Politecnico di Milano

Scuola di Ingegneria Industriale e dell'Informazione
Corso di Laurea Magistrale in Computer Science and Engineering

Surveying Financial Markets Prediction

A Focus on Genetic Programming Applications

Relatore:

Prof. Pier Luca Lanzi

Correlatore:

Dott. Danilo Attuario

Tesi di Laurea di:

Paolo Antonini

matricola 858242

Anno Accademico 2016–2017

IF I HAVE SEEN FURTHER

IT IS BY STANDING ON YE SHOLDERS OF GIANTS.

ISAAC NEWTON, 1675

Contents

List of Figures

List of Tables

Abstract

In this survey work, we explored the possibility of predicting financial markets using machine learning techniques, with a focus on the application of genetic programming (GP) to the definition of profitable trading strategies for stock prices, stock indexes, and foreign exchange rates (FX rates). Financial markets are complex systems, the study of which has spanned over the years, both from a theoretical standpoint and from a more practical perspective. While many, in the theoretical discussion, support the essential unpredictability of financial markets, others suggest that stock prices and FX rates are in fact predictable, and developed analytical models to prove so. Building mainly on technical analysis, the study of historical prices in the belief that patterns do tend to repeat themselves, computer scientists over the years applied different machine learning techniques to produce relevant predictions and strategies. Based on our research, the machine learning techniques that most commonly are applied to make daily forecasts are GP and artificial neural networks (ANNs), with fuzzy logic-based methods gaining popularity in recent years. Results of the different techniques are comparable, in that they are mixed and inadequate to establish a clear outperforming method. We selected GP applications to focus on. In most research papers that we reviewed, GP is used as the search algorithm to find profitable trading strategies in stock and foreign exchange markets. We present and discuss the different applications from a number of different perspectives, starting from the motivations that guided researchers to adopt GP, then moving to the definition of inputs and building blocks of the GP algorithm, and concluding with the solutions adopted to deal with the issue of adapting the trading models to the dynamic conditions of financial markets. GP appears to be a relevant approach to the prediction of financial markets, with clear advantages over other methods. The research conducted so far is rather solid, and the open issues and proposed future developments well-defined.

Sommario

In questo lavoro, esploriamo la possibilità di prevedere l'andamento dei mercati finanziari usando tecniche di *machine learning*, presentando una rassegna della letteratura pertinente; ci concentriamo, in particolare, sull'applicazione di *genetic programming* per la definizione di strategie di investimento profittevoli che riguardino prezzi di azioni, indici azionari e tassi di cambio tra valute. I mercati finanziari sono sistemi complessi, lo studio dei quali si è sviluppato nel corso degli anni sia da un punto di vista strettamente teorico, sia da una prospettiva più applicata. Da una parte, molti sostengono la sostanziale impredicibilità dei mercati finanziari, dall'altra alcuni suggeriscono che prezzi e tassi possano essere effettivamente previsti, e hanno sviluppato modelli analitici per dimostrarlo. Partendo dall'analisi tecnica (*technical analysis*), cioè dallo studio delle serie storiche dei prezzi nella convinzione che alcuni pattern si ripetano nel tempo, tecniche di *machine learning* sono state impiegate per produrre previsioni e strategie rilevanti. Sulla base della nostra ricerca, le tecniche principalmente utilizzate per le previsioni giornaliere sono *genetic programming* e reti neurali, con metodi basati su logica *fuzzy* che hanno acquistato popolarità negli ultimi anni. Le differenti tecniche forniscono risultati comparabili, positivi solo in parte e tendenzialmente inadeguati a stabilire quale sia il migliore approccio da utilizzare. Abbiamo scelto di focalizzarci sulle applicazioni di *genetic programming*. Nella maggior parte dei documenti di ricerca analizzati, *genetic programming* viene utilizzato per trovare strategie di investimento favorevoli nel mercato azionario e monetario. Discutiamo le differenti applicazioni da diversi punti di vista: le motivazioni che hanno guidato la scelta dei ricercatori di adottare *genetic programming*, la definizione dei dati in ingresso e delle primitive dell'algoritmo di ricerca genetica, e le soluzioni adottate per risolvere il problema di adattare i modelli di previsione alle condizioni dinamiche dei mercati finanziari. Riteniamo che *genetic programming* sia un metodo di previsione adatto allo scopo. La ricerca condotta finora è abbastanza solida, e i punti ancora aperti e le proposte di ulteriore sviluppo ben definiti.

1
Introduction

One of the earliest and most enduring models of the behaviour of stock prices is the Efficient market hypothesis (EMH), whereby price changes are essentially unpredictable, as they fully incorporate the expectations and information of all market participants. What proponents of the theory suggest is that investors aggressively take advantage of even the smallest informational advantages at their disposal, and in doing so, they incorporate such information into market prices, thus quickly eliminating the profit opportunities that gave rise to their aggression. Therefore, prices always fully reflect all available information and no profits can be achieved from information-based trading. A long-standing example of economic logic illustrates the belief of those who support EMH, and it goes like this. An economist and a companion (Lo and MacKinlay, 1999, chapter 1), or otherwise a finance professor and a student (Malkiel, 2003), come upon a \$100 bill lying on the ground. As the student reaches down to pick it up, the professor says, "Don't bother: if it were a real \$100 bill, someone would have already picked it up". The \$100 bill is an illustration of the (supposedly) unattainable returns.

On the opposite side, those who reject EMH (i.e., those who believe that the \$100 bill is actually there and one can pick it up) maintain that markets are predictable, at least up to a certain degree, and that above-average, risk-adjusted returns can be earned. To this end, academics and financial practitioners developed different approaches to make such prediction possible. Some suggest that the intrinsic value of a company may not be fully reflected in the prices of the company's stock: any such deviation, if detected, can be a source of consistent earnings (fundamental analysis). Others believe that stock prices repeat themselves in patterns over time, and developed mathematical and graphical instruments to capture those patterns (technical analysis). Still others integrate principles of psychology and sociology in the analysis of financial markets, in the belief that market agents are not as rational as it may be assumed, and that their behaviour directly affects market moves (behavioural analysis).

Our work is part of the line of research that believes financial markets to be, in fact, predictable using quantitative analysis of past prices. The approach, therefore, is that of technical analysis. Computer science, and in particular machine learning, greatly improved the possibilities of researchers to study financial data series and derive models with some degree of forecasting ability. Modern studies make extensive use of evolutionary (e.g., genetic programming) and soft computing (e.g., artificial neural network and fuzzy logic-based systems) techniques, and sometimes their integration, to produce relevant and profitable trading models. Most commonly, all such techniques receive as inputs and process past prices and trading volume of the stock under investigation (or stock index, or foreign exchange rate), and return either a forecast of future prices, or a trading strategy (intended as a sequence of buy and sell signals) that can be profitable on the market. In particular, our goal in this document is to survey the research work made so far in the field, by reviewing the motivations, results, and successes of each different machine learning techniques. To facilitate the review work, we follow the classification of studies proposed by Park and Irwin (2007). Standard and model-based bootstrap studies are particularly important for their establishing a new, modern method of analysis, based on more rigorous parameter optimisation and statistical testing. Then, genetic programming, artificial neural networks, and fuzzy logic-based methods are analysed as state-of-the-art forecasting techniques.

