the system had been unable to form any sort of model, there was no system to test on out-sample data.

6.9.2 Gold

Unlike for the Copper futures market, the GA was able to find models that appeared stable over the in-sample test data. The cumulative profit and loss curve is constructed as before. However it is clear from the equity curve that something has gone drastically wrong, and that the system has reliably failed to build any model that is appropriate.

Figure 6.11: Cumulative Profit and Loss: Gold



There are a number of points that are worthy of discussion regarding this experiment with commodities:

1. It is clear that neither of the commodities has even come close to operating successfully on out-of-sample data.

2. The commodities markets tend to be significantly more volatile than the interest rate (Government bond futures) markets.

3. Many commodities, such as wheat, have a seasonal nature. One would not expect this to be the case with metals to the same degree, but many practitioners routinely use time-of-year arguments to explain certain happenings in their belief system about markets[Zwei86]. A typical example might be "The market is often depressed in August because people are on holiday" The truth content of this statement is debatable, but that is not the point. Traders think that these factors affect the market, and so they affect the market.

4. Certain *a priori* statements can be made about the behaviour of the system from an analysis of the content of the rules - namely that the system is trend following in nature. This is discussed in greater depth in Section 6.11.1. At the time of writing, the gold market was notorious for having range-traded (had no trend) for 3 years, after undergoing a structural change that brought many years of downward trending to an end. This can only help explain the later half of the system's non-performance on gold, as clearly some other factor(s) must have been responsible for the first half of the test data.

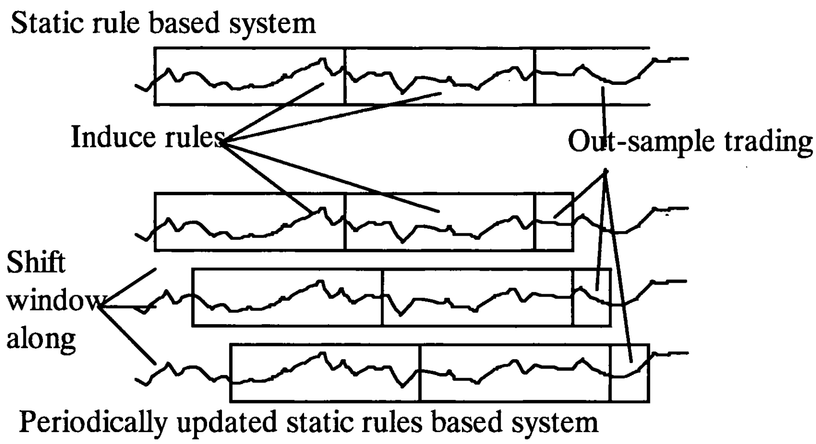
5. It could be suggested that the Gold rules could simply be traded "the other way": buy when the system says sell and vice-versa. At first glance, this appears to be a possibility, but with a little more thought it is clear that this is unacceptable. The rules that have generated this equity curve have been profitable over a 10 year period, and to attempt to trade them in reverse is to accept that over the previous 10 year period, the system would have lost money. The equity curve is dramatic in that very few trades ever make a profit when they are closed, but it is difficult to justify the decision to trade the system the other way - presumably, it is not out of the question that one might decide to reverse the signals once more when the system doesn't appear to be working anymore. At this stage, the dealer is then trading the trading system, *but the trader will only reverse the interpretation of the signals after sustaining a loss*. Therefore, the trader commits himself to strategy of taking strings of losses. The psychology of trading a system is an extremely important part of the trading process which is often ignored during system development.

6.10 Stationarity

Instead of using the rules to generate speculative trading signals, they can be used to characterise the markets that they were induced on. In turn, the system becomes a tool for examining structural changes in markets[4]. Rules that are induced from a specific section of market history are effectively a description of trading strategies that would have worked over this period. For instance, an extended period of non-trending low volatility will yield very different rules to those induced from the crazed final stages of a bull market.

The rules induced from two different sections of market history can be compared to assess how the nature of the financial markets has changed in this time. This comparison can be carried out in a number of ways, from looking at the composition of the rules to actually trading the rules on a reference section of market history. A third approach has been taken here that provides the means of examining the evolution of markets, and looking at the "shelf-life" of the rules that the GA finds. This methodology can be posed in terms of comparing the out-sample trading performance of a static set of rules from the start of the data with that of a set of rules that is being periodically updated. This is shown in Figure 6.12.

Figure 6.12: Stationarity Methodology



Rules are induced using 10 years of market data. The first set of rules to be induced are derived from the earliest data available. These rules are then traded on the out-

[4] Stationarity is this context is taken to mean that the processes that give rise to a time-series are unchanging. It is *not* the statistical notion of no systematic change in the mean or variance of a time-series once strictly periodic cycles have been removed.

sample market history for a period of one year. At this point, the 10 year wide 'window' is moved along by one year and a new set of rules found from this slightly different section of market history. The overlap is significant, but any differences that exist between the original and the new rules must be derived from the effect of removing the oldest part of the history and adding 1 year's more recent data. Then, both the original rules and the new rules are traded on the following year's history, and the P&Ls compared. This is an efficient scheme in terms of data-usage: the trading comparison clearly only ever takes place on out-sample data, which is then re-used for inducing the next set of models.

This approach enables the monitoring of how the original rules compare with the most up-to-date rules, which could, *a priori*, be expected to be better attuned to current market conditions. The static and updated P&Ls are shown in Figure 6.13.

Figure 6.13: Divergence of static and updated rules: US Bond, US Bond rules

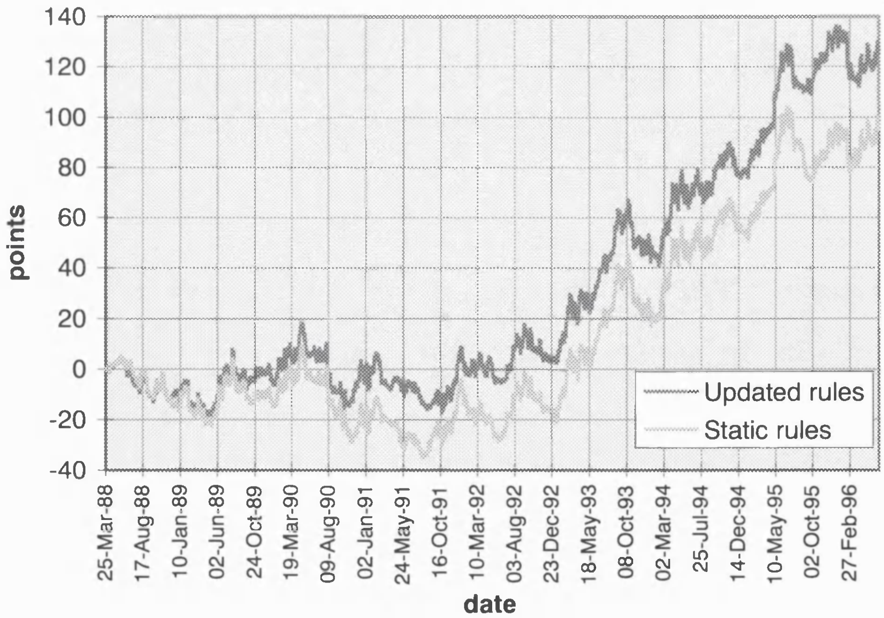
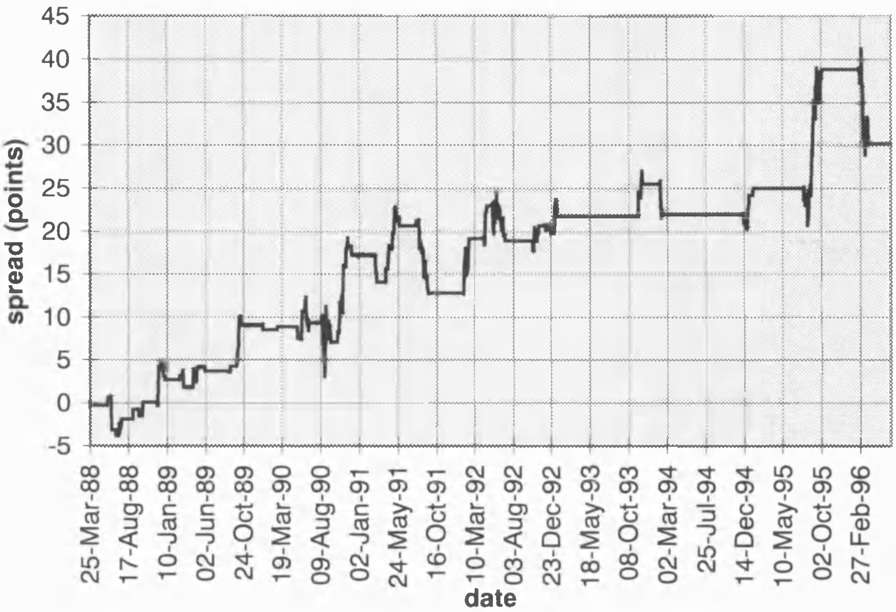


Figure 6.14: Divergence of P&Ls



The updated rule P&L can be subtracted from the static rules P&L to yield a graph of the spread: see Figure 6.14. The progressive widening of this spread indicates that the static rules are becoming less and less effective. The horizontal sections of the graph are not where the markets are not changing: they are when the system is not closing trades so no differential is evident between the performance of the two sets of rules, even though the subtraction is carried out with the marked-to-market P&Ls.

6.10.1 Comment

This work on stationarity is very interesting. Economists accept the existence and validity of some of the results from chaos theory, namely that the market exhibits memory effects and persistence. However, the economists maintain that these effects are all untradably small in size - any profits that would be made from this approach would be consumed by trading commissions, bid-ask spread and execution slippage.

What this tool offers is a method of monitoring how the market is changing, after the effects of slippage have been removed. Each rule that is produced has demonstrated that it is profitable *after* trading costs have been taken into account. As the market changes, so will the exact nature of the rules that are induced. Market changes that are untradable will be ignored, as rules that attempt to exploit untradable patterns will not survive the validation process. Therefore, only rules based around tradable processes will remain, and the evolution of these can be monitored by their differential trading performance on out-sample data.

6.11 Discussion

Much of this discussion will revolve around a defence of the work against an attack by a hypothetical economist, as this is a convenient forum for presenting criticisms and examining the results and consequent inferences. In addition, this will allow the development of some earlier points that were only mentioned in passing, such as why an analysis of the rules leads to the *a priori* statement that the system follows trends and consequently cannot trade ranging markets effectively.

