1

Introduction

Over the last decade, a considerable literature on biologically inspired algorithms (BIA) has emerged. These powerful algorithms can be used for prediction and classification, and have clear application for use in financial modelling and in the development of trading systems. Financial markets represent a complex, ever-changing, environment in which a population of investors are competing for profit. Biological entities have long inhabited such environments, and have competed for resources to ensure their survival. It is natural to turn to algorithms which are inspired by biological processes to tackle the task of survival in a financial jungle.

The primary objectives of this book are twofold: to provide readers with an up-to-date introduction to a broad range of BIAs, and to illustrate by means of a series of case studies how the algorithms can be applied for the purposes of modelling financial markets, for the prediction of corporate failure, and for the prediction of credit ratings. Although we cannot provide any guarantees that these technologies provide an easy route to financial riches, we hope this book will spark new ideas in the minds of readers to encourage them to undertake their own work in the fascinating nexus of computer science and finance.

This book is aimed at two audiences: those in the finance community who wish to learn about advances in biologically inspired computing and how these advances can be applied to financial modelling; and those in the computer science community who wish to gain insight into the domain of financial modelling and trading system design. Strong emphasis is placed in this book on evolutionary methodologies, particularly *Grammatical Evolution* [174]. This book is also suitable for use on advanced undergraduate or postgraduate courses, on quantitative finance or computational intelligence. No prior knowledge of either BIAs or financial prediction is assumed.

1.1 Biologically Inspired Algorithms

Biological systems are a notable source of inspiration for the design of optimisation and classification algorithms, and all of the methodologies in this book have their metaphorical roots in models of biological and social processes. These processes are as diverse as the operation of the central nervous system, biological evolution, the mapping of genes to proteins, the human immune system, and models of social interaction amongst organisms. BIAs do not seek to perfectly imitate the complex workings of these systems, rather they draw metaphorical inspiration from them to create mathematical algorithms which can be used in an attempt to solve hard, real-world problems, such as modelling financial markets. Figure 1.1 provides a broad taxonomy of some of the primary methodologies discussed in this book. A vast number of hybrid models which combine elements from more than one of these methodologies can also be constructed.

It is not possible to undertake a complete discussion of all of these in a single text, and we concentrate on neural network and evolutionary algorithms, while providing an introduction to BIA technologies drawn from social and immune metaphors. A brief overview of some of these technologies is provided in the following paragraphs, with a more detailed discussion of them being provided in later chapters.

1.1.1 Artificial Neural Networks

Artificial neural networks (NNs) is a modelling methodology whose inspiration arises loosely from a simplified model of the workings of the human brain. Both learn from their environment and encode this learning by altering the connections between individual processing elements, neurons in the case of the human brain, nodes in the case of NNs. NNs can be used to construct models for the purposes of prediction, classification and clustering. NNs are a non-parametric modelling tool, as the model is developed directly from the data.

1.1.2 Evolutionary Computation

Evolutionary algorithms draw inspiration from the processes of biological evolution to *breed* solutions to problems. These problems may be as diverse as determining the coefficients for a non-linear regression model, or determining the components of a financial trading system. The algorithm commences by creating an initial population of potential solutions, and these are iteratively improved over many 'generations'. In successive iterations of the algorithm, fitness-based selection takes place within the population of solutions. Better solutions are preferentially selected for survival into the next generation of solutions, with diversity being introduced in the selected solutions in an attempt to uncover even better solutions over multiple generations. BIAs that