After noting that genetic programming (GP) and GP-related literature have some advantages on the other methods, ranging from the interpretability of the candidate models to a more structured and approachable *corpus* of research papers, we focus our attention on the adoption of such evolutionary approach to the issue of predicting stock prices, stock indexes, and foreign exchange rates (FX rates). In our focus, we attempt to capture and discuss both classical and modern studies that make use of genetic programming (GP) to obtain profitable trading models for stock and foreign exchange markets.

The present document is structured as follows. In chapter 2 we outline the financial framework all reviewed studies deal with, by further explaining the concept of Efficient market hypothesis (EMH) and the three already mentioned analytical models. Chapter 3 is a compact survey of current literature in the field of financial markets prediction, using machine learning techniques; as it was already mentioned, we present each class of approaches separately, for the sake of an orderly and structured discussion. Then, in chapter 4, we focus our attention on the applications of genetic programming (GP) to financial markets prediction, and discuss the main takeaways, as well as the proposed future developments, in chapter 5. The present work is concluded by chapter 6.

2

Theoretical Economic Framework

The prediction of financial markets has been attracting scientific interest for a long time. Predictability itself has been discussed by academics from a theoretical as well as practical standpoint. The most common targets of scientific enquiries are stocks prices, stock indexes, and foreign exchange rates (FX rates), for the reasons that we propose in section 2.1. Next, in section 2.2 we briefly describe how the stock market works, to give more comprehensive a view of the framework the scientific research works in. Finally, in sections 2.3 and 2.4 we describe the most relevant achievements in the analysis of stock prices and FX rates, from a theoretical perspective: namely, the Efficient market hypothesis (EMH), a longstanding economic theory by which the stock market is essentially unpredictable, and the three analytical models that have been proposed by academics and practitioners who reject the EMH.

2.1 Classification of Financial Instruments

Financial instruments comprise the whole range of financial contracts made between two institutional units. More specifically, a financial instrument, or asset, is a store of value, over which ownership rights are enforced and from which their owners may derive economic benefits by holding or using them over a period of time. Multiple different categorisations of financial instrument have been proposed over time, based on the prevailing needs, national regulations, and scope of application. A relevant classification is offered in *IFRS 9* (2014) by International Financial Reporting Standards Foundation (IFRS Foundation), a prominent international standard-setting body. However, this document is meant as a technical regulatory framework for accountants, which makes its content too specialised for our goals.

The document that better serves our purposes is *Balance of payments and international investment position manual* (2008, chapter 5), by the International Monetary Fund (IMF). The main goal of this document is to promote cross-country comparability of data and analyses in financial statistics, by providing and explaining concepts and definitions. Three broad categories of instruments are identified (table 2.1), based on the legal characteristics of the relationship between the involved parties:

- *Equity and investment fund shares*: their distinguishing feature is that the holders own a residual claim on the assets of the institutional unit that

Categories	Types
Equity and investment fund shares	Equity Investment fund shares
Debt instruments	Special drawing rights Currency and deposits Debt securities Loans Insurance, pension, and guarantee schemes Other accounts receivable/payable
Other financial assets and liabilities	Monetary gold Financial derivatives and employee stock options

Table 2.1: Classification of financial instruments, in accordance with International Monetary Fund.
International Monetary Fund, *Balance of payments and international investment position manual* (2008)

issued the instrument; roughly said, this kind of instruments represents fractional ownership of a corporation. Prominently, shares (also called *stocks*) fall into this category.

- *Debt instruments*: these instruments are those that require the payment of principal, possibly with an accrued interest, at some point in the future. The amount to pay is usually set according to a predefined formula, which means that the creditor has a more limited risk exposure. Relevant representatives of this category are bonds, but also currency, loans and insurance schemes fall in this class.

- *Other financial assets and liabilities*: these are instruments that cannot be otherwise categorised. An example are financial derivative contracts, that are financial instruments linked to another specific financial instrument, indicator, or commodity (commonly called *underlying asset*, or simply underlying) and through which specific financial risks can be traded in their own right in financial markets. Transactions and positions in financial derivatives are treated separately from the values of any underlying items to which they are linked. The best known derivative instruments are options and futures: but for some minor differences, in both cases, the two counterparts agree to trade a specified amount of an underlying item (financial or non-financial) at some point in the future and at an agreed-upon price.

A few distinguishing characteristics, that prompt the choice of researchers to focus on stock prices, should be evident from the previous brief description. The composition of investment funds is generally complex, and occasionally not fully or timely disclosed, resulting in major difficulties in the prediction of their performance. The only possible exception are those funds that are specialised at closely tracking some large stock market index, like the Standard & Poor's 500 (S&P 500) and the Dow Jones Industrial Average (DJIA), which means replicating as tightly as possible the performance of the benchmark index.

Debt instruments are largely fixed or connected through a formula to some other variable, such as a market interest rate or the price of a selected item. The result is that the instrument behaviour is easily predictable, should the related variable be known: the challenge of forecasting is simply shifted. The same applies to financial derivatives: their value is associated, by nature, with the one of a specified underlying entity. Like in the case of debt instruments, it only makes sense to predict the value of the underlying (if at all possible).

In light of the above, researchers focused their attention on investigating the predictability (and thus on the prediction) of stock prices and stock market indexes. Probably due to their importance at a macroeconomic level, FX rates have been prompting almost the same degree of scientific interest as stock price prediction.

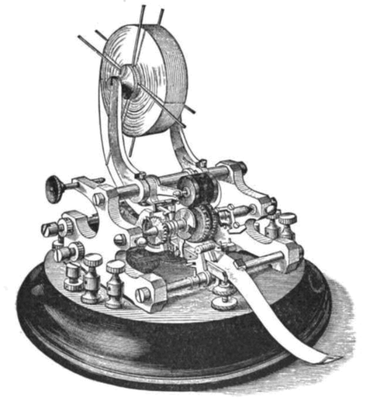

Figure 2.1: Drawing of Edison's stock ticker tape machine, invented by Thomas Edison in 1869. In use between around 1870 through 1970, ticker tape machines were used to transmit stock price information over telegraph lines. It consisted of a paper strip that ran through a machine, which printed abbreviated company names as alphabetic symbols (*tickers*) followed by numeric stock transaction price and volume information.
W. Maver. 1892. *American Telegraphy: Systems, Apparatus, Operation,* though https://commons.wikimedia.org/wiki/File:Edison_stock_ticker.png

2.2 *Stock Market Description*

The detailed characteristics and legal aspects of stocks and trading are far beyond the scope of this work; nevertheless, it is apt to spend some words on the main mechanisms of stock markets. In general, the shares of a company may be sold by shareholders to other parties; marketplaces for trading shares, as well as other financial products, are set up by stock exchanges. In order to have their stocks traded, a company may list its shares on an exchange, in their own country, or elsewhere (most commonly, the US). A unique alphabetic symbol, called *ticker*, is used to identify the company in the exchange market[1].