6.11.1 Information Content of the Rules

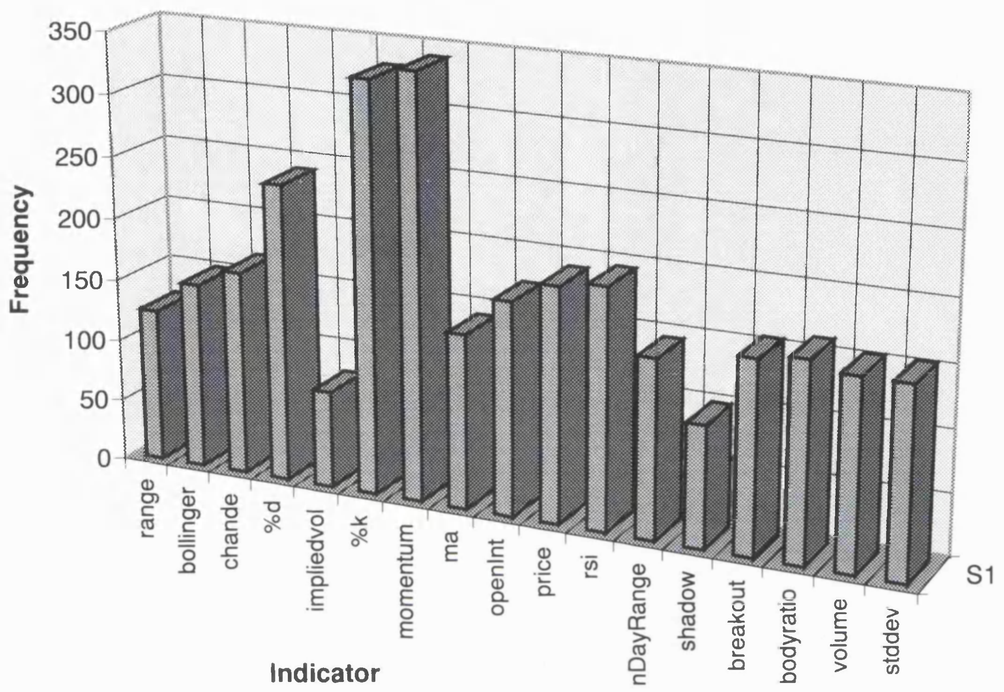
Insight can be gleaned into the behaviour of the system if the rules that the GA finds are examined. From an analysis of the distribution of the indicators used in the rules, it is clear that the GA finds more rules that are based on momentum than any other type of indicator. See Figure 6.15. The three highest peaks are momentum, %k and %d, which are all momentum based indicators. The predominance of these types of indicator indicates that it is easiest to find momentum based rules that operate adequately on both the training and validation sets. This is because from the rule framework that the GA is given, momentum is the most reliably profitable aspect of the market's dynamics.

This is an interesting result, for several reasons:

1. This means that the most reliably profitable action was to go with, not anticipate, market movement - momentum is quite literally, a "velocity of price movement" over a specified time-period. An inference from this is that the market price has a degree of 'inertia'. In other words, what the system has discovered is that the most reliable way to make profits in the interest rate markets is to go with the trend.
2. The longer maturity markets are more volatile and hence the daily movements are correspondingly larger. This could partially explain the effect observed in the

maturity experiment - that the system makes greater returns on longer maturity markets.

Figure 6.15: Frequency of indicator selection



3. This observation is also consistent with the behaviour of the system on gold futures. For much of the out-sample data that the system was tested on, the market was range trading. This will prevent a trend following system from operating effectively.

4. The idea that market motions persist is consistent with results from chaos theory: Measurements of equity's Hurst exponents reveal significant levels of persistence[Pete91].

It is also of interest to observe that the auxiliary data streams such as open interest, volume and implied volatility are not greatly used: the bulk of the indicators used act directly on the market price. In fact, the least used indicator is implied volatility, and there are several possible explanations for this:

1. This indicator has a very low information content. It is often argued, even by their most loyal devotees that "indicators should not be used on their own". If this is the

case it is debatable how much information any individual indicator presents. To present significantly less information than "an average indicator" would hint that implied volatility is nearly utterly useless in this framework.

2. A more pleasing (and likely) explanation would be that it is difficult to use implied volatility in a useful way in the supplied framework. The behaviours of indicators are compared to each other, and when certain relationships between the indicator's dynamics exist, the rule fires. If an indicator cannot be sensibly compared to many other indicators, then the fraction of the rule-space that mentions this indicator will necessarily be meaningless, and rules built from this area of the space are unlikely to be robust. Implied volatility is not directly comparable with other indicators, and consequently the fraction of the rule-space that mentions implied volatility is meaningful is small. It is interesting to note that sometimes, incompatible indicators *are* compared to each other: the trading range breakout indicator is only ever 0 or 1, and usually 0. The GA often uses this as a "switch" for examining whether another indicator is greater or less than 0.

3. The system is attempting to find rules for timing trade entries and exits. To attempt to do this with any other input than the price series incorporates an unnecessary level of ambiguity about the state of the market. For instance, if implied volatility suddenly rises, then this could mean that a large move in an unspecified direction was likely. A rule for going long based on such information will lose dramatically as often as it gains dramatically. However, information of this nature might be useful in a trading system that made use of options, or of delta-neutral trades[5].

6.11.2 Criticisms of the Method

There are several criticisms that could be levelled at this work which are important to explore. These fall into two main categories: criticism of the results and interpretation, and criticism of the viability of the system in terms of management management and the psychology of trading.

[5] Delta is how the net present value of the trade varies with the market price. A delta-neutral trade's net present value is consequently unaffected by price movements.

The results are not risk adjusted: This is a very important issue. Financial theory asserts that excess returns (ie. returns greater than that made by an index such as the FTSE100 or S&P500) can only be made by taking on more risk, by gearing up investments. The economists would argue that what has been done here is that the system is really under-performing the market, but that it is geared up so that the percentage returns are greater than those made by the indices. If this was the case, then risk-adjusting the returns would expose the extent of market under-performance. This argument is misplaced and risk adjusting the returns is a debatable proposition:

1. For this system to work depends on the market not behaving as economists believe. Therefore, the behaviour of this system must be in some sense disjoint to financial theory. It is of debatable wisdom then to attempt to apply (linear) financial theory to an area of non-linear behavioural study that must, by definition, lie outside the domain of linear econometrics. The robustness of linear models collapses when taken even slightly outside the assumptions of the problem domain - this study *relies* on assumptions of randomness, equilibrium pricing, investor rationality, and market efficiency being infringed, and the market price consequently having exploitable dynamics. To then suddenly assert that all the assumptions are valid again is dubious.

2. This study contributes to a growing volume of work advocating a paradigm shift to a non-linear view of economics. Arguments by analogy are often weak, but in this case can vividly illustrate the problem. Assume that the financial theory of the day operates around the motion of the planets. If someone then devised the current financial theory, it would be ludicrous to expect them to "moon adjust" their results to gain credibility. Typically, when paradigm shifts take place, the existing theory that is replaced is a special case of the new theory, and importantly - effects become visible that were invisible from the viewpoint of the previous paradigm[6]. Consequently, for a non-linear system, a non-linear version of risk-adjustment must be carried out.

[6] For instance, Newtonian mechanics is a special case of the Theory of Relativity, but concepts such as time-dilation or gravitational lensing are invisible from within Newtonian mechanics[Cham82].

3. The most widely used measure of risk, the variance of the P&L, depends on what length scale is used. The length scale must then be chosen *ad hoc* over which to find the variance[Pete91]. For the variance to have meaning, the distribution of returns must be normal or Gaussian. It has been acknowledged since the 60's [Osbo59] that the distribution of returns is fat-tailed - i.e. not normal. Mandlebrot [Mand82] asserts that if the markets display stable Paretian behaviour then their variance is infinite. Moreover, as the system only ever stops and reverses, it is always in the market, so the volatility of the P&L must be a simple function of the volatility of the market. According to the financial theorists, market volatility exhibits GARCH behaviour, where periods of persistent high volatility are followed by persistent low volatility, with random changes between the two. The variance is then an inadequate measure of risk, just as the average number of a die is 3.5.

The whole exercise is simply curve fitting: This is another very serious criticism, and one that is extremely difficult to counter with experimentally supported argument. The argument goes that as the system is developed, the developer runs tests, and makes modifications on the basis of the results of those tests. The first time those tests are run, the test is genuine, but thereafter, a feedback loop exists through the experimenter, and the validity of the results is progressively compromised as more and more experiments are run.

The defence against this criticism is that all the project development was done on the Eurodollar 3 month market. Only once the system was operational was data from other markets used. The P&Ls shown in Figure 6.6 were the first use of those data streams, the maturity experiment was also blind, as were the commodity experiments. This defence is not complete, as economists would then assert that all the markets are highly correlated, and so there has been an implicit use of future data once the system was complete. However, it is clear from Figure 6.6 that even if the markets are highly correlated, the system responds to them in different ways. If the system behaves differently to different but correlated markets, then either it is acting randomly or it is responding to some behavioural component of the markets that does not show up in the correlation calculations.

The system is behaving randomly: It is likely that some aspect of the behaviour of this system is effectively random, as it is being driven by a market that, at times, is probably near random. However, if the system's behaviour is completely random then it is unlikely that a confidence level of 86% would emerge for the existence of some dependence on maturity. While this value is not so high as to place the existence of this effect beyond all doubt, it is too high to be dismissed out-of-hand as the product of a random effect.

The system operates completely systematically on gold. The probability of sustaining a series of losses as long as is shown in Figure 6.11 as the result of random trading is very low. It must be emphasised that this is not the system simply losing equity through commissions, bid-ask spread and slippage - these factors contribute approximately 10% of the total losses made. There are simply practically no winning trades, and this demonstrates that the system is, in some sense, operating consistently.

The results are only one sample path: This point is closely related to the previous criticism - if a system has a stochastic component to its behaviour, then it will clearly take some time to assess its behaviour. It is true that this is one sample path, but there are some points that are important to recognise:

1. No out-sample test conducted as part of this investigation uses less than 1100 days of market data - this constitutes simulated trading on 12 markets over 3 time-frames over a period in excess of 4 years. Through an appeal to reality, most managers would feel that they would be in a position to make an informed assessment of a trader's performance after this time.

2. The system has consistent behaviour: The dynamics are similar for the closed cumulative profit and loss curves for the initial investigation and the maturity experiment(Figures 6.7, 6.10) despite the fact that these two P&Ls come from experiments whose target markets only overlap to a limited extent.

3. To a very real degree, this is a criticism that is independent of the quality of the work: once data from a market has been used, there is no way to generate completely independent sample path for that period. This is more a criticism of the fundamental methodology of analysing time-series and attempting to find exploitable determinism that is precluded by financial theory, than a criticism specifically aimed at this work.

The system is untradable: The validity of this statement depends entirely on one's risk tolerance and what one's definition of success is. Although the concept of risk as an analytical tool was criticised earlier in this chapter, it is useful as an intuitive notion of the likelihood of simply losing money. For how long and to what degree can the dealer, and more importantly, his management, handle holding a losing position? As for success, what is the benchmark? Until the technical analysts agree with the financial theorists about what the system has to do to be a success, this is not a question that can easily be answered, short of "put us in the top 5% of hedge funds".

A case has been made that the system should be used on the longer maturity markets. Similarly, it appears that the system trades the US markets consistently better than the other markets that have been tested. To test this hypothesis thoroughly will unfortunately take around 4 years for the market to generate new, untainted data.

6.12 Summary

The main points of this chapter were as follows:

- The problem was to develop a genetic algorithm based system that could find and exploit tradable behaviour in interest rate futures markets.
- The data available was daily open, high, low closes for a range of European, Japanese and American government bond futures markets. Additional data such as trading volumes and open interest were also available.
- It appears that the system works more effectively on the longer maturity markets. It is difficult to account for this effect from a conventional econometric viewpoint.
- The system was unable to form reliable models for either of the commodity futures markets tested.
- From an analysis of the rules, it appears that price movements have a tradable degree of persistence.
- A new methodology for examining the stationarity of markets is proposed.