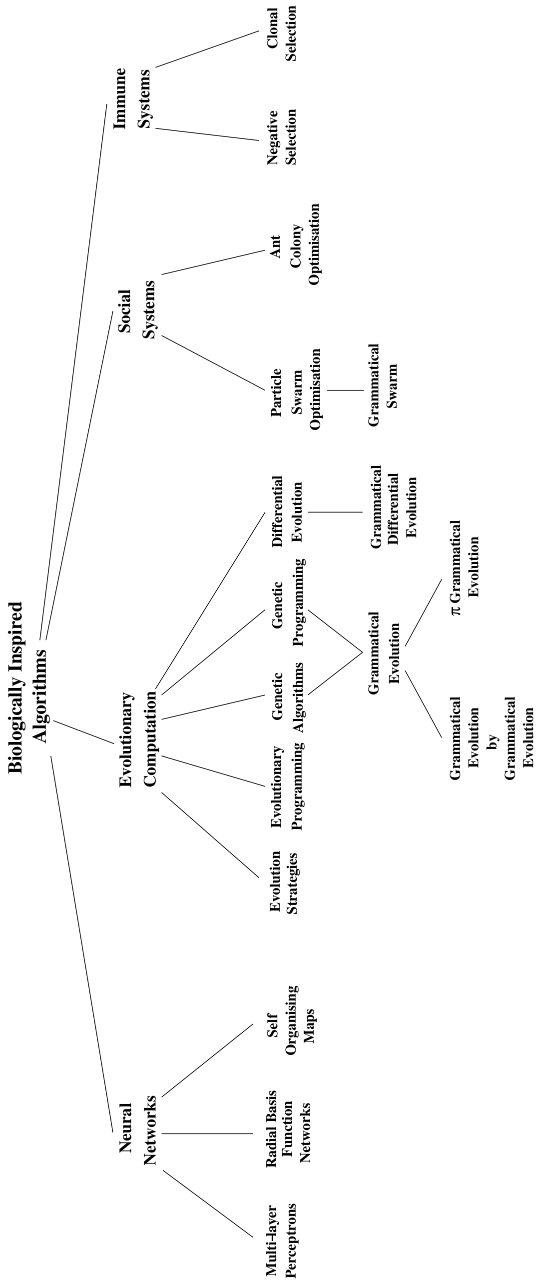

Fig. 1.1. A taxonomy of the biologically inspired algorithms that are discussed in this book

employ an evolutionary approach include genetic algorithms (GAs), genetic programming (GP), evolutionary strategies (ES) and evolutionary programming (EP).

A significant recent addition to BIA methodologies is grammatical evolution (GE), an evolutionary automatic programming methodology, which, for example, can be used to evolve rule sets or financial trading systems. GE incorporates a *grammar* which governs the creation of these rule sets. The idea of a grammar is inspired by the biological process of the mapping of genes to proteins.

1.1.3 Social Systems

The social models considered in this book are drawn from a *swarm* metaphor. Two popular variants of swarm models exist, those inspired by the flocking behaviour of birds and fish, and those inspired by the behaviour of social insects such as ants. The essence of these systems is that they exhibit flexibility, robustness, self-organisation, and communication between individual members of the population. The swarm metaphor has been used to design algorithms which can solve difficult problems by creating a population of problem-solvers, and allowing these to communicate their relative success in solving the problem to each other. Higher-performing individuals attract the attention of others, who test variants on their problem-solving strategy in an attempt to improve it.

1.1.4 Artificial Immune Systems

The human immune system is a highly complex system, comprised of an intricate network of specialised tissues, organs, cells and chemical molecules. The capabilities of the natural immune system are to recognise, destroy and remember an almost unlimited number of foreign bodies, and also to protect the organism from misbehaving cells in the body. To assist in protecting the organism, the immune system has the capability to distinguish between *self*, and *non-self*. Artificial immune systems (AIS) draw inspiration from the workings of the natural immune system to develop algorithms for optimisation and classification. Practical applications of AIS models to pattern-recognition tasks include the identification of potentially fraudulent credit card transactions, the identification of the 'state' of the stock market, and the identification of financially at-risk companies.

1.2 Computer Trading on Financial Markets

Computerised or automated trading on financial markets is not a new phenomenon. Computers have been used for *program trading* for many years. In program trading, the object is usually to uncover and eliminate anomalies between financial derivatives and the underlying financial assets which make up those derivatives.[1] A typical example of program trading is *index arbitrage* which involves the automated purchase or sale of a basket of stocks which make up a market index, in conjunction with the simultaneous sale or purchase of a derivative product such as stock index futures, in order to profit from the price difference between the basket and the derivative product. In theory the transaction generates risk-free returns, but in practice it relies on estimates of dividend income from companies, an estimate of the rate of return available on invested dividends, and the ability of the computer to make the purchases/sales at the prices which produced the arbitrage opportunity. Program trading accounts for a considerable portion of trading on major stock exchanges. For example, it is estimated that program trading volume accounted for approximately 50.6% of the total trading volume on the New York Stock Exchange (NYSE) in 2004 [163].