1. For example, Apple Inc. is identified as `AAPL`, and Amazon.com is characterised by `AMZN` ticker symbol.

The price at which stocks are traded is strictly the result of the balance of demand and supply forces, like any other material or immaterial good. The supply is the number of shares offered for sale at any one moment, while the demand, by contrast, is the number of shares investors wish to buy at that same time. The price of the stock moves in order to achieve and maintain equilibrium. Although theories and models of price determination have been discussed for long, the matter has not been settled yet.

As the price fluctuates during the trading day, stock market data are typically timestamped records of the opening, highest, lowest, and closing prices of the stock over a given period of time (usually a day, but the periodicity may be much tighter or looser); another kind of data typically provided is volume, that is the number of stocks traded in the period. Since companies may perform financial operations that directly and drastically change the price of a stock, then it is common for data providers to retroactively compute adjusted prices, to make up for such operations. A company may perform two different actions to directly change the price of a stock, based on a decision by the board of directors: paying dividends or splitting stocks. The payment of dividends is the distribution of a portion of the company's earnings to shareholders. Dividends are commonly paid in the form of cash distributions on a periodic basis, according to the number of stocks each shareholder owns[2]. After the dividends are distributed to shareholders, the price of each stock is reduced by the amount of the dividend itself. A stock split, instead, happens when a company increases the number of shares that are outstanding on the market, by issuing more shares to current shareholders. To keep the market capitalisation constant, the stock price is adjusted accordingly: in a 2-for-1 stock split, for example, an additional share is given for each share held by a shareholder, and the share price is halved. A stock split is usually done by companies that have seen their share price increase to levels that are either too high or beyond

2. For example, on 1st August 2017 Apple Inc. issued a dividend payment for an amount of \$0.63 per share. An investor who owned 100 `AAPL` shares at the time of issue, earned \$63.00 in total.

the price levels of similar companies in their sector; the primary motive is to make shares more affordable to investors.

A trader who wants to invest in a company's stocks can perform two different trading actions. When a trader buys an amount of a stock, she is 'opening a long position'[3], in the hope that its price will raise, and she will be able to sell those shares at a higher price, making profit. The opposite trading action is short selling ('short position'), where an investor borrows some shares from a lender and sells them to another party. In case the stock price falls, she will be able to buy it back and return it to the lender to *cover* the short sale, thus again making profit on the difference between the price at which the trader sold the asset and the price at which she bought it back.

3. In finance, a position is the amount of a security, commodity or currency in which a physical person or legal entity has direct financial interest.

Similar mechanisms regulate also the foreign exchange market (FX market), where currencies are traded. In a typical foreign exchange transaction, a party purchases some quantity of a currency by paying with some quantity of another currency; in addition, derivative contracts are possible. Like in the stock market, the shifts in the FX rates make profits possible to traders.

2.3 *Efficient Market Hypothesis*

The predictability of stock prices has been widely discussed in the academic literature, both from a theoretical standpoint, with the development of the principles and models presented in this chapter, and from a more practical and technical perspective, as we will see in the next chapters of this document. Although it is known that stock prices are the result of supply and demand equilibrium, little can be said about the actual price determination, that is how people decide the maximum price at which they are willing to buy or the minimum at which they are willing to sell. Two contrasting views stand in the field: the one of proponents Efficient market hypothesis (EMH), and that of its opponents.

In a seminal paper on the topic, Fama (1970) defined a market to be efficient when 'prices always fully reflect available information', as a consequence of the full rationality of all market agents. In other words, financial assets are always priced correctly, given what is publicly known, at all times. This results in stock prices approximately describing a random walk through time: in response to the arising of new information, price changes happen immediately and independently of any past variation[4], and therefore the best possible prediction of future returns is nothing more than the historical mean. By the end of the twentieth century, the idea of efficiency has been relaxed, to allow for anomalies and 'bubble' periods. Nevertheless, even when this is the case, Malkiel (2003) argues that there is no way in which investors can reliably exploit any anomalies or patterns that might exist, because they can be detected only after the fact. Malkiel also suggests that, even in the case where stock markets are not fully efficient and prices follow a mathematically perfect random walk, such deviations are so small that any excess return is unlikely, at the least: statistical significance and economic significance, in his view, should be decoupled.

Eugene F. Fama, along with Lars Peter Hansen and Robert J. Shiller, is 2013 Nobel Prize winner in Economic Sciences for his empirical analysis of financial asset prices. `https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/`

4. Hence the name of the theory: stock markets are extremely efficient in reflecting information about individual stocks and about the stock market as a whole.

On the other side, different authors question the idea of market efficiency, on theoretical and empirical grounds. The core assumption of EMH is the rationality of market agents, whereby market participants make unbiased forecasts about the future and make decisions that balance the attractiveness of economic opportunities and the presence of risk. Thaler (1999) dismisses such assumption as unrealistic, at the least. In multiple works, Shiller (2003; 2015, chapter 9) points out that, even though EMH cannot be expected to be so egregiously wrong that immediate profits should be continuously available, nevertheless such theoretical model has its place as illustration of an ideal world, and cannot be held in its pure form as accurate descriptor of actual markets: speculative bubbles are there to prove so. Thaler (1999) goes even further, proposing a simple argument against the efficiency of stock markets: the very high volume of shares traded daily cannot be justified in a world where traders are rational and mutually aware of their rationality. The rationale is quite elementary. Why should an investor A buy some shares of some company, when there is an investor B ready to sell them? Does either A or B know some piece of information, unbeknown to the counterpart? Therefore, in a fully rational world, significantly fewer transactions should take place.

2.3.1 *Efficiency of Foreign Exchange Markets*

The efficiency of FX markets has been debated in a similar way. Meese and Rogoff (1983a, 1983b) advance evidence that the predictability of FX rates is severely limited, and that a random walk model performs as well as any estimated model in out-of-sample forecasting. In later works, Diebold and Nason (1990) explain that, while statistically significant rejections of a random walk model may routinely occur and can be identified in an *ex post* analysis, in out-of-sample forecasting a random guess cannot be improved upon, at least in the short term. LeBaron (1999) discusses evidence and extent of FX rates predictability in connection to central banks activity, which may actually introduce noticeable trends into the evolution of FX rates and create opportunities for market participants to profit: however, the results show that FX rate predictability in the United States is dramatically reduced, in those periods when the Federal Reserve is not intervening actively in the market.

In contrast, some researchers are able to provide evidence of some form of FX rate predictability, thus rejecting the random walk hypothesis. Kilian and Taylor (2003) detect nonlinear dynamics in FX rates, that allow dismissing the random walk model, but only at horizons of 2 to 3 years, while at a shorter term the predictability is lost. Neely, Weller and Dittmar (1997) found consistent evidence that, by using genetic programming over the period 1981-1995, FX rates could be predicted and generate economically significant out-of-sample returns. However, by admission of the authors, their results do not represent a compensation for the systematic risk a trader should bear.