Chapter 7:

The (D-)efficient Market Hypothesis

In this chapter, the assumptions behind the Efficient Market Hypothesis are examined and the implications that this has for financial theory and the validity of different investment strategies. Empirical evidence for and against the Efficient Market Hypothesis is presented, along with a discussion of the implications of the results of the financial experiments presented in this thesis.

7.1 The Efficient Market Hypothesis

The Efficient Market Hypothesis is a theory that can be expressed in terms of the information required to make excess profits in the financial markets without simply leveraging capital[King96]. It has three versions, each with varying strengths of assertion. They are summarised in the table below[King96]:

Table 7.1: Efficient Market Hypothesis Summary

Hypothesis form	Impossible to profit from the analysis of
Strong	i). Private (insider) information ii). Public information (dividends, eps etc) iii). Price histories
Semi-strong	i). Public information (dividends, eps etc) ii). Price histories
Weak	i). Price histories

There are some subtle points to bring out about each of these:

The Strong EMH: This precludes the generation of risk-adjusted excess returns through analysis of any sort whatsoever - the current price of a security reflects all that is knowable about it[LoHa73]. This view is (understandably) unpopular with analysts in the investment industry, as it means their jobs are meaningless and pointless[Malk90]. From a less emotive position, this view is perhaps excessive as a consequence of this hypothesis is that even those with inside information cannot profit from their knowledge, which is probably not the case.

The Semi-Strong EMH: The term Efficient Market Hypothesis is usually understood to refer to this form. It accepts that it is possible to make excess returns with private information, and the price reflects all public information - the current price forms as an aggregate of the independent opinions of analysts about the correct price. This view is more acceptable to analysts because the efficiency of the market is now due to their efforts, not in spite of them. It is clear from pension fund league tables that it is hard to reliably out-perform indexes such as the S&P500, and indeed many funds operate by tracking such indexes. Malkiel [Malk90] claims that through the 70's and 80's "approximately 70% of the equity portfolios have been out-performed by the S&P500", and that "the proof of the pudding is in the eating".

The Weak EMH: This form only holds that it is impossible to profit from examining the history of a security. The usual line of argument is that "the information is already in the price", and if any exploitable structure or pattern exists within the history, then the actions of those trying to profit from them will destroy the patterns. This *is* the case for cyclical patterns, but this line of argument is not watertight as it makes no provision for the presence of auto-catalytic effects, which are usually assumed to be untradably small in size. These "bubbles" of rising prices are regarded as anomalies by financial theory[Arth95], but it will be shown that they are a natural consequence of relaxing the assumptions about investor's expectations being homogeneous. Not only that, but that these bubbles can be of considerable size.

To further examine the validity of all forms of the EMH, it is natural to examine the assumptions that underpin this hypothesis.

7.2 Expectations, pricing, tuiip bulbs, bubbles and GARCH

This section is largely derived from [Arth95], which makes many points relevant to this thesis. There are a number of assumptions important to financial theory [Arth95,Pete91]:

1. Investors are rational.

2. Investors have homogeneous expectations.

An in-depth discussion of the rationality of investors is beyond the scope of this work, but a body of work exists questioning this assumption [KaCo90,Plum96]. We can also note in passing that the concept of the rational investor appears inconsistent with the casual observation of a trading floor when the market is dropping rapidly. Indeed, one of the hallmarks of a successful trader is the ability to remain rational under pressure[Schw84]. Instead, the second assumption will be discussed.

7.2.1 Are expectations really homogeneous?

Risk constraints and time-horizons are external factors specific to each investor, in addition to a suite of managerial and political issues, so it is unlikely that individual investing agents could react in a homogeneous manner even if they have homogenous expectations derived from homogeneous information sets. (It will be ignored that some market participants spend a great deal of money enhancing their information sets). It is irrelevant whether market participants behave differently due to their different circumstances, or because they really do have different expectations: the result is the same, they behave differently. For any single trade to take place, there must be a dealer trying to buy and another trying to sell. The price then crystallises as an aggregate of the buying and selling actions of the investing agents. However, the actions that give rise to this price are heterogeneous in nature, and so it is difficult to positively establish that expectations really were homogeneous in the first place. If expectations really are homogeneous -why are these agents trading with each other? Indeed, in the semi-strong EMH, the price forms from the *independent* value assessments made by analysts. This implies that heterogeneity *must* exist in the analysts' minds.

One very interesting observation that Arthur[Arth95] makes is that "actions taken by economic decision makers are typically predicated upon hypotheses or predictions about future states of the world that is itself in part the consequence of these hypotheses or predictions". In short, any hypothesis about the future of the market that is acted on, will to some degree, be self referential. An example of this is the Head and Shoulders chart formation already presented in Section 4.2. Olser and Phang have demonstrated that in certain currency markets, this pattern has predictive power[OsPh95]. Interestingly, it does not matter whether charting works or not - currency dealers are aware that lots of other dealers will be aware of Head and Shoulders formations, and hence the appearance of this pattern will influence the market price. This idea will be developed further, as a consequence of this self-referencing of trading hypotheses is that pricing is now *logically indeterminate*, as the following exposé will demonstrate.

7.2.2 Why homogeneous expectations lead to problems with pricing

It is straightforward to show that that the price of a security will settle at a level of[Arth95]

$$p_t = \beta(E[p_{t+1} \mid I_t] + E[d_{t+1} \mid I_t])$$

where p_t is the price today, β is a discount factor derived from the risk free rate, $E[p_{t+1} \mid I_t]$ is the investor's (shared) expectation of tomorrow's price given an information set I_t, and $E[d_{t+1} \mid I_t]$ is the expected value of tomorrow's dividend. But herein lies a problem. The formation of today's price contains the expectation of tomorrow's price, and this will, by a similar argument depend on the "expectation of expectation" of the following day's price. So now the investor is in a difficult position - he must determine what the expectation of expectation of price will be from the information set I_t. Not only must each market participant assess what information is in I_t but assess what other investors will make of it. For this reason, Arthur calls I_t a Rorschach inkblot[1]. In addition, there are no objective means for an investor to know another investor's assessment of the valuing of future dividends. Indeed, there may be

[1] It is interesting to note Arthur's terminology here - on page 40 of the "The Money Game", 'Adam Smith' uses a Rorschach inkblot to convey a similar point, in a book first published in 1967.

several plausible, defensible but mutually inconsistent viewpoints about the future state of the market inferable from I_t. Given that an investor does not know the others' expectations, and knows that these expectations will be based on *their* expectations of others' expectations, how can the investor arrive at their own expectation? As Lord Keynes once remarked, stock selection calls for assessing what average opinion thinks the average opinion will be[Smit76].

It is important here to point out that this analysis makes no assumptions about the information processing capabilities of the investing agents. They can be *arbitrarily intelligent*, but, as Arthur puts it, still unable to "implement deductive rationality".

To further examine the effect of the assumption of homogeneity on markets, Arthur has conducted a pair of experiments with artificial stock markets[Arth95]. The market is composed of a pool of a number of investing agents, each with a number of expectational models - that is, expectations about the future movement of the market. Expectational models are discarded when they are invalidated by the market's behaviour. Expectations that the price represents fundamental value would be valid (and self-validating) as long as these expectations were held by others. However, if a sufficient density of self-fulfilling alternative hypotheses exists, and the price trends upwards by chance, then a temporary price bubble might occur as more and more agents buy into the trend and sustain it artificially. The fundamentalist pricing can then become inaccurate, and this can clearly be important when dealers are marked-to-market[2] on a daily basis. A brief overview of these two experiments will now be presented.

7.2.3 Artificial Stock Market with Assumptions of Homogeneity

The artificial stock market was initialised with all agents having the same expectational model of fundamental-value pricing, and Arthur found that it was very difficult for non-fundamental expectational hypotheses to "get a footing". Therefore, if the market is sufficiently dominated by fundamental value pricing, then it is stable. This is conventional finance theory - that prices converge to their equilibrium level.

---

[2] Marking to market is the process of calculating the net present value of a portfolio (or trade) from that day's closing market rates and prices.

7.2.4 Artificial Stock Market without Assumptions of Homogeneity

The experiment was initialised as before, but the expectational models were taken from a random distribution surrounding the fundamentalist expectations. Arthur found that trend-following beliefs would appear capriciously and with sufficient density to become self-fulfilling. The dynamics of the artificial market were now pervaded by rallies and crashes, which are features of "real" markets. Arthur also observes that these artificial markets display GARCH (generalised auto-regressive conditional heteroscedastic) behaviour, where persistent high volatility is followed by persistent low volatility, with random shifts between the two. This makes no sense in standard economic theory, and similarly, the homogeneous expectations version of this experiment displays no GARCH behaviour. The fact that the artificial market displays GARCH behaviour when expectations are heterogeneous, but does not when expectations are homogeneous, is a novel, unexpected and important result.

7.3 Bubbles and Tulips

Of course, the academic finance community have been aware of bubbles for a long time, but contend that they are a small and infrequent anomaly in the markets, which cannot be reliably profited from. Bubbles do not have to be small. The classic example of a collapse of a price bubble occurred in Holland in 1637[Smit76]. In the previous four years, the value of tulips had registered a 20-fold increase for little more reason than it was fashionable to have a tulip with a certain type of virus and there were more buyers than sellers. Some people began to cash-in their profits, the bubble burst, and the price of tulips came crashing back down again. The example above is more than 350 years old: it would be a simple matter to assert that today's market practitioners would not engage in something that was so patently unsustainable. However, markets do not change very much - by definition they are a mechanism for people to buy things from sellers, and the motives of those participating in the markets have not changed fundamentally either: the exchange of tradables and the pursuit of profit. In the last few years there have been a number of stock market crashes(i.e. bubbles bursting), such as October '87, October '89, October '97 and at the time of writing, a bubble exists in the US stock market, which analysts have been projecting the end of for four years. However, given that a bubble exists - what should a fund manager do?

If a fund manager decides not go with the fund management herd (who are going with the bubble), and is wrong, then he must then out-perform the market just to get back on even terms. If he goes with the herd and the herd is wrong, the relative standing of his fund to the others with not be seriously affected. A dealer can be right in the long run - the bubble will burst, but if he is marked to market frequently, has to meet benchmarks, or be in the top 25% of pension funds, then being right in the long run can occur long after "being rationalised" in the short term for underperformance.

There are significant psychological, managerial and political factors, extrinsic to the process of trading, that have an impact on the actions of dealers and hence affect the market price. These can be as varied as stop losses(being stopped out of a trade pushes the market further in the direction it was going in), having to demonstrate competence to management and shareholders by ensuring that the portfolio contains the latest 'hot-stock' (which pushes the price up further) and even anticipating what very large investment institutions are about to do. The list is endless. Traders must then make some decisions based on criteria other than financial econometrics [Plum96,KaCo90,Schw84], and as is revealed in the previous section, theories based on homogeneous expectations will only work if they are by far the dominant strategy[Arth95]. The presence of rallies and crashes and GARCH processes in the financial markets indicates that they may not be. To study this further, non-linear methods will be used.