A second, less publicised use of computers is to construct trading systems which assume trading risk in the search for superior, risk-adjusted, returns. These systems are the focus of interest of several of the case studies in this book.

1.3 Challenges in the Modelling of Financial Markets

Modelling of financial markets is challenging for several reasons. Many factors plausibly impact on financial markets including interest rates, exchange rates, and the rate of economic growth. We have no hard theory as to how exactly these factors effect prices of financial assets, partly because the effects are complex, non-linear and time-lagged. For example, a change in interest rates may impact on the foreign exchange rate for a currency, in turn effecting the level of imports and exports into and from that country. Another difficulty that arises in financial modelling is that unlike the modelling of physical systems we cannot conduct controlled experiments. Only one sample path through time is available for our examination, as we only have one history of market events. Additionally, some factors which can effect financial markets are inherently unpredictable such as earthquakes, the weather, or political events. Taken together, these difficulties imply that our ability to predict market movements will always be imperfect.

[1] A derivative is a financial instrument whose value is based on that of another financial instrument such as a share. For example, investor A may sell an option on a share to investor B. This option gives investor B the right to buy (or sell) that share to investor A, at a specified price for a specified time. As the value of the underlying share changes, the value of the financial derivative (the option) will also change.

1.3.1 Do Prices Follow a Random Walk?

The very attempt at modelling financial markets for profit meets with the scorn of some financial economists. Two pillars of traditional financial economics are that market prices of financial assets follow a *random walk* and that markets are *efficient*. One of the earliest studies suggesting that prices in markets might follow a random walk was undertaken by [123]. The traditional definition of a random walk is a process in which the changes from one time period to the next are independent of each other, and are identically distributed. If prices in financial markets did follow a random walk, this would imply that the size and direction of a past change in price provides no insight into the size and direction of the next change in price of that financial asset. In other words, there would be no auto-correlation in the time-series of prices from a financial market.

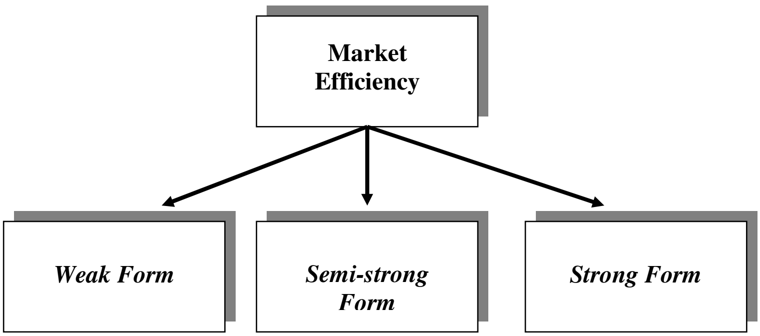

Fig. 1.2. Three forms of market efficiency

Closely related to the concept of a random walk (and often confused with it) is the efficient market hypothesis (EMH). Although a random walk in share prices could arise for a variety of reasons, it is consistent with a proposition that current prices fully reflect the market's aggregate assessment of any existing information which could impact on the price of a financial asset. If a market is informationally efficient, in that all information is impounded accurately and instantly into prices once it becomes available, then there is no scope to make excess returns from trading on such information. The more information-efficient a market is, the more random the sequence of price changes that it will produce, as prices will only alter when new information emerges. As the nature of new information is by definition unpredictable, a share's price is equally likely to rise or fall in the future. Three versions of market efficiency were posited [73] (Fig. 1.2).