2.4 *Financial Markets Analytical Models*

Traditionally, those who reject EMH rely upon two basic approaches, to make relevant decisions about their investments and the actions to undertake: fundamental analysis and technical analysis. Fundamental analysis is the analysis of a business's financial statements and market conditions, in the attempt to identify possible profitable discrepancies between the 'true' value and the actual value of a company's stocks. Technical analysis, instead, assumes that a security's price already reflects all publicly available information, and instead focuses on the statistical analysis of price movements, which are deemed to cyclically repeat their past patterns. Although the tenets of EMH put in doubt the validity of both, practitioners make use of the two, oftentimes integrating the results. More recently a new approach, called behavioural analysis, was introduced. This kind of analysis tries to bind together finance and principles of psychology and sociology, in the belief that sentiment and mood of market actors (i.e., investors) have a direct and significant impact on market trends.

2.4.1 Fundamental Analysis

Fundamental analysis is arguably the most traditional approach, common among scholars (Lo, Mamaysky and Wang, 2000) as well as practitioners (Malkiel, 2012, chapter 5). The main goal of fundamental analysis is to assess the proper value of a company's shares, based on the evaluation of underlying factors that affect the actual business of the company itself. Ou and Penman (1989) explain the foundational assumption of fundamental analysis: stock prices imperfectly reflect the value of firms (the so-called *fundamentals*, or *intrinsic value*). This results in stock prices continuously giving an overestimation or underestimation of the true value of a company; nevertheless, stock prices do have a tendency to slowly gravitate towards the fundamental values, so the investor who detects any such deviation may make profit.

The process of fundamental analysis is based on the evaluation of a wide range of information sources, both quantitative and qualitative. Financial statements from the company itself, such as balance sheets and income statements, give factual insight on the performance of a firm, and thus on its value. In addition, qualitative characteristics of the company are to be taken into account as well, because an assessment of the business model, or of the corporate governance and procedures, is likely to give hints on the future evolution of a company in the market. A thorough fundamental evaluation should include also the study of industry and market conditions, because those are the factors that may be operative on the environment the company works in. Macroeconomic factors that are commonly considered are, for example, competition degree and market share of the company, and economic indicators such as inflation and FX rates.

Financial statement analysis, understandably, is slow and difficult to automatise; moreover, the insights generally give a long-term outlook, rather than a quickly exploitable result. In contrast, measures of the global state of the economy are more easily accessible, through proxies like commodity prices (Kilian and Park, 2009) and FX rates.

2.4.2 Technical Analysis

In literature, little consensus has been reached about the very definition of technical analysis. Murphy (1999, chapter 1) defines technical analysis as the study of past market action (which, in the stock market, includes price and volume movements), primarily through the use of charts, for the purpose of forecasting future price trends. Lo, Mamaysky and Wang (2000) make a distinction between technical analysis itself and quantitative finance: while the former employs the tools of geometry and pattern recognition[5], the latter relies on the more rigorous principles of mathematical analysis and statistics. Park and Irwin (2007) take a midway position, stating that technical analysis includes a variety of techniques ranging from strictly mathematical approaches to visual chart pattern recognition.

5. To the point that it is called *charting*, sometimes with a non-negligible degree of contempt and mockery.

What all such viewpoints do have in common is the opinion that past market actions can be used to predict future trends; the main differences, instead, stand in the methods that practitioners use to make their predictions. Most commonly, the tools in the hands of a technical analyst are technical indicators, which are metrics derived from market activities of a stock (or any other financial asset). Based on generally simple functions of historic prices and volume, technical indicators provide an insight into the current state of the market, for example by evidencing a situation where the stock under examination is overbought[6], and a price drop is to be expected.

6. Overbought refers to a situation in which the demand for a certain asset pushes its price to levels that are not justified by its intrinsic value. When an indicator suggests an overbought situation, it is generally interpreted as a sign that the price of the asset is becoming overvalued and may experience a drop. The exact opposite situation is when an asset is oversold: its price becomes undervalued, and a price raise should be expected.

A thorough survey conducted by Park and Irwin (2007) shows how 56 among a total of 95 reviewed studies, published in the period 1988-2004, find positive results regarding technical trading strategies. However, such strategies are in general profitable until the late 1980s, but not later. Timmermann and Granger (2004) propose that new technical trading rules may be profitable for their first users, but once they are documented in the academic literature and come to be known, they quickly lose their effectiveness or even reverse their profitability. This happens because the very exploitation of trading opportunities evidenced by technical rules affects the prices heavily enough for the rule itself to be invalidated.

Despite the findings just described, which cast a shadow of doubt on our chance of success, this approach still remains appealing, in the computer scientist's perspective, because an algorithm can quickly compute any such indicator and analyse many of them, in order to discover possible patterns. For this reason, the technical analysis approach is at the core of many researchers' work.

2.4.3 Behavioural Analysis

Richard H. Thaler is 2017 Nobel Prize winner in Economic Sciences for his contributions to behavioural economics. `https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2017/`

Upon the turn of the century, a new attitude towards financial analysis grew stronger: more and more empirical results were showing that EMH was flawed, at least in its strictest interpretation; therefore, scholars started disputing the very foundational assumption of EMH, that is the rationality of investors. Behavioural analysis is a new approach to financial markets, that introduces the idea of integrating principles of psychology and sociology in the analysis of financial markets, to explain basic facts about the aggregate stock market as well as individual trading behaviour, inexplicable in the traditional framework.

Behavioural analysis holds that at least some financial agents may not be fully rational. According to Barberis and Thaler (2002), this has two main consequences, that are the basic principles of behavioural analysis itself.

First, when irrational traders interact with the market, irrationality can have a significant and long-standing impact on prices, that rational investors cannot correct without a considerable amount of risk or cost, or both. This contrasts with the traditional framework, where it is assumed that any irrational mispricing creates an attractive opportunity for rational investors, resulting in a quick and risk-free correction. In behavioural analysis, this principle is not true, not least because of transaction costs of the operations.

The second founding principle of behavioural analysis are those models that try to investigate and explain the systematic biases that arise when people form their beliefs, expectations, and preferences. Barberis and Thaler (2002) describe some typical human behaviours that may heavily affect investors' conduct in the financial markets. As an example, people are overconfident in the correctness of their own judgement, leading to poor abilities at evaluating the probability of events to happen or not (e.g., the probability of a stock to maintain a given trend, or to revert it). Belief perseverance is another instance of typical human irrationality: in general, people are reluctant to search for evidence that contradicts their belief, and even in case they find such evidence, they treat it with undue scepticism[7].

7. With some degree of irony, the two authors mention as an illustration of this point the adherence or opposition to the theory of efficient markets: those who believe in EMH happen to continue to believe in it, even though compelling evidence to the contrary has emerged.

In this context, irrationality of investors expresses in unsustained optimism or pessimism in the performance of a stock or of the market as a whole. The final idea of behavioural analysis is that the general prevailing attitude of investors is bound to, and to a certain extent influences and anticipates, the upcoming price development in a market. As a result, detecting the overall investors' sentiment can be helpful to predict future trends. To overcome self-evident issues related to, for example, representativeness and completeness of such sentiment detection, Baker and Wurgler (2007) suggest that a practical approach may be combining several measures, every one of which, taken singularly, is deemed to be an imperfect proxy of investor sentiment. Over the years, researchers proposed several such proxies; at least five different categories may be identified:

- *financial market-based measures*: investors generally agree that the market itself may convey useful information on the optimism and attention of investors; trading volume is regarded as a good indicator of investors' attention (Gervais, Kaniel and Mingelgrin, 2001); another interesting indicator is Volatility Index (VIX) by Chicago Board Options Exchange (CBOE) (Baker and Wurgler, 2007), which measures market volatility in terms of expected range of movement in the S&P 500 index over the next month[8].