7.4 Hurst Exponents

The Hurst exponent, H, is a means of classifying a time-series[Hurs51]. It was developed by Hurst, a hydrologist who was studying the changes in water levels at the Nile River Dam. It was natural to assume that the distribution of reservoir levels would be normally distributed, as it was the product of a system with many degrees of freedom. However, Hurst found that the distribution of levels was not normally distributed, and so developed a classifying test statistic, the Hurst exponent, that did not depend on assumptions of normality. It is a test capable of classifying time-series as random, mean-reverting, or persistent (trending).

Unlike many statistical tests, the time-series does not need to be normal distributed - this test can distinguish between random non-normal series and non-random non-normal series. It is for this reason that this test is valuable. Market analyses that depend on assumptions of normality are to a degree tautological - as Hull[Hull93] notes on the empirical testing of Black-Scholes option pricing:

> "The first problem is that any statistical hypothesis about how options are priced has to be a joint hypothesis to the effect that (1) the option pricing formula is correct; and (2) markets are efficient. If the hypothesis is rejected, it may be the case that (1) is untrue; (2) is untrue; or both (1) and (2) are untrue."

It is not unfair to say that the option pricing formula will be derived from "conventional financial theory" and its assumptions/reliance on the notion of market efficiency, and hence the tautology arises. While market efficiency does not necessarily imply that the market prices exhibit a random walk, (and therefore that assumptions of normality will hold), the price being a random walk (and hence being normal) does imply efficiency. Many statistical tests do depend on assumptions of normality, and this is a stronger requirement than the market just being efficient. For example, the derivation of the Black-Scholes pricing formula makes repeated use of generalised Wiener processes that follow geometric Brownian motions. An implied assumption here is that the markets must then be normal and not merely efficient.

These discussions are operationally irrelevant for the Hurst exponent, as few assumptions are made about the behaviour of the series under analysis, and no assumptions are made about market efficiency. The brief discussion given above is presented to lend credence to the application of a non-market specific technique to the market. The Hurst Exponent was originally published in [Hurs51], but an informal description is given here.

This technique works by examining how the rescaled range of the series(the high-low range divided by the standard deviation of the series over that interval), R, increases with time, T. For a random-walk series, the rescaled range increases with the square root of time: $R = aT^{0.5}$. The longer the period over which the range measurement is made, the larger the measured value - but the functional relationship between the two variables remains approximately constant. If the series is mean reverting, the range expands slower than if the series was random ($R=aT^{H}$, $H<0.5$) as the series has a greater tendency to revert to its original value. Similarly, if the series is persistent, or trending, the range expands faster than if the series was random ($R=aT^H$, $H>0.5$), as the series has a greater bias to carry on moving in an established direction. This is summarised in Table 7.2[Pete91].

Table 7.2: Hurst Exponent Summary

Series behaviour	**Rate of rescaled range expansion**	**Hurst Exponent**
Random	Square root of time	0.5
Mean reverting	Slower than random	$0 \le H < 0.5$
Persistent(trending)	Faster than random	$0.5 < H \le 1$

7.4.1 Hurst Exponents and the Capital Markets

Peters[Pete91] has carried out a survey of the Hurst exponents of the monthly returns of a range of securities from various markets. This is shown in Table 7.3.

A Hurst exponent of 0.5 indicates that the underlying market is following a random walk, while a Hurst exponent greater than this indicates the presence of trending behaviour. Clearly none of these securities except the Singapore dollar has a Hurst exponent of 0.5. There are two points that are important here:

1. The Singapore dollar/US dollar exchange rate is a random walk, because the Singapore dollar is pegged to the US dollar. The randomness of this series reflects the random timing of the trades required to bring the Singapore dollar back into line.

2. Clearly, the more information is available (the greater the period of data available), the lower the level of noise associated with the Hurst exponent sampling. Peters uses some rather *ad-hoc* reasoning to determine the amount of data that is required, and (fortunately, perhaps?) determines that he has enough. However, Peter's motive is to make the case that the Hurst exponents for the financial markets (except in cases of pegging) are not 0.50. The most convincing demonstration of this is in the scrambling experiment he conducts on the S&P500.

Table 7.3: Hurst exponents and the Capital Markets

	Security	Hurst Exponent
Individual Stocks	IBM	0.72
	Xerox	0.73
	Apple Computer	0.75
	Coca-Cola	0.70
	Anheuser-Busch	0.64
	McDonald's	0.65
	Niagra Mohawk	0.69
	Texas State Utilities	0.54
	Consolidated Edison	0.68
Stock Market Indices	S&P 500	0.78
	MSCI UK	0.68
	MSCI Germany	0.72
	MSCI Japan	0.68
Dollar exchange rate	Sterling	0.61
	DeutscheMark	0.64
	Yen	0.64
	Singapore Dollar	0.50
Government Bonds	Treasury Bill (yield)	0.65
	US Bond (yield)	0.68

It can be seen from Table 7.3 that the Hurst exponent for monthly returns of the S&P500 is 0.78. When the series is scrambled, Peters reports that the exponent drops to 0.51 - i.e. random. This is a clear indicator that the scrambling process has destroyed the persistent, trending behaviour that was originally present in the S&P500, and consequently a demonstration that the S&P500 series is not a Markov process with no memory. That market movements are independent is one of the foundations of financial theory.

7.4.2 Approximate Entropy

The Hurst exponent is not the only means of assessing whether time-series are random or not. Working as what Stewart calls a "freelance mathematician"[Stew97], Pincus developed a test of randomness called Approximate Entropy (ApEn)[Pinc95] that, like the Hurst exponent, makes no assumptions about the process that generates the time-series. It works by examining blocks of (binary) data, for instance 101, and if this block is usually followed by a 1 then the process that generated the series must have a degree of predictability. Similarly, if the block is followed equally often by a 0 as it is by a 1, then the series is unpredictable with respect to this block. The Approximate Entropy measure is the average unpredictability over all blocks. Pincus used this technique to examine whether stocks in the S&P500 are random or not, and concluded that they were far from random.

There was one intriguing exception to this, however. "Interestingly," says Pincus, "in the 1987 to 1988 time frame, there was a unique single two-week period in which ApEn was nearly maximally irregular—precisely the two weeks immediately preceding the stock market crash of 1987."

7.5 In Defence of Market Efficiency

Ironically, one of the most impressive arguments for market efficiency - or at least, the difficulty of beating the index, comes from the dealers themselves: the movements of stock market indices such as the FTSE100(UK), S&P500(USA), Dow Jones Industrials(USA), DAX50(Germany), CAC40(France), Nikkei 225(Japan) etc. are almost exclusively derived from the buying and selling of the large institutional investors: pension funds, insurance and life assurance funds, hedge funds and unit trusts. Very few other market participants have the financial resources to have any significant impact on the markets. It is the institutional dealer's aggregate behaviour that determines what happens to the market indices. When the institutions are buying, the demand drives the price up, when they are selling, the price comes down. They drive the movement of the index.

Given that dealers must operate from within a framework of commissions, bid-ask spread and slippage, there are costs incurred in participating in the market. So for every pound that one dealer out-performs the market, another dealer must under-perform by more than a pound, as the brokers and market-makers will take their cut. Consequently, it is to be expected that more dealers will under-perform the market than not, and indeed this is borne out by pension fund league tables[Malk90]. Even those funds that track the index passively[3] under-perform due to operating costs and commissions, bid-ask spread and slippage incurred in keeping the components of the portfolio in the correct ratio. Curiously, it is difficult to imagine half, or even any dealers being confident of under-performing over the next 12 months, but this is an inevitable consequence of the way in which the market indices are calculated.

Note that even though it is difficult to reliably out-perform the market indices, this does not necessarily mean that the market is efficient in the sense that the economists mean - the market may digest and value new information very rapidly and adjust to new prices, and out-performing the index may be difficult - but this does *not* necessarily imply that the price is either fair, at equilibrium, a product of investor rationality or following a random walk.

7.5.1 Empirical Studies of Market Efficiency

A large body of literature exists on empirical tests of whether markets and exchanges are efficient or not. Given some of the discussion already presented in this thesis, it is perhaps unsurprising that the results are mixed, depending on what tests are applied and to which markets.

Much empirical evidence has been published that is insufficient to refute theories of market efficiency (read: the market is effectively efficient) in the Johannesburg [ThWa95] and Canadian[AlHa93] stock markets, the Australian interest rate futures market[Tan92], and the Tokyo foreign exchange spot market[LaNa92]. However, other authors have claimed that the German stock market has behaviour inconsistent with even the weak form of the efficient market hypothesis[Stoc90]. Surveys of emerging markets, such as the Istanbul[ZyBK95], and Warsaw [GoRi95] stock exchanges have been carried out, but almost by definition, an emerging market will

[3] The index is usually an average (sometimes weighted) of the market price of a number of stocks. Passive tracking funds operate by owning stocks in the ratio that they exist in the calculation. If the weights are changed, a stock splits or is replaced in the index, then clearly the portfolio must be updated.

not be 100% efficient and indeed this is one of the conclusions that the authors draw[ZyBK95].

7.5.2 Types of Test

These rejections or otherwise of the various forms of the EMH clearly depend to some extent on which tests are carried out. To explore this further, specific experiments that have different theoretical roots will be examined:

The first is a test of technical analysis: Donaldson[Dona90] concludes that the London stock exchange is inefficient because he observes that the FTSE100 index reacts to the proximity of a 00 resistance level (2300, 2400 etc). The price appears to have difficulty penetrating such levels from below, but once breached, the price moves up further and faster than expected. This critical level information is "rationally irrelevant" as the scale is arbitrary, but can reduce forecast errors and so Donaldson[Dona90] rejects all forms of the EMH. At this point, many financial theorists would argue that excess profits cannot be made from the existence of such effects, once trading costs have been taken into account, and so this is not a 'useful' refutation of the EMH.

The second experiment is one that stems from financial/statistical approach: Ali and Hasan[AlHa93] attempt to question the validity of the efficient market hypothesis with statistical, rather the non-linear tests. They study the Canadian stock market with vector auto-regression (VAR) techniques, and unsurprisingly do not find evidence against the EMH. Any experiment searching for exploitable *linearity* in the market is unlikely to be successful, as either:

1. These are exactly the effects that would prevent the Efficient Market Hypothesis from being formulated or widely believed in the first place.
2. They would have been found, exploited, and consequently arbitraged away.

In general, it appears that (non-linear) tests of the EMH by experiments whose assumptions are disjoint from those of financial theory, will tend to strain the EMH, while statistical overviews of market behaviour or experiments that share assumptions with the EMH will be joint hypotheses, and probably fail to reject the EMH with any confidence.

7.6 Summary

The main points of this chapter were as follows:

- The Efficient Market Hypothesis is introduced and literature for and against it is surveyed.
- The assumption of homogeneous investor expectations is discussed, and evidence is presented that relaxing this assumption leads to price bubbles and GARCH behaviour in markets.
- Hurst exponent analysis of financial markets reveals significant trending components.

Chapter 8:

Assessment and Conclusions

In this chapter, the thesis goals are reviewed and each of the projects critically assessed and conclusions drawn. Results that have an impact on the Efficient Market Hypothesis are discussed, and the chapter concludes with a discussion of genetic algorithm rule induction..