Under weak form efficiency it is considered that the price of a share at any point in time reflects *all* the information contained in its price history. This would imply that excess risk-adjusted returns cannot be obtained by attempting to construct a model which uses information on past share prices, or past transaction volumes, to predict future share prices. The semi-strong form of efficiency suggests that a share price at any point in time reflects all publicly available information which could impact on the share's price. The strong form of market efficiency implies that share prices include all information relevant to the price of a share, including both public and private (non-public) information.

1.3.2 Attack of the Anomalies

If the semi-strong form of the EMH was correct, there would be no scope to construct a model of a financial market using publicly available information, which would generate excess risk-adjusted returns. In spite of the initial research which lent broad support to the EMH, there is a growing body of research in more recent times which suggests that subtle patterns do exist in time-series of financial asset prices, and that these prices do not follow a random walk [143, 144]. Among the anomalies that have been noted are the existence of serial correlation in weekly and monthly stock returns, particularly for small capitalisation (small company) stocks. Generally, three patterns of serial correlation in stock returns are recognised: short-term reversals (looking at returns over a few weeks), medium-term inertia, and longer-term reversals.

A considerable body of empirical evidence suggests that short-run volatility in share returns is clustered. A large change in price tends to be followed by another large change in price, but the direction of this change is difficult to predict. In other words, prices tend to be volatile when they have just been volatile, leading to patterns of short-run price reversals. Under medium-term inertia, good (or bad) performance of a stock over 3-12 months is typically indicative of continued good (or bad) performance in the next few months. This could provide scope for the implementation of *momentum* investment strategies, where investors seek to buy (sell) shares which have recently trended upwards (downwards). Over longer time periods (3-5 years), there is evidence of negative serial correlation in share returns, suggesting that stocks that have performed well over the last several years are more likely to under-perform in the future. This lends support to the common idea of *contrarian* or *value* investment strategies where investors buy out-of-favour stocks (those whose share price has underperformed that of their peers in recent years). A posited explanation for this negative serial correlation is the *over-reaction hypothesis* that investors are subject to waves of optimism and pessimism which cause prices to swing temporarily away from their underlying value for individual firms or whole sectors [53, 63]. Other examples of asset-price anomalies are reported in [37] and [54].

Interpretation of the results of studies reporting anomalies has been controversial [74, 75], but they are consistent with a hypothesis that market efficiency is a relative term. Under this premise, as market participants uncover new information processing mechanisms (such as BIAs), market efficiency is enhanced as market participants apply the new information processing methodology. It is plausible that novel, powerful computational techniques which can uncover new price-relevant information could prove profitable. However markets represent a competitive, adaptive environment and are likely to rapidly impound available information into asset prices. Just as $100 bills do not last long on the sidewalk, traders who constitute markets have a vested interest in searching for and exploiting any edge which could lead to profit. Financial modelling in such an environment can be compared to an arms race whereby each player rapidly cancels out any advance of another. The advantage offered by a new modelling technique is therefore likely to be short-lived.

There are close parallels between the challenges of the environment of financial markets, and those from which several of the biologically inspired algorithms discussed in this book are drawn. Evolutionary algorithms, swarm algorithms and immune algorithms are drawn from environments where, just as in financial markets, there is continual adaptation and where there is competition for resources between individuals.

1.4 Linear Models

The goal in modelling a system is typically to gain insight into its behaviour, to determine which factors impact on the output of the system, and to determine how influential each of these factors is. A second goal is to enable prediction of the future output of the system under different conditions. A simple linear model has the general form:

$$Y = \alpha + \beta_1 X_1 + \ldots + \beta_n X_n \tag{1.1}$$

where Y is the dependent variable, $X_1, \ldots, X_n$ are independent (or explanatory) variables (in a simple model there may be only one independent variable, in a multiple regression model there will be several), $\beta_1, \ldots, \beta_n$ are regression coefficients, and α is a constant which allows the model to produce a value for Y even if all the dependent variables have a zero value.

In constructing a linear model, the first step is typically to posit a *cause-and-effect* relationship based on prior theory or intuition between the dependent variable and one or more explanatory variables (or inputs). In other words: 'I think x and y impact on the value of z'. Two questions then spring to mind:

i. Is the assumption that x and y effect z supported by empirical evidence?
ii. How great is the effect of x and y respectively on z?