8. `http://www.cboe.com/VIX`

- *survey-based sentiment indexes*: this kind of indexes directly measures investors' sentiment by means of surveys or polling newsletters issued by selected experts; AAII Investor Sentiment Survey (AAII) by the American Association of Individual Investors, and Investors Intelligence Sentiment Index (II) and are two such indexes; Brown and Cliff (2004), for example, evidence a high correlation between sentiment levels and price changes, but also show that sentiment has little predictive power for near-term future stock returns.

- *relevant sources of textual data*: some researchers propose to use text mining and sentiment analysis techniques to extract information about investors' mood from specialised and non-specialised social networks (e.g., StockTwits, Twitter) and newspapers (e.g., *Wall Street Journal*, *New York Times*, *Financial Times*); Dougal et al. (2012) report direct evidence that the writing of specific journalists in *Wall Street Journal* has a casual effect on the short-term returns of the DJIA, while Bollen, Mao and Zeng (2011) present how the public mood, measured on a large-scale collection of daily Twitter posts, is correlated, and in fact can be used to predict, DJIA changes over time.

- *internet search behaviour of households*: from the observation that people start their decision-making process by gathering relevant information (Simon, 1955), it is possible to conclude that search volume data from search engines may be a good proxy of investor attention; Google Trends is a widely used tool to get this kind of data (Preis, Moat and Stanley, 2013).
- *non-economic factors*: maybe the most elusive sources of market sentiment data include profoundly non-economic factors such as news that are unrelated to any particular company or industry segment (Kaplanski and Levy, 2010), weather (Hirshleifer and Shumway, 2003; Cao and Wei, 2005), sports events (Edmans, García and Norli, 2007).

While authors suggest that behavioural analysis approach may be promising, and notwithstanding its clear relatability[9], availability and granularity of data is still a limiting factor. Survey-based indexes are generally proprietary and publicly unavailable. Sentiment analysis and text mining on news and social media is a complex process: other than data availability[10], the very techniques to perform such analysis are not fully settled yet, with much development ongoing, and would deserve a whole research work. Non-economic factors seem to be too elusive to be taken into account.

9. After all, behavioural analysis deals with sentiment and real-life data, rather than obscure mathematical functions that manipulate historical price series.

10. In recent years, major newspapers have put up paywalls, to restrict the free access to contents (e.g., *New York Times*: `http://www.nytimes.com/2010/01/21/business/media/21times.html`). Likewise, social media (e.g., Twitter: `https://developer.twitter.com/en/docs`) limit their data availability.

3
Financial Markets Prediction in Empirical Studies

In this chapter, we are going to report a survey of the advances of empirical research in the field of financial market prediction. Our review work is partly based on the surveys by Park and Irwin (2007) and Cavalcante et al. (2016); our interest is to highlight the motivations that led researchers to apply different forecasting models, the strengths and shortcomings of their approach, and the overall performance of each group of methods.

The surveyed studies investigate the usefulness of technical analysis in a variety of markets for the purpose of either uncovering profitable trading rules or testing market efficiency, or both. Most studies concentrate on stock markets (stock prices and stock indexes), both in and outside the US, and foreign exchange markets (FX markets) (especially considering foreign exchange rates (FX rates) to or from US dollar). Most commonly, data are sampled on a daily basis, leading to a prediction at a one-day time granularity; however, this is not always the case.

The chapter is organised as follows. In each of the following sections we group the different research works based on the approach they take. Each section, then, describes separately the prediction of stock prices, indexes, and FX rates; unless otherwise stated, a one-day time granularity should be assumed. The high-level grouping of methods that we can identify is the following:

- standard and model-based bootstrap studies (section 3.1);
- genetic programming (section 3.2);
- neural networks (section 3.3);
- fuzzy logic-based methods (section 3.4).

It is important to emphasise that in this survey we are considering only quantitative technical analysis methods. Therefore, both fundamental and behavioural analysis methods are excluded from this work. Fundamental analysis is hard to automatise and highly subjective (section 2.4.1), as it requires the assessment of multiple statements in plain text from the company. The only aspect of fundamental analysis that can be easily adopted in a quantitative approach (and a few researchers do so) is the incorporation of macroeconomic data, such as commodity prices and financial indicators.

On the other hand, behavioural analysis methods appear to be less mature (section 2.4.3), and in general provide an outlook of the market, based on agents' sentiment, rather than a measurable impact on the day-to-day market movements.

3.1 *Standard Studies*

In standard studies, single or small groups of technical trading rules are optimised based on a specific performance criterion, and out-of-sample verification is performed to test the actual profitability of the rule. Common rules that are tested in these studies include moving average crossovers and filter rules. The former states that a short-term moving average (e.g., five days) of the prices generates a buy signal when it crosses upward a long-term moving average (e.g., 200 days), and conversely generates a sell signal when it crosses downward the long-term moving average; the latter is that a buy or sell signal is generated when today's closing price rises above its most recent low, or falls below its most recent high, respectively. Nowadays, with computing power increasingly available at always cheaper price, this kind of studies appear to be outdated and obsolete: nevertheless, their relevance lies in the fact that the research presented below has imposed a new, more scientific method for the prediction of financial data and for the evaluation of models and results.

Park and Irwin (2007) identifies in the work by Lukac, Brorsen and Irwin (1988) the first 'modern' empirical study, in that systematic statistical tests of significance on trading returns are performed, along with adjustment for transaction costs and risk. While their results are of little interest to us, as the authors work on commodity futures, two aspects are worth noting. First, to test the quality of their technique, they update the parameters of the rules yearly, based on the past three years; this guarantees adaptive optimal parameters and out-of-sample testing. Second, they test both gross returns and net returns to be different from zero.

Foreign Exchange Rates Park and Irwin (2007) report that several studies in this category investigate FX markets. The technical rules studied in those works yield positive annual net returns for most FX rates, after adjustment for transaction costs. However, profits of technical trading rules in FX markets seem to decline over time. Olson (2004), in particular, identifies the best-performing moving average crossover rule for each successive period of five years from 1971 to 1995, and tests their profitability in the following period of five years. On the 18 different currencies under scrutiny, the author finds that risk-adjusted trading rule profits have declined over time from an average of over 3 % in the late 1970s and early 1980s to about zero in the 1990s, both in-sample and out-of-sample. Olson concludes that the inefficiency that led to positive returns between the 1970s and the 1980s must have been temporary.

Stock Prices and Stock Indexes Taking a similar approach, Taylor (2000) tests moving average crossover rules on a number of US and UK stocks and indexes. The author is able to prove, with statistical evidence, that positive gross returns can be achieved, in general, on indexes, but not on stocks. Taylor takes a further step, computing the breakeven transaction costs[1] for the considered data series: indexes feature a mean breakeven transaction cost of 0.58 %, much higher than those of single stocks, that is around 0.08 % (notice, however, that the computations are performed on different periods of time, for the different data series).