8.1 Thesis Objectives

The goals of this thesis, as stated in chapter 1 were:

1. To investigate the relationships between the intelligent techniques and the applications. This took two forms: the comparing and contrasting of various techniques for specific problems, and the exploration of individual techniques across a range of applications.
2. To bring previously proprietary intelligent financial systems knowledge and research to the public domain.
3. To further the field of intelligent systems through the development of new techniques and the enhancement of existing methods.
4. To make novel observations, statements and hypotheses about the nature of the problem domain(s). Most of the projects have been concerned with trading financial markets.

Each of the projects undertaken in the course of this research had all of these goals, but in addition, also had the requirement of addressing the business problem. Each of these projects will now be critically assessed, and conclusions presented.

8.2 Neural Networks for Residual Value Forecasting

8.2.1 Project Summary

The objective of this project was to use neural networks to forecast the residual value of new vehicles in 3-4 years time, to an accuracy of £50. A database was available of the prices of new and second-hand vehicles in a range of conditions, at monthly intervals for 3½ years. A synthetic depreciation series was constructed for proof-of-concept purposes. A pair of neural network time-series forecasting models were compared to a pair of linear benchmarks for this data.

8.2.2 Assessment

This project did not meet these objectives and instead has brought some important points to light on the application of intelligent systems. There were several reasons why the system did not operate as desired, and many of these stem directly from the choice of technique: namely neural network time-series forecasting. These assessments must ultimately stem from the behaviour of the models under out-sample testing (see Table 8.1):

Table 8.1: Residual value forecasting errors

Technique	Variant	RMS testing set forecast error (£)
Regression	Linear regression	478
	Exponential regression	593
Neural network	Time-series forecasting	627
	1st difference forecasting	3639

It is clear that for this application, for these implementations, that the linear techniques out-perform the more complex non-linear models. This is an important and positive result, as it is a valuable reminder that one does not have to recourse to building large, complex systems to get most of the information out of the data.

8.2.3 Project Conclusions

1. Complexity of technique/model should be appropriate for the complexity of the problem. The preprocessing that was carried out transformed the problem into one that had little forecastable non-linearity. The linear models work better because they implicitly use domain knowledge that is not available to the neural networks. If the synthetic series wasn't approximately a straight line, regression would not have been used. Regression was used because it was going to work fairly well. A neural network must process a reasonable amount of information before recognising that the system is near linear. If knowledge of the problem can be used to effect transformations that result in the forecast being trivial (as in this case), then the solution to the problem then becomes one of "undoing" the preprocessing to yield a forecast.

2. Residual value forecasting is a feasible proposition, but the original project target of an RMS error of £50 was perhaps somewhat optimistic. This study gives an informed basis from which to make expectations and assessments of the performance of forecasting technologies.

3. Attempting to forecast first differences of the time-series was ineffective, owing to the sensitivity of the reconstructed depreciation path to systematic errors. An inference from this experimental result is that when little data is available, model frameworks should be chosen to maximise their robustness, use as much domain knowledge as possible and have as few free parameters as possible.

4. It is important not to interpret this to mean that it is not possible to use neural networks effectively for this problem, simply that for the final form of the forecasting, after the preprocessing had been done, the linear models were more appropriate and consequently were more robust and accurate. If other data was available or a different approach had been taken to modelling depreciation, then it is possible that neural networks would be the most appropriate technology.

8.2.4 Future Work

1. It has been hypothesised in this thesis that the information required to make forecasts sufficiently accurate to meet the £50 RMS error simply was not present

in the data used in this project. A valuable project would be to attempt to enhance the information content available to the machine through the use of macro-economic data. It is possible that neural networks would be useful for this application.

2. An interesting project would be to construct an agent-based Monte-Carlo simulation of the depreciation of vehicles, given the forces of supply and demand that will operate at each stage of the vehicle's life. This is of particular interest as if a vehicle has a strong anticipated residual value, it will have lower hire charges and so probably be in demand. However, at disposal 3 or 4 years later, the market will be flooded with ex-fleet hire vehicles of that model.

3. Given the existence of a device for forecasting the aggregate depreciation across a vehicle range, a description of a method of making individual model residual value forecasts is described in section 3.8.1.

8.3 Genetic Algorithms for Trade Filtering

8.3.1 Project Summary

The aim of this project was to model expert trading knowledge without recourse to knowledge elicitation exercises. Genetic algorithm rule induction was used on a history of past chart trades to find rules that capture the knowledge content of the original trade history and to automate the discovery of new types of chart patterns.

8.3.2 Assessment

This project was very successful and comprehensively met its brief of capturing, reproducing and extending expert knowledge. This claim can be substantiated by the induction of a number of new trading rules on out-sample data that are statistically significant to greater than 95% confidence, and that the distribution of rule Z-scores is positively skewed. This demonstrates that chart patterns must have information content (consistent with [OsPh95]) and that the construction of useful trading rules can be successfully carried out with genetic algorithm rule induction. As a result, this project meets thesis objectives 1,2 and 4.

However, there are some limitations of this work that it is important to be aware of:

1. It is difficult to assess to whether or not the data was complete, honest and reflective of "real-life", or whether it had been massaged in some way. For instance, it is likely that stationarity would be an issue (i.e. that the expert's ideas and opinions changed over the two decades that the trading history spans), and possible that some of the history had been mislaid.
2. The applicability of the induced trading rules is narrow as it is only the trader who has all the information necessary to interpret the symbols in a rule. Some of these symbols such as "Bull" are self explanatory, but most are not, such as "Context: DES".
3. There are problems with performing repeatable empirical experiments on the financial markets. This will be discussed in greater depth in section 8.6, as this is an issue that is common to three of the four projects that have been carried out.

8.3.3 Project Conclusions

1. It is apparent that simple rules have higher Z-scores associated with them and appear more robust in general. This is due to two reasons: i) If a rule is simple, broadly speaking it will fire more often than one that has many clauses that must be satisfied. This leads to the rule being tested more thoroughly and hence given a higher significance; ii) If a rule is general it is less likely to be over-fit and consequently simply an artefact of the data. This leads to the more general result that simple trading rules are more likely to be effective than more complex rules.
2. Charts have information content but it is not possible to prove from the available information whether or not they can be used to beat the market.
3. This experiment has been a successful application of intelligent systems, and genetic algorithm rule induction has proved to be an appropriate approach for this problem.

8.3.4 Future Work

The most useful development to this work would be to re-test it on trades that have been entered since the original experiments were carried out. This would be a useful test to carry out as there is no possibility of data snooping, which is often a serious criticism of this type of study. This would also give the opportunity to assess whether charts have sufficient information content to make appreciable returns and/or beat the market.

8.4 The Continually Adaptive Trading Engine

8.4.1 Project Summary

The continually adaptive trading engine was an equity trading device, designed to attempt to remain attuned to the current market conditions. The system used genetic algorithm style operators on a rule framework to find and improve simple trading rules for generating buy and sell signals. The results of this work have been encouraging but there are some non-trivial outstanding issues concerning the funding of positions that need to be addressed before it will be clear whether the system is practical or not. In addition, as the system trades equities, it will receive a stream of dividend payments in addition to returns due to speculative trading returns.

8.4.2 Assessment

The objective has been partially met - it is evident that the system evolves and changes, and usually remains attuned to the market and trades effectively. However, it is incapable of always trading equity markets effectively, as is demonstrated by long flatish sections of the proxy-to-equity curve (Figure 5.5), although little risk-capital is lost in these periods.

In order to attempt either an actual deployment of the system or to construct an actual cumulative P&L curve requires strategies to be defined for fund management and the funding of positions . At present these are somewhat speculative and the precise details not known. It is possible that costs associated with any fund management strategy will rob the system of sufficient profits for the system to be relevant. However, the mean return per trade is greater than LIBOR, so if the system is making some combination of approximately LIBOR and super-LIBOR then it is likely, though not certain that LIBOR would be modestly out-performed. A more difficult problem is to establish which of the current funding options, with what parameters, is the best strategy without contaminating the data. It is also possible that doing effective fund management for this system is actually a more complex task than developing the trading engine itself. In addition, as the system trades equities it will receive a stream of dividend payments. The data required to calculate the effect of these payments on the P&L curve is not currently available.

Until the fund management and dividend payment issues are resolved, the impact of this work is somewhat limited due to the speculative nature of the results. However, it is interesting that one of these effects (fund management) is likely to reduce the system's profitability, while the other (dividends) will increase it.

8.4.3 Project Conclusions

1. It is likely that the system could make modest (approximately 1%) returns over LIBOR - the system makes modest returns as it stands, ex dividends, but also has large amounts of unused risk capital that could be lent in the money markets for additional returns. The assertion that the system would beat LIBOR can be made as the average return per trade after costs is 8.7%, which is greater than LIBOR (typically 4-5%).

2. The system cannot always evolve with the market, but does not take bad losses during these times. This is apparent from the lack of serious drawdowns in the proxy-to-cumulative-P&L curve (Figure 5.5). Given that the system cannot sell short (profit from a falling market) this is a good result, as to stay out of the market is the way to minimise losses.

3. The system is unable to trade the FTSE index, and enters no trades on this market. This is interesting as a common econometric benchmark is to beat a buy-and-hold strategy on such an index.

4. The results look encouraging at this stage for genetic algorithm rule induction.

8.4.4 Future Work

This work would benefit greatly from two extra pieces of work, as this would resolve the speculation that exists about whether the system is viable or not:

1. Resolution of the funding strategies: This will require implementation of a fund management strategy, and determination of the costs involved. A number of

options were proposed and discussed in section 5.6. In addition, as the overnight rate is not fixed, the price series for the relevant period will be required. LIBOR can be calculated from the price of Eurodollar futures[Hull93].

2. The inclusion of dividend payments into the cumulative P&L calculations will increase both the validity of the simulation and the returns made.

8.5 The Bull-Bear Trading Engine

8.5.1 Project Summary

Genetic algorithms were used to induce rules for speculatively trading government bond futures markets in Europe, Japan and the USA. The system was given a comprehensive and thorough treatment of trading simulation, and tested on a number of relevant markets. There appeared to be some dependence on system effectiveness with maturity and so an experiment was conducted to test this hypothesis. The system was also applied to metals commodities, but neither of these experiments worked well. A technique for monitoring the evolution of the character of markets was proposed.

8.5.2 Assessment

This project's objectives have been met partially. It is clear that the system broadly makes returns across the required markets and that the composite P&L has a positive drift component. The system is more effective over homogenous set of markets, namely the US Treasury bond futures. It is important to note that the actual percentage returns made are dependent on the level of deleveraging that the system is traded with. As is discussed in Section 6.11.2, the risk adjustment of this system is not a trivial task and so calculating the risk adjusted return is currently an open issue.

It was demonstrated that the effectiveness of the system has some dependence on maturity to a confidence of 86%. This result combined with the observation that most rules are momentum based can easily be incorporated into existing non-linear economic arguments concerning the existence of price bubbles and the GARCH behaviour of markets - effects that financial economists willingly admit occur. This is covered in greater depth in Section 8.6.1. These results fulfil goal 4.

That the system did not work well on commodities indicates that a palpable difference exists from the interest rate futures markets on which the system was developed and which it trades acceptably. This result would hint that momentum and trend trading would work markedly better on government bond futures than on copper or gold futures. This, and the statement that trend/momentum trading works best on long maturity futures markets are statements of widespread applicability.