To answer these questions, we can collect sample data, vectors of explanatory variables and the associated value of the dependent variable, and then either manually or using a computer package determine the values for the regression coefficients which produce a model whose output closely agree with actual known output for each vector of input data. If values can be found for the regression coefficients such that the linear model is successful in explaining a high portion of the variation in the dependent variable (z), where the signs (+ or −) of the regression coefficients concord with theory/intuition, we consider that the model is good, and that our hypothesis that x and y impact on z is plausibly supported by the collected data.

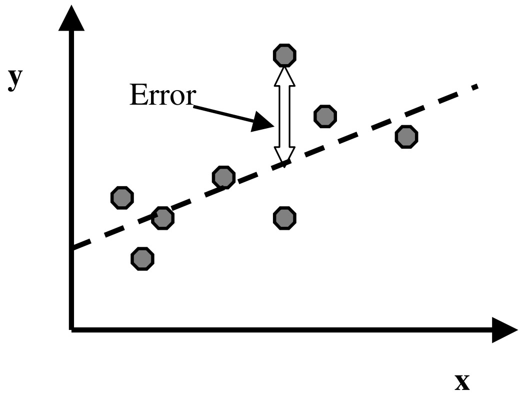

Fig. 1.3. The least-squares regression line is constructed by minimising the sum of the squared errors for each data-point. The error for each point is the difference between its actual value of y, and the predicted value of y according to the regression line

The Error Measure

To determine the values of the regression coefficients, the modeller must define an error measure so that the error between the model's predicted output and the actual output value can be calculated. This error is then used as feedback to alter the regression coefficients in order to reduce the error measure. Typically in basic regression models the goodness-of-fit measure is the sum of the squared errors between the predicted and the actual outputs (Fig. 1.3). If we are willing to make assumptions concerning the distribution of the error terms resulting from the model's predictions, i.e. that the errors

- have a mean of zero,
- are normally distributed,
- are independent, and
- have constant variance,

then a series of statistical statements can be made including the construction of confidence intervals for the model's predictions and for the values of the regression coefficients. One issue of particular interest is whether the regression coefficients are significantly different from zero.

Modelling with Biologically Inspired Algorithms

In applying the various forms of biologically inspired algorithms, we are undertaking the same basic modelling process, although the actual mathematical form of the resulting model may be considerably more complex than that of the simple linear regression model. If we want to predict a future share price or other financial variable, can we identify plausible sets of explanatory (input) variables based on theory or intuition? If so, we can test our hypothesis that there is a link between the explanatory and dependent variables by using historical market data.

Although the choice and implementation of modelling methodology (linear regression, artificial neural networks, etc.) can play an important role in determining the quality of the final model, it is only one component of the modelling process. Other vital decisions faced by the modeller address questions such as:

- What data should be used to construct the model?
- Is the cause-and-effect relationship plausible?
- Does this data need to be preprocessed before it is included in the model?
- What error measure is appropriate?

No modelling methodology will compensate for poor decisions in these steps, and each of these issues will be discussed in later chapters.

1.5 Structure of the Book

The remainder of this book is divided into three parts. In Part I a range of biologically inspired algorithms are introduced and explained. These offer the potential to develop useful financial models. However, despite the undoubted power of these algorithms, their successful implementation requires the careful selection of explanatory variables, and the careful validation of the results arising from the developed models. Therefore the book contains a section on model development (Part II), which covers a range of practical issues which arise in the creation of financial models. Finally, in Part III, a series of case studies are provided to illustrate several potential financial applications of biologically inspired methodologies. A number of these cases concentrate on the construction of trading systems in equity and foreign exchange markets. The utility of the methodologies is further demonstrated through their application to a range of other tasks, including the prediction of corporate failure, and the prediction of corporate bond ratings.

http://www.springer.com/978-3-540-26252-7

Biologically Inspired Algorithms for Financial Modelling
Brabazon, A.; O'Neill, M.
2006, XV, 277 p., Hardcover
ISBN: 978-3-540-26252-7