1. Breakeven transaction costs are defined as the percentage one-way trading cost that eliminates the additional return from technical trading.

3.1.1 *Model-Based Bootstrap Studies*

This class of studies extends the previous approach, by evaluating technical trading rules against model-based bootstrap samples. The method consists of fitting a model (or a number of different models) to the data series under scrutiny; using such models, several bootstrap samples are generated. Technical rules are applied both to the original time series and to the samples: a statistical comparison of the results can emphasise the robustness of the results.

Stock Indexes A wide stream of research started with the seminal work by Brock, Lakonishok and LeBaron (1992). They analyse the profitability of two common trading rules on the Dow Jones Industrial Average (DJIA) index: the moving average crossover rule, and the trading-range break, where signals are generated when prices hit new highs or lows. The findings of the authors are that such rules are effective predictors of price movements, and that the so-obtained profits are not compatible with a random walk or common time series models. For statistical inference, Brock, Lakonishok and LeBaron use bootstrap to confirm and improve the robustness of results: the returns conditional on buy and sell signals from the actual DJIA data are compared to the returns from simulated series generated by a fitted model from the null hypothesis class under investigation, with the residuals being randomly re-sampled with replacement. The null models considered are: autoregressive process of order one, generalized autoregressive conditional heteroskedasticity in-mean model (GARCH-M), exponential GARCH[2] (EGARCH). The technical strategies explored by the authors outperform the null models at a significant statistical level over the period from 1897 to 1986. Also, mean returns generated over a 10-day period, based on the two considered trading rules, are about 0.80 %, against the 0.17 % of buy and hold (B&H) over the same period of time. Transaction costs are not taken into account in this study: as a result, the work by Brock, Lakonishok and LeBaron is not enough to prove the actual economic profitability of the system.

2. These are statistical models for time series data. In particular, GARCH models are especially used in econometrics, and describe the variance of the current error term as a function of the actual sizes of the previous time periods' error terms.

Other authors extended and improved the work by Brock, Lakonishok and LeBaron (1992). Ratner and Leal (1999) report that the same two rules studied by Brock, Lakonishok and LeBaron are actually profitable in emerging markets (Taiwan, Thailand, and Mexico), even in later years and after transaction costs are applied; similar results are reported also by other researchers. Sullivan, Timmermann and White (1999) test the same and other trading rules, both on DJIA and on Standard & Poor's 500 (S&P 500), getting poor out-of-sample performance over the period from 1987 to 1996, which leads them to conclude that the efficiency of stock markets may have

Study	Target	Transaction Costs	Risk Adjustment	Results
Brock, Lakonishok and LeBaron (1992)	Stock indexes	No	No	Positive
LeBaron (1999)	FX rates	–	–	–
Ratner and Leal (1999)	Stock indexes	Yes	No	Declining
Sullivan, Timmermann and White (1999)	Stock indexes	Yes	No	Declining
Taylor (2000)	Stock prices and Stock indexes	–	No	–
Day and Wang (2002)	Stock indexes	Yes	No	Declining
Neely (2002)	FX rates	–	–	–
Olson (2004)	FX rates	No	Yes	Positive

Table 3.1: Summary of standard and model-based bootstrap studies.

improved, after the analysis by Brock, Lakonishok and LeBaron (1992); Day and Wang (2002) report similar results and draw the same conclusions.

Foreign Exchange Rates As it was already mentioned (section 2.3.1), LeBaron (1999) and Neely (2002) report a remarkable correlation between US official intervention and returns to a typical moving average rule, where such official intervention is expressed in terms of millions of dollars purchased or sold by the US Federal Reserve. LeBaron (1999), in particular, studies daily and weekly rates, and evaluates the performance of moving average crossover strategy using the bootstrap procedure described by Brock, Lakonishok and LeBaron (1992). Neely (2002) expands the works by considering a wider set of FX rates to US dollar and hourly (intraday) data, which allows him to confirm that there is a link between official intervention and higher returns, but also to conclude that the causality relationship between the intervention and the FX rate predictability may be less obvious than expected. As a matter of fact, high returns due to the trading rule in general *precede* such interventions, by less than 24 hours.

3.2 *Genetic Programming*

Genetic programming (GP), introduced by Koza (1992), is an optimisation procedure where functions, or computer programs by extension, are evolved over generations, until a single one is selected to be the fittest, based on some fitness measure and subject to some time or complexity constraint. When applied to technical trading, the building blocks of the functions to evolve consist of various functions of the past prices, as well as other numerical and logical primitives. The main advantage that GP features is that the search space of logical combinations of trading rules can be much wider than in the standard methods presented above; researchers do not need to pre-select parameters and structure of the function to test.

The strengths of GP rely on the fact that it can be effectively used in a wide range of problems, even where the space of possible solutions is too large to be handled efficiently by standard procedures. Moreover, the generated solutions are a composition of interpretable pieces (Bhattacharyya, Pictet and Zumbach, 2002; Dempster and Jones, 2001), which makes the resulting trading rules more relatable to the trader. In this sense, the solutions themselves are a simulation of real traders' strategies: a systematic technical trader rationally chooses her trading strategies from an arsenal of popular technical trading rules, combining them to meet some pre-specified evaluation criteria.

Results show that stocks and indexes are predictable, at least to some degree. However, once transaction costs are taken into account, profitability of trading rules is lost, in general. On the contrary, different studies confirm that FX rates are indeed predictable, and trading rules are profitable.

Stock Indexes Allen and Karjalainen (1999), among the first to apply GP in the field of stock market trading, use GP to learn technical trading rules for the S&P 500 index, using daily prices from 1928 to 1995. For a more robust evaluation of the performance of the models, ten successive training periods are employed: every five years, a 5-year training and 2-year selection period begins, with test periods starting from the end of the selection period and always ending in 1995. Selection period is used to avoid overfitting: only those trading rules that generate higher excess returns than the best rule so far are retained and passed to the next generation. With a rather limited function set to work on (the authors include only arithmetic, Boolean, and logical operations), and using excess returns over the B&H strategy as fitness function, the GP procedure fail to generate consistent excess returns over a simple B&H strategy, once adjustment for transaction costs is applied. Excess returns are calculated using several alternative one-way transaction costs, but even when such costs are at the lowest (0.10 %), average out-of-sample excess returns are negative for six out of the 10 periods. The authors, however, find some evidence of predictability in returns, as the rules tended to be in the market during periods of high returns and out of the market during periods of low returns.

Neely (2003) improved the work by Allen and Karjalainen (1999), in that he implements a risk-adjusted fitness measurement of the returns. Although the resulting rules may have lower returns than the B&H strategy, lower volatility may permit the returns to be leveraged up to exceed the B&H return with similar risk[3]. Different criteria to adjust the returns are used, but the trading rules fail to consistently and significantly outperform the B&H strategy by any of those risk-adjusted measures. Therefore, the results by Neely (2003) are consistent with those by Allen and Karjalainen (1999) and Efficient market hypothesis (EMH) in general. A few shortcomings of both works are evident, as they share the setting and approach. First, the function set appears to be quite limited, which does not allow trading rules to grow and capture the full complexity of the data at hand. Also, in a situation where profits are the main goal, it may be appropriate to consider short positions, which enable profits also on price drops.

3. This is possible when trading on margin is allowed, that is essentially investing with borrowed money.

Stock Prices Potvin, Soriano and Vallée (2004) work on single Canadian stocks, rather than using a composite index like S&P 500. Unlike previous works, the authors include a few technical indicators in the function space to be used by the GP algorithm. Once again, only long positions are considered. Another shortcoming of the research work is that transaction costs are not considered. However, a few interesting results are shown. With the exception of a few stocks, positive returns can be achieved using the strategy identified by the system. Also, those few stocks that generate negative returns show relevant discrepancies in the transaction volumes and prices over the training and testing periods. To conclude, the authors observe that, when the B&H approach generates negative or small positive returns, the profits due to the various trading rules are generally positive, and in most cases more profitable than the B&H ones.

Being able to understand and replicate the statistical behaviour of real markets is regarded to be very valuable, in the perspective of making forecast and, ultimately, making profit. To gain a better understanding real financial markets, intended as complex systems where multiple agents interact, artificial markets have been used by researchers. These are simulations of stock markets, where artificial agents are allowed to trade shares, just like traders do; such trading affects the (simulated) prices of the stocks in the artificial market. By comparing the statistical properties of real and simulated stocks[4], it is possible to evaluate to what degree the simulation is able to capture the complexity of real stock markets. Martinez-Jaramillo and Tsang (2009) propose a method to simulate an artificial market, composed of different kinds of agents. In addition to noise traders, which operate with no specific strategy, and fundamental agents, which adopt a deterministic and fixed strategy, the authors introduce technical traders: these are represented as decision rules, evolved with GP and based on technical indicators. It is important to notice that all technical actors are evolved together (coevolution), because the interaction itself is critical to the traders' performance. Extensive testing allows concluding that the prices generated in the artificial market are realistic, in terms of statistical properties, especially in the case where different groups of technical traders were assigned different groups of technical indicators to evolve with.

4. A few properties that are tested are the insignificant autocorrelation of returns, and volatility clustering, whereby large changes in price tend to be followed by further large changes, and small changes tend to be followed by further small changes. Another important property of stock returns is the non-Gaussianity, in terms for example of high kurtosis (heavy tails).

Foreign Exchange Rates Neely, Weller and Dittmar (1997) apply GP to generate trading rules for the FX market. A single training-validation-test period is defined for all six FX rates that are considered, from 1974 to 1995. The function set used to generate the solutions is quite limited, as it includes arithmetic, Boolean, and logical operations only. In the out-of-sample period from 1981 to 1995, find strong and consistent evidence of economically significant excess returns after transaction costs. Further work by the authors (Neely and Weller, 2001) confirms the results, although trading profits appear to gradually decline over time, reaching zero or going negative during the 1990s.

FX markets are the target of the work by Bhattacharyya, Pictet and Zumbach (2002). Their solution is based on strongly typed genetic programming (STGP) (Montana, 1995), a variation of traditional GP where type constraints are enforced. Moreover, semantic and domain-related restrictions are introduced, with the goal of reducing the complexity of solutions and improving the robustness of result: primitives and composition functions are introduced subject to the symmetry condition of FX markets[5]. The GP models are evaluated against a risk-adjusted fitness measure, which results in models that are less prone to overfitting. In addition to this, the results show

5. A trading model that is optimal for the USD/EUR rate, is expected to be optimal also for the inverted EUR/USD, since both rates correspond to the same market. Therefore, the recommended signals of a consistent and sensible trading model must be anti-symmetric with respect to price changes.

Study	Target	Transaction Costs	Risk Adjustment	Results
Neely, Weller and Dittmar (1997)	FX rates	Yes	No	Positive
Allen and Karjalainen (1999)	Stock indexes	Yes	No	Negative
Neely and Weller (2001)	FX rates	Yes	No	Declining
Bhattacharyya, Pictet and Zumbach (2002)	FX rates	Yes	Yes	Positive
Neely (2003)	Stock indexes	Yes	Yes	Negative
Potvin, Soriano and Vallée (2004)	Stock prices	No	No	Positive
Martinez-Jaramillo and Tsang (2009)	Stock prices	–	–	–

Table 3.2: Summary of genetic programming studies.

that semantics and risk-adjustment give more stable solutions (i.e., less variance of the returns and fitness measure among different applications of the GP procedure), to more simple trees (i.e., with fewer nodes), and to generally improved out-of-sample performance (while in-sample performance remains similar, leading to the conclusion that overfitting is reduced). However, a comparison with B&H strategy is not provided, so a complete evaluation of the approach is not possible.

3.3 *Artificial Neural Networks*

Artificial neural networks (ANNs) are among the most common methods used to explore the predictability of stock prices, thanks to their inherent ability to reproduce and model nonlinear processes, even without any prior knowledge about the input data distribution. The wide variety of available architectures and types of neurons, along with their wide popularity, puts any effort to review the literature at the risk of being incomplete and insufficient. Therefore, especially in this section, our goal is to only report some relevant literature contributions, based on the innovations they introduce; a significant, if relatively old, review work was proposed by Atsalakis and Valavanis (2009b), where the authors have surveyed over 100 scientific articles that apply soft computing techniques to solve the financial market forecasting problem.

With the exception of some early applications, results are in general positive for all targets we consider. A research trend that we can identify, especially in recent studies, is the integration of evolutionary computation techniques and ANNs: in most of the cases, this was done to overcome the high sensitivity of neural networks to noisy data, as in fact are financial data, and speed up the training phase.

Stock Prices White (1988) takes his place among the first researchers who applied ANN to the financial field. The network he devises, five neurons organised in a single hidden layer, fails to predict daily stock returns for IBM Corporation (IBM). This work, despite being primeval and of limited scientific utility in modern times, shows how the interest on ANN and financial markets origins far back in time.

One of the main drawbacks in the network proposed by White (1988) is that it was too simple (both in size and structure) to correctly capture the features of financial data series. Saad, Prokhorov and Wunsch (1998) suggest using more complex architectures, that allow to better capture short-term variations in stock prices. Focusing on ten US stocks, the authors compare the performance of three different neural networks: time delay, recurrent, and probabilistic neural networks. Their goal is trend prediction, which can be formulated as forecasting whether monthly returns will be above or below an arbitrary threshold; in particular, they focus on limiting false alarms, that is inaccurate prediction of great returns. The recurrent neural network architecture, structured as a single hidden layer of recurrent neurons, with one input of the normalised daily closing price, shows to be slightly better than the other two: in the period from October 1995 to March 1996, the recurrent network predicted about half false alarms, compared to the other two architectures.

Stock Indexes In more modern times, researchers moved their interest to the application of ANNs to the prediction of stock indexes, rather than single stock prices. The wide research by Gençay (1998a, 1998b) aims at proving that the predictability of financial data series is nonlinear, and that there may be a long-term dependency between current prices and the past. In particular, Gençay (1998a) reports that a feedforward neural network is able to predict DJIA, and generate a positive net return, after transaction fees are accounted for. In detail, across six out-of-sample sub-periods between 1963 and 1988, the trading rules generate annual net returns of 7 % to 35 %, and easily outperform a B&H strategy. Most notably, short-selling is allowed in this research work. Gençay (1998b), and Gençay and Stengos (1998), improves the previous work by incorporating past trading signals from technical trading rules (e.g., moving average rule), lagged returns and volume indicators as additional regressors into the same feedforward neural network, confirming the previous results.

The work by Kim and Han (2000) proposes to use a genetic algorithm to determine the connection weights of an artificial neural network that predicts Korea Composite Stock Price Index (KOSPI) price changes, and to discretise the input features to such network. The feature discretisation is applied to convert the input values of a number of technical indicators into categorical features, as a form of dimensionality reduction. The optimisation of the thresholds for feature discretisation and of the connection weights between the layers of the neural network, thanks to the genetic algorithm, is simultaneous. Testing the performance on the last three months of each year, from 1989 to 1998, shows an overall out-of-sample accuracy of around 61.70 %. This result outperforms the two comparative systems: the first one is a neural network trained with back-propagation, and fed with input data scaled into the range [0, 1], whereas the second is a genetic algorithm-trained neural network with, once again, scaled input data.

The idea of a genetic algorithm that optimises the weights of an artificial neural network is followed also by Kwon and Moon (2007), however with some specificities. The considered neural network architecture is recurrent, in this case; this allows the authors to represent the weights of the network as a 2D-matrix, that a genetic algorithm is able to optimise. The input data are a number of technical indicators, based on stock prices and volume. The authors test their prediction system on 36 US stocks, and on each year from 1992 to 2004, with the goal of measuring the total profit. In most cases, the proposed method was able to perform significantly better than the B&H strategy. In both previous cases (Kim and Han, 2000; Kwon and Moon, 2007), a genetic algorithm was introduced to speed up the training of network weights, especially in situations where the input data is large and noisy.

In a recent work, Wang and Wang (2015) proposed a solution where principal component analysis (PCA) and stochastic time effective neural network (STNN) are combined, to forecast the next-day values of various stock indexes (Shanghai, Hong Kong, United States). STNNs, in particular, are a type of feed-forward neural network where the input data are weighted based on their age and volatility. This serves to fix the relative importance of each piece of data, according to the distance in time from the current datum, and also depending on the volatility of the data series. An evaluation of the results shows that the prediction error is higher, where there are large return volatilities (essentially, where indexes fluctuate more violently). PCA is used, in turn, to reduce the dimensionality of the input. With a mean absolute percentage error (MAPE) consistently lower than 2 %, the network proposed by the authors outperforms the benchmark systems for all the indexes considered. As a comparison, the authors considered a STNN without PCA, and a common back-propagation ANN.

Foreign Exchange Rates Another target that researchers considered for prediction with ANNs are FX rates. Gençay (1999) applied his feedforward neural network (Gençay, 1998a) to forecast FX rates returns; on average, the network is able to predict the sign of FX rate change 58 % times.

Yao and Tan (2000) show how weekly FX rates between US dollar (USD) and five other major currencies can be predicted using a time delay neural network (TDNN), which processes moving averages of the FX rate at different time lengths. In particular, they conduct two experiments. First, the authors test the forecasting ability of a TDNN (five input nodes, three hidden nodes, and one output node) which receives as inputs the past five weekly rates, with poor results: the network tends to output the most recent value, simply delayed by one week. Therefore, the strategy resulting from the prediction is unable to outperform the B&H approach. To overcome the issue, a second experiment is performed: rather than the delayed weekly rates, Yao and Tan compute five different moving averages of the daily prices, from the past week up to the past six months, once again sampled at weekly frequency. Their goal is to smooth out much of the irregularities affecting the FX rates and better detect the trends of the market. In this second setting, the network shows good out-of-sample performance: with the exception of the FX rate to Japanese yen (JPY), the strategy is in general able to outperform the B&H approach, especially when B&H gives negative or small positive returns. Moreover, the authors notice that the performance on the five currencies shows a degradation in forecasting after about six months, thus resulting in the need for the network to be retrained.

In a well-written and properly justified work, Yu, Lai and Wang (2008) propose an ensemble learning model, based on radial basis function net-

Study	Target	Transaction Costs	Risk Adjustment	Results
White (1988)	Stock prices	No	No	Negative
Gençay (1998a, 1998b)	Stock indexes	Yes	No	Positive
Saad, Prokhorov and Wunsch (1998)	Stock prices	No	No	Positive
Gençay (1999)	FX rates	No	No	–
Kim and Han (2000)	Stock indexes	No	No	–
Yao and Tan (2000)	FX rates	No	No	Positive
Kwon and Moon (2007)	Stock indexes	No	No	Positive
Yu, Lai and Wang (2008)	FX rates	Yes	No	Positive
Sermpinis et al. (2013)	FX rates	No	No	–
Wang and Wang (2015)	Stock indexes	No	No	–

Table 3.3: Summary of artificial neural networks studies.

works (RBFNs). It is suggested that standard feed-forward neural network models may suffer from problems like slow training and a general tendency to get stuck at local optima; on the other hand, RBFNs can overcome the above drawbacks. In turn, an ensemble learning model is introduced to address the inherent instability of ANNs predictions, in particular when data is noisy or insufficient; in other words, ensemble learning is deemed to effectively produce more stable and robust results. A further RBFN is used to combine the forecasts of the single predictors, where diversity is enforced by varying the number of nodes in the hidden layer and the parameters of the radial basis function. With a training on the monthly data of four major FX rates from 1971 to 2000, and an out-of-sample test period from 2001 to 2006, the ensemble model is able to significantly and consistently outperform individual RBFNs in terms both of percentage error and of prediction accuracy of movement direction.

A RBFN is proposed also by Sermpinis et al. (2013), who use particle swarm optimization (PSO) to find its structure, along with its optimal combination of parameters. The authors give different FX rates, commodity prices and stock prices as input to the network, which outputs the rates of return of three FX rates (EUR/USD, EUR/GBP, and EUR/JPY). PSO, a population-based heuristic search algorithm, is evolved with a multi-objective fitness function, as three goals should be achieved at the same time. First, a low prediction error is desired; however, this does not guarantee that the prediction is economically significant (i.e., returns are not guaranteed to be profitable); therefore, the overall return is maximised; finally, a high number of hidden neurons is penalised, as the authors are interested in finding the simplest neural network that achieves the previous goals. Results show that the solution proposed by the authors is able to outperform all the considered benchmarks. However, a comparison with the B&H strategy is not reported.

3.4 Fuzzy Logic-Based Methods

An approach to the prediction of stock prices or the definition of a trading decision support system, that has been gaining increasing popularity in recent years, is the integration of fuzzy logic into a neural network. The goal of fuzzy logic (Zadeh, 1973) is to provide an effective way to describe the behaviour of a system by means of approximate variables (*linguistic variables*), in order to overcome the complexity of the system itself. In a fuzzy logic context, then, a variable x may be characterised