The third goal of this thesis was to develop new intelligent techniques: the proposed novel methodology for assessing changes in the character of financial markets is currently underdeveloped but could be the basis of a useful and robust market characterisation technique. Given that the rules have information content, there will be information to be interpreted from the changing spreads between the static and updated rule P&Ls. However, no attempt has been made to assess precisely what this information is, beyond a basic statement of whether the current market is changing to be more or less like the market that gave rise to the original rule set. This technique is computationally demanding which will limit its applicability, but this technique is potentially a precursor to a new type of market characterisation tool.

This project has therefore met, at least partially, all of its research objectives.

8.5.3 Project Conclusions

1. The system is more effective in longer maturity markets. This effect was demonstrated to a confidence of 86% following blind out-sample experiments on the US Treasury bond futures markets of varying maturity.

2. From an analysis of the composition of the trading rules, it appears inferable that price movements have a tradable degree of persistence. That price movements have persistent behaviour is consistent with results from non-linear dynamics[Arth95] and chaos theory[Pete91]. This is an important result because this result also accounts for the effects of bid-ask spread, commissions and slippage, which in conjunction with random daily price returns, are usually hypothesised as preventing speculative trading from being a viable activity.

3. The previous two points can be combined to argue that price movements in longer maturity markets display a greater degree of persistence. This conclusion can easily

be incorporated into a framework of homogeneous expectations, GARCH markets and price bubbles, by making the (reasonable) assertion that pricing bubbles can be larger in longer maturity markets. The Bull-Bear engine exploits bubbles (persistent behaviour) and so operates more effectively in longer maturity markets.

4. This project was successful in finding profitable trading rules and systems. Genetic algorithm rule induction has therefore proved itself to be an effective technique for this application. In addition, conclusions from this experiment have contributed to non-linear models of financial markets.

8.5.4 Future Work

1. A worthwhile experiment to conduct would be to run the system without modification on the market history that has occurred since the original experiments were carried out. Thorough validation of the trading rules is one of the more problematic aspects of experiments of this nature, and testing on market history produced since system development terminated, completely prevents any possibility of inadvertent "data-snooping".
2. A system was proposed for assessing changes in the character of financial markets by comparing the out-sample trading performance of rules induced from different periods of market history. So far, only a basic and rather conjectural interpretation of the results has been attempted, and a worthwhile strain of research would be to investigate the operation and interpretation of this technique further.

8.6 Experiments on the Financial Markets

Three of these projects have been concerned in some way with attempts to trade financial markets aggressively, based on the analysis of past information. It is important to bear in mind that there some limitations to this methodology:

1. It assumes future is like the past. The validity of this assumption is debatable as some factors are obviously changing - the automation of exchanges, new financial products, new currencies, and changes to legal and political constraints to name but a few. The question is "are the important characteristics of markets changing slowly enough for the analysis of data histories to be useful?" What is deemed

"important", and "slow enough" are clearly dependent on factors such as the experimental technique chosen and what the objectives are.

2. Financial markets are not scientific test beds: a single market price path is available, and data migrates in-sample extremely rapidly [King96]. For instance, if a section of market history is used for testing a trading rule, then it is invalid to use any information from that experiment on any subsequent trading rule test that uses that same period of market history - to do so would be to effectively use information from the future in the second experiment.

3. There are also significant problems with the validation of trading rules, as for each market only a single price series is available to conduct experiments on. Consequently it is very difficult to make robust empirical assessments of whether a rule works or not. For instance, before accounting for trading costs, it would be expected that half of all possible utterly bogus (i.e. zero information) trading rules would appear to make a profit at the end of the test run. However, it is often not clear from inspection whether a rule has information processing capabilities or not. Rules can be required to perform acceptably across disjoint test sets, but as extra constraints are imposed on the characteristics of the cumulative profit and lost curve, such as tolerable draw-down, rule longevity, or rule applicability across markets, the experiment begins to be propelled into the realm of curve-fitting. The only answer is to carry out out-sample testing on completely uncontaminated data for reasonable periods of time.

8.6.1 Financial Anomalies - A Consistent Story

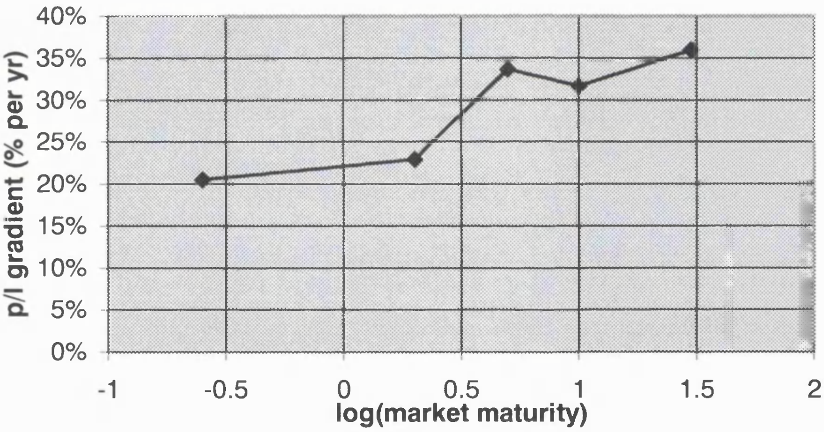
The experiments presented in this thesis provide evidence both for and against the EMH. From the survey of the literature that was presented earlier, this is not a surprising outcome. However, the evidence can be arranged in an attempt to form coherent picture. In particular, the Bull-Bear results and some of the points presented in Chapter 7 can be brought together into a common explanatory framework that can account for several anomalies of financial theory.

It can be seen from Peters' work[Pete91] that in general, most traded securities have significant non-random aspects to the price dynamics, and exhibit trending behaviour. This is in conflict with the econometric assumption that the movement of market prices follows a random walk. This assumption of the price following a random walk is also in conflict with the idea that it is possible to reliably profit from price bubbles.

A reasonable inference from Arthur's work[Arth95] is that relaxing assumptions of homogeneous expectations leads to financial markets rallying and crashing, and the market's volatility exhibiting GARCH behaviour. That markets display GARCH behaviour is an econometric construct, but that is in fact anomalous to the assumptions of financial theory.

The most important result of this thesis is that the Bull-Bear trading engine works better on longer maturity markets. See Figure 8.1. This effect has been demonstrated to 86% confidence, and there is no econometric explanation for this result. It is clear from analysing the rules that most are momentum based.

Figure 8.1: P/L gradient vs. Market Maturity: US Treasury Futures



This maturity effect can be incorporated into a non-linear, behavioural view of markets that is consistent with Arthur's and Peters' work. Markets trend - this is not really under dispute but is proved by Peters. If assumptions of homogenous expectations are dispensed with then price bubbles occur (and the simulation takes on behavioural qualities displayed by "real" markets). The Bull-Bear trading engine profits from going with price movements. Longer maturity markets are more volatile[1], and so bubbles can be larger in these markets. As the Bull-Bear engine profits from going with price

[1] The price of the bond is the sum of the coupon payments over its lifetime and the return of the principal. However, these cashflows must be discounted at the current market rates, and the longer the bond, the larger the number of payments that need discounting, and so interest rate changes have a greater effect on the prices of long bonds.

bubbles, it works better in the longer maturity markets. This view of financial markets is summarised in Table 8.2.

Table 8.2: Market Models

Conventional Financial View	**New View**
Markets are random	Markets trend (Hurst exponents > 0.5)
Expectations are homogeneous	Expectations are heterogeneous
GARCH is anomalous	GARCH natural consequence
Bubbles are anomalous	Bubbles are natural consequence
Long maturity markets are noisier	Bubbles can be larger in long maturity markets
Bull-Bear result inexplicable	Bull-Bear exploits bubbles

8.6.2 But don't overdo it

It is important not to neglect the evidence against rejecting the EMH:

1. The Genetic Algorithms for Trade Filtering project demonstrates that charts have information content, but this is not under dispute. What is unresolved is whether they contain sufficient information to out-perform a buy-and-hold strategy. This is indeterminable from the available information, and so the EMH cannot be rejected from this information.

2. The Continually Adaptive Trading Engine makes returns approximately 100 basis points (1%) over LIBOR. It is unclear whether this constitutes beating the market or not: if the markets are efficient, one should not be able to make money by borrowing in the interbank loan markets, and then simply investing it in the stock market, thus borrowing at around 5% and then making about 10% on stocks. For this reason, it is unclear whether making super-LIBOR returns is beating the market or not.

3. The Bull-Bear Trading Engine fails to operate on commodities. This is consistent with the market being efficient. The analysis implications of the system's behaviour on gold is ambiguous: the system failed to work, as is consistent with the EMH, but also behaves completely systematically which is not. However, for reasons explained in Section 6.9.2, this system cannot simply be traded the other way.

Each of these results are either consistent with the EMH, or insufficient to reject it.

8.6.3 Conclusions on the Efficient Market Hypothesis

The ultimate conclusion of this discussion of the Efficient Market Hypothesis must be that it is untrue. Just as Fortune concluded[Fort91], that there is overwhelming empirical evidence against the EMH. Tests of the EMH usually fail to reject it if the test in question is somehow derived from financial theory in the first place. Tests that have roots in ideas different from finance often find a hole in the theory.

The work presented here finds yet more holes in the EMH:

1. The Bull-Bear Trading Engine successfully trades most Government bond futures markets;
2. The presence of some relationship between the effectiveness of momentum based trading systems and market maturity has been demonstrated to 86% confidence.
3. The Continually Adaptive Trading Engine appears to out-perform LIBOR.

There is a body of literature already in existence sufficient to refute the EMH, and although work presented in this thesis makes a valuable contribution to this stance, it is insufficient to end the debate once and for all.

One reason for this lies in whether the experiments are "useful refutations" of the EMH or not. An experiment may punch a hole in the EMH, but if it enables investors to underperform the market, then the economists are rightly allowed to say "So What?". A similar response would greet any failures of the EMH that cannot be exploited once transaction costs have been taken into account. However, if an experiment makes excess returns then other problems arise - an example of this is the risk adjusting of returns (Section 6.11.2). These experiments are underpinned by auxiliary assumptions that leave the final state of the discussion in a peculiar state. As Fama wrote in summary in one of his papers[Fama90] "Whether the combined explanatory power of the variables - about 58% of the variance of annual returns - is good or bad news for market efficiency is left for the reader to judge".

Markets are not efficient, but they are sufficiently efficient for the EMH to be a working assumption - it is not a trivial process to devise a speculative trading system that systematically extracts profits from the markets. As Malkiel[Malk88] notes, there is a great deal of evidence supporting both sides of the argument, and "reports of the death of the Efficient Market Hypothesis appear premature". However, this statement was made 10 years ago, but the same academic jousting is taking place and options are still priced with variants of the Black-Scholes pricing formulae.

Some of the results presented in the course of this thesis suggest that the EMH is untrue. A large number of papers have been published in reputable journals by competent researchers that demonstrate that the EMH is invalid. A hypothesis cannot be restored to health by a second body of literature that shows that it has not been refuted[Cham92], and so as a result of these investigations, it must be concluded that the EMH is invalid.

8.7 Intelligent Systems

8.7.1 What works where?

The main research goal underlying this thesis is to explore relationships between the intelligent techniques and business applications. This has been carried out in two ways:

1. Neural networks and linear regression were compared for the task of residual value forecasting. Linear regression was found to work well and outperformed the neural network models. This was discussed in section 8.2.
2. Genetic algorithm rule induction was explored for a series of applications related to financial trading, and worked well in two of three projects and has given encouraging results in the third.

The effectiveness of the various techniques to the application problems have been summarised in Table 8.3.

Table 8.3: The Application of Intelligent Systems

Application **Technique**	Residual value forecasting	Trade filtering	Adaptive trading	Trading system induction
Genetic Algorithms		☺	😐	☺
Neural networks	☹			
Linear regression	☺			

Key:

☺ : Technique effective; 😐 : Results inconclusive; ☹ : Technique ineffective

8.7.2 Genetic Algorithm Rule Induction

Building overlapping meshes of models has proven to be a powerful means of modelling complicated systems. One reason for this is that simple, mutually inconsistent sub-models can be developed for application in different circumstances. As any rule will only be applicable in a limited range of circumstances, any rule will only need to have correspondingly limited explanatory or predictive power. This is particularly useful in dealing with the financial markets, as the nature of the relationships can change abruptly or even reverse. For instance, a raise in the interest rate can either result in a strengthening or weakening of the currency, depending on factors such as whether the interest rate rise was expected or not, the current market mood and the general health of the underlying economy.

Genetic algorithm rule induction can be thought of as finding pockets of predictability, or alternatively, the building of an overlapping meshwork of implicit partial models of the problem domain. For the purposes of discussion, a rule will be regarded as being of the form:

if <criteria> then <model statement>

The default assumption is that nothing can be done and that no meaningful rules can be expressed within the rule grammar about the target system. The genetic algorithm then attempts to find circumstances where this is untrue. The key to this approach is that the model statement is only applicable to the target system when the criteria are met. This gives rise to one of the main strengths of this approach: no attempt is ever made to build either a complete or consistent model of the target system. Instead, only aspects of the target system that are modellable by the framework are modelled. Those aspects of the behaviour that are random or difficult to model with the supplied framework will not give rise to reliable models of the system, and hence will rarely survive the validation procedure. Consequently, when the target system enters this unmodelled mode, the rules will largely remain mute (few rules have their <criteria> met) and the system implicitly acknowledges that the current behaviour is beyond comprehension. This is not a facility that is usually open to neural networks.

It is likely that completely spurious models will sometimes survive the validation process, but estimating the fraction of rules with information content is difficult. This is because there is little empirical difference between a worthless rule, a rule that does have value but where the system failed to exhibit behaviour where the rule could be successfully applied, and a rule that did have value but the system under scrutiny changed over the test period. In addition, even simple rules can exhibit complex behaviour, due to the complexity of their 'environment'[Broo90]. Given that the market must drive itself to the edge of predictability[WeGe93], the behaviour of even simple rules can be extremely complex. However, another strength of rule induction systems is that they can be interrogated to find out what rules gave rise to a particular action. Moreover, if apparently reliable rules are induced that are deemed to be nonsensical then these can be removed and prevented from influencing the final operation of the system.

These techniques are powerful, and with this comes a requirement for sensible use. It is extremely easy to simply export the responsibility for understanding the problem onto the computer, but this is an approach that will rarely work well. Rule induction has proved to be a particularly worthwhile technology, not only because it offers the possibility of autonomous model building, but also because it allows the researcher to decode and assess the resulting model. In a business environment this can be extremely important. Even if the machine found the model, it is advantageous to be able to tell management and clients about how the system works, and to be able to reassure them that everything is under control. The future is extremely promising for the application of powerful, transparent, well-chosen intelligent systems to business.

Appendix A - References

[AavL87] E. Aarts and P. van Laarhoven, "*Simulated Annealing: Theory and Practice*", John Wiley and Sons, New York, 1987.

[AdCh94] H. Adeli and N. Cheng, "*Augmented Lagrangian Genetic Algorithm for Structural Optimization*", ASCE Journal of Aerospace Engineering, v7:1, 104-118, 1994.

[AlHa93] S.M. Ali and M.A. Hasan, "*Is the Canadian Stock Market Efficient with Respect to Fiscal Policy? Some Vector Autoregression Results*", Journal of Econometrics and Business, 45, 1, February 1993.

[AlPh93] P.M. Allen and H.K. Phang, "*Evolution, Creativity and Intelligence in Complex Systems*", Vol. 62, Interdisciplinary Approaches to Non-linear Complex Systems, H. Haken and A. Mikhailov (eds.), Springer Verlag, 1993.

[Arth90] W.B. Arthur, "*Positive Feedbacks in the Economy*", Scientific American, February, 1990.

[Arth92] W.B. Arthur, "*On Learning and Adaption in the Economy*", Santa Fe Institute, 92-07-038.

[Arth95] W.B. Arthur, "*Complexity in Economic and Financial Markets*", Complexity, 1, 20-25, 1995.

[BeJa90] R. Beale and T. Jackson, "*Neural Computing – an Introduction*", Adam Hilger, Bristol, 1990.

[Bent98] P. Bentley (Ed.), "Evolutionary Design by Computers", Academic Press, 1998. (In Press)

[Bren73] R. Brent, "*Algorithms for Minimization without Derivatives*", Prentice-Hall, Englewood Cliffs, 1973.

[BrLL91] W. Brock; J. Lakonishock; B. LeBaron, "*Simple Technical Trading Rules and the Stochastic Properties of Stock Returns*", Santa Fe Institute, 91-01-006.

[Broo90] R.A. Brooks, "*Elephants don't play chess*", Journal of Robotics and Autonomous Systems, Vol. 6, 1990, p. 3-15.

[CAP97] CAP Blackbook, Used Car values, CAP Motor Research, May 1997.

[Cham92] A.F. Chamlers, "*What is this thing called Science?*", Hackett Publishing Co. Inc., 2nd Edition, 1982.

[ChKr94] T.S. Chande and S. Kroll, "*The New Technical Trader*", Wiley, 1994.

[Das94] S. Das, "*Swap and Derivative Financing*", Revised Edition, Irwin Professional Publishing, New York, 1994.

[Davi91] L. Davis, "*Handbook of Genetic Algorithms*", Van Nostrand Reinhold, New York, 1991.

[Deb91] K. Deb, "*Binary and Floating Point Function Optimization using Messy Genetic Algorithms*", Illinois Genetic Algorithms Laboratory (IlliGAL), Report no. 91004.

[Deb93] K. Deb, "*Multimodal Deceptive Functions*", Complex Systems 7:2, 131-153, 1993.

[Debo94] G. Deboeck (ed.) "*Trading on the Edge: Neural, Genetic, Fuzzy Systems for Chaotic Financial Markets*", John Wiley and Sons, 1994.

[Dona90] R.G. Donaldson, "*Is the London Stock Exchange a Rationally Efficient Market?*", John M. Olin Program for the Study of Economics Organisation and Public Policy, 58, October 1990.

[Eco92] "*Chartists Tame Chaos*", The Economist, 15 August 1992.

[Eman96] R. Emanuels, "*Practical Applications using Neural Networks*", ABN Amro Bank, July 1996.

[Fama90] E.F. Fama, "*Stock Returns, Expected Returns, and Real Activity*", Journal of Finance, 45, 4, September 1990.

[Fama96] E.F. Fama, "*Multifactor Explanations of Asset Pricing Anomalies*", Journal of Finance, 51, 1, March 1996, P. 55-84.

[Fort91] P. Fortune, "*Stock Market Efficiency: An Autopsy?*", New England Economic Review, March-April 1991.

[GoFe94] S. Goonatilake and K. Feldman, "*Genetic rule Induction for Financial Decision Making*", *in "Genetic Algorithms in Optimisation, Simulation and Modelling*", J. Stender, E. Hillebrand and J. Kingdon (eds.), IOS Press, Amsterdam, 1994.

[GoKh95] S. Goonatilake and S. Khebbal (eds), "*Intelligent Hybrid Systems*", John Wiley and Sons, 1995.

[Goon94] S. Goonatilake, "*An Intelligent Hybrid System for Financial Decision Support*", PhD Thesis, University College London, 1994.

[GoRi95] B. Gordon and L. Rittenburg, "*The Warsaw Stock Exchange: A Test of Market Efficiency*", Comparative Economic Studies, 37, 2, Summer 1995.

[GoTr95] S. Goonatilake and P. Treleaven (eds), "*Intelligent Systems for Finance and Business*", John Wiley and Sons, 1995.

[Gold89] D.E. Goldberg, "*Genetic Algorithms in Search, Optimisation Machine Learning*", Addison-Wesley. Read, Mass., 1989.

[Gold90] D.E. Goldberg, "*An Investigation of Messy Genetic Algorithms*", Department of Engineering Mechanics, University of Alabama, TCGA Report No. 90005, May 1990.

[Hart75] J.A. Hartigan, "*Clustering Algorithms*", Wiley, 1975.

[Harv92] I. Harvey, "The SAGE Cross: The Mechanics of Recombination for Species with Variable-length Genotypes", Parallel Problem Solving from Nature 2, North Holland, 1992.

[Hays81] H.T. Hayslett, "*Statistics*", Made Simple Books, Butterworth Heinman, 1981.

[Holl75] J.H. Holland, "*Adaptation in Natural and Artificial Systems*", Ann Arbor, MI: University of Michigan Press, 1975.

[Hopf82], J.J. Hopfield, "*Neural Networks and Physical Systems with Emergent Collective Computational Abilities*", Proc. Nat. Acad. Sci., vol. 79, pp. 2554-2558, 1982.

[Hull93] J.C. Hull, "*Options, Futures and other Derivative Securities*", 2nd Edition, Prentice Hall International Editions, 1993.

[Hurs51] H.E. Hurst, "*Long-term Storage Reservoirs*", Transactions of the American Society of Civil Engineers, 116, 1951.

[KaCo90] H. Kahn and C.L. Cooper, "*Mental Health, Job Satisfaction, Alcohol Intake and Occupational Stress among Dealers in Financial Markets*", Stress Medicine, Vol. 6, 285-298, John Wiley & Sons, 1990.

[KaKw92] B.L. Kalman and S.C Kwasny, "*Why Tanh? Choosing a Sigmoidal Function*", International Joint Conference on Neural Networks, Baltimore, MD, 1992.

[Kauf87] P. Kaufman, "*The New Commodity Trading Systems and Methods*", John Wiley, 1987.

[Kauf95a] P. Kaufman, "*Kaufman on Market Analysis - Volume*", KDY Seminars in Print, April 1995.

[Kauf95b] P. Kaufman, "*Smarter Trading*", McGrawHill, 1995.

[Keyn36] J.M. Keynes, "*General Theory of Employment, Interest and Money*", MacMillan, London, 1936.

[KiCS] K. Kiriakopoulos, N. Christofides and G. Salkin, "*A Tree Based Algorithm for Risk Management of Interest Rate Derivative Portfolios*", Centre for Quantitative Finance, Imperial College of Science, Technology and Medicine.

[KiFe94] J. Kingdon and K. Feldman, "*Genetic Algorithms and Applications to Finance*", University College London, 1994.

[King96] J. Kingdon, "*Feed-Forward Neural Networks and Genetic Algorithms for Automated Time-Series Modelling*", PhD Thesis, University College London, 1996.

[Koho84], T. Kohonen, "*Self-Organisation and Associative Memory*", Springer-Verlag Inc., Berlin, 1984.

[Koza92] J.R. Koza, "*Genetic Programming: On the Programming of Computers by Natural Selection*", MIT Press, 1992.

[Koza94] J.R. Koza, "*Genetic Programming II: Automatic Discovery of Reusable Programs*", MIT Press, 1994.

[Koza98] J.R. Koza et. al., "*Proceedings of the Third Annual Genetic Programming Conference*", University of Wisconsin, July 22-25, 1998.

[LaNa92] J.P. Lajaune and A. Naka, "*The Tokyo Spot Foreign Exchange Market Consistent with the Efficient Market Hypothesis?*", Review of Financial Economics, 2, 1, Fall 1992.

[Lang82] O. Langford, "*A Computer Assisted Proof of the Feigenbaum Conjectures*", Bulletin of the American Mathematical Society 6, 1982.

[LeBa91] B. LeBaron, "*Technical Trading Rules and Regime Shifts in Foreign Exchange*", Santa Fe Institute, 91-10-044.

[Leba93] B. LeBaron, "*The Joint Dynamics and Stability of Stock Prices and Volume*", Department of Economics, University of Wisconsin, 1993.

[Lipp87] R.P. Lippman, "*An Introduction to Computing with Neural Nets*", IEEE ASSP Magazine, pp. 3-22, April 1987.

[LoHa73] J.H. Lorie and M.T. Hamilton, "*The Stock Market: Theories and Evidence*", Homewood, IL: Richard D. Irwin, 1973.

[Lora95] A. Loraschi. et. al., "*Distributed Genetic Algorithms with an Application to Portfolio Selection*", SIGE Consulenza, S.P.A., Milano, Italy, 1996.

[Mahf95] S.W. Mahfoud, "*A Comparison of Parallel and Sequential Niching Methods*", ICGA-95.

[Malk88] B.G. Malkiel, "*Is the Stock Market Efficient?*", Princeton Financial Research Centre Memorandum, 97, October 1988.

[Malk90] B.G. Malkiel, "*A Random Walk down Wall Street*", W.W. Norton, New York, 1990.

[Mand82] B. Mandlebrot, "*The Fractal Geometry of Nature*", New York, Freeman, 1982.

[Mark52] H.M. Markowitz, "*Portfolio Selection*", Journal of Finance 7, March 1952.

[Mast93] T. Masters, "*Practical Neural Network Recipes in C++*", Academic Press, San Diego, 1993.

[McCa69] J. McCarthy and P.J. Hayes, *"Some Philosophical Problems from the Standpoint of Artificial Intelligence"*, Machine Intelligence 4, B. Metzler and D. Michie(eds.), Edinburgh University Press, Edinburgh, 1969.

[MiPa72] M. Minsky and S. Papert, *"Perceptrons"*, MIT Press, 1972.

[MiSw91] P. Miron and P. Swannell, *"Pricing and Hedging Swaps"*, Euromoney Publications, 1991.

[Morg96] *"Credit Derivatives"*, J.P. Morgan Securities, 1996.

[MoUt] J. Moody and J. Utans, *"Principled Architecture Selection for Neural Networks: Application to Corporate Bond Rating Prediction"*,

[Moze93] M.C. Mozer, *"Neural Network Architectures for Temporal Sequence Processing"*, in A.S. Weigend and N.A. Gershenfeld, *"Time Series Prediction: Forecasting the Future and Understanding the Past"*, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addision Wesley, 1993.

[NeSi76] A. Newell and A.H. Simon, *"Computer Science as Empirical Equiry: Symbols and Search"*, Communications of the ACM, Vol. 19, No.3, March 1976.

[OCPT97] M. Oussaidene, B. Chopard, O.V. Pictet and M. Tomassini, *"Parallel Genetic Programming and its Application to Trading Model Induction"*, Parallel Computing, 23, 1997, 1183-1198.

[Osbo59] M.F.M. Osborne, *"Brownian Motion in the Stock Market"*, in P. Cootner(ed.), *"The Random Character of Stock Market Prices"*, Cambridge, MA: MIT Press, 1964. (Originally published in 1959)

[OsPh95] C.L. Osler and P.H. Phang, *"Head and Shoulders: Not Just A Flaky Pattern"*, Federal Reserve Bank of New York Staff Reports, Number 4, August 1995.

[Pann] W.H. Panning, *"Birds, Bees and Bonds: Genetic Algorithms and Investment Strategy"*, MetLife Investment Corp.

[PDDC96] O.V. Pictet, M.M. Dacorogna et al., *"Genetic Algorithms with collective sharing for Robust Optimisation in Financial Applications"*, Olsen & Associates, Research Institute for Applied Economics, 1996.

[Pete91] E.E. Peters, *"Chaos and Order in the Capital Markets"*, Wiley, 1991.

[PFTV88] W.H. Press, B. Flannery, S. Teukolsky and W. Vetterling, *"Numerical Recipes in C"*, Cambridge University Press, New York, 1988.

[Pinc95] S. Pincus, *"Approximate Entropy (ApEn) as a Complexity Measure"*, Chaos, 5, 110-117.

[Pitt95] G. Pittaras, *"Pairs Trading: An Application of the Kalman Filter Methodology"*, Internal Document, European Bank for Reconstruction and Development, 1995.

[Plum96] T. Plummer, *"The Troubled Trader"*, Hambros Fund Management PLC, 1996.

[Pola71] E. Polak, *"Computational Methods in Optimization"*, Academic Press, New York, 1971.

[Prin91] M.J. Pring, *"Technical Analysis Explained"*, International Institute for Economic Research, 3rd Edition, McGraw-Hill, 1991.

[Prin93] M.J. Pring, *"Martin Pring on Market Momentum"*, International Institute for Economic Research, Braun-Brumfield, 1993.

[RaMC95] E.M. Rauch, M.M. Millonas and D.R. Chialvo, *"Pattern Formation and Functionality in Swarm Models"*, Physics Letters A, 1995.

[RZCB95] A.N. Refenes, A.D. Zapranis, J.T. Connor and D.W. Bunn, *"Neural Networks in Investment Management"*, in *Intelligent Systems for Finance and Business*, S. Goonatilake and P. Treleaven, John Wiley & Sons, 1995.

[RuHW86] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *"Learning Internal Representations by Error Propagation"*, in Parallel Distributed Processing, ed. J.L. McClelland, vol. 1, pp. 318-362, MIT Press, 1986.

[RuPM92] J.R. Rust, R. Palmer and J.H. Miller, *"Behaviour of Trading Automata in a Computerised Double Auction Market"*, Santa Fe Institute, 92-02-008.

[Sait95] A. Saitta, *"Range Breakout Trading in Treasury Bonds"*, Technical Analysis of Stocks and Commodities, April 1995.

[Salo96] *"Inside the Yield Curve"*, Salomon Brothers International Limited, 1996.

[Saun94] E.M. Saunders, *"Testing the Efficient Market Hypothesis without Assumptions"*, Journal of Portfolio Management, 20, 4, Summer 1994.

[Schw84] J.D. Schwager, *"Market Wizards"*, Harper Collins, 1984.

[Sher92] C.J. Sherry, *"The Mathematics of Technical Analysis"*, Probus, Chicago, 1992.

[Smit76] 'Adam Smith', *"The Money Game"*, Vintage Books, 1976.

[Smit93] M. Smith, *"Neural Networks for Statistical Modelling"*, Van Nostrum Reinhold New York, 1993.

[Span95] A. Spanos, *"On Theory and Testing in Economics: Modelling with Non-experimental Data"*, Journal of Econometrics, 67, 1, May 1995.

[SrDe95] N. Srinivas and K. Deb, *"Multiobjective Optimization Using Non-dominated Sorting in Genetic Algorithms. Evolutionary Computation,* v2:3, 221-248, 1995.

[Stew96] I.N.S. Stewart, *"Mathematical Clues to Nature's Rules"*, Faraday Lecture, The Royal Society, 1995.

[Stew97] I.N.S. Stewart, *"Putting Randomness in Order"*, New Scientist, 10th May 1997.

[Stoc90] D. Stock, *"Winner and Loser Anomalies in the German Stock Market"*, Journal of Institutional and Theoretical Economics, 146, 3, September 1990.

[Tan92] T.Y. Tan, *"Event Studies of Efficiency in the Australian Interest Rate Futures Market"*, Economic Record supplement, 1992.

[TCRR94] P.A.F. Turner, D. Corne, G. Ritchie and P. Ross, *"Obtaining Multiple Distinct Solutions with Genetic Algorithms Niching Methods"*, Department of Computer Science, University of Edinburgh, 1994.

[Tesf96] L. Tesfatsion, *"How Economists can get Alife"*, Department of Economics, Iowa State University, 1996.

[ThWa95] A.R. Thompson and M.J.D. Ward, *"The Johannesburg Stock Exchange as an Efficient Market: A Review"*, Journal for Studies in Economics and Econometrics, 19, 3, November 1995, p. 33-63.

[ToWh92] S. Tokinaga and A.B. Whinston, *"Applying Adaptive Credit Assignment Algorithm for the Learning Classifier System based upon the Genetic Algorithm"*, IEICE Trans. Fundamentals, Vol. E75-A, No. 5, 1992.

[TrGo92] P.C. Treleaven and S. Goonatilake, *"Intelligent Financial Technologies"*, Parallel Problem Solving from Nature: Applications in Statistics and Economics, by the Statistical Office of the European Communities, 1992.

[TsUh97] L.H. Tsoukalas and R.E. Uhrig, *"Fuzzy and Neural Approaches in Engineering"*, Wiley, 1997.

[Veri94] J. Verity, *"Silicon and Software that Mine for Gold"*, Business Week, 09-05-1994, New York.

[ViGo96] J.W. Viner and S. Goonatilake, "*A Study of a Genetic Algorithm Based Continually Adaptive Trading Engine*", Proceedings of Intelligent Technologies in the Human Related Sciences, 1996.

[WaHG95] R.F. Walker, E.W. Haasdijk and M.C. Gerrets, "*Credit Evaluation Using a Genetic Algorithm*", in *Intelligent Systems for Finance and Business*, S. Goonatilake and P. Treleaven, John Wiley & Sons, 1995.

[WaSh95] S. Ward and M. Sherald, "*The Neural Network Financial Wizards*", Technical Analysis of Stocks and Commodities, December 1995.

[WeGe93] A.S. Weigend and N.A. Gershenfeld, *"Time Series Prediction: Forecasting the Future and Understanding the Past"*, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addision Wesley, 1993.

[Weis91] S.M. Weiss and C.A. Kulikowski, *"Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems"*, Morgan Kaufmann Publishers Inc., San Mateo, California, 1991.

[Zade84] L. Zadeh, "*Making Computers Think Like People*", IEEE Spectrum, August, p. 26-32, 1984.

[Zwei86] M.E. Zweig, *"Winning on Wall Street"*, Warner Books, New York, 1986.

[ZyBK95] E.J. Zychowitz, M. Binbasioglu and N. Kazancioglu, *"The Behaviour of Prices on the Istanbul Stock Exchange"*, Journal of International Financial Markets, Institutions and Money, 5,4, 1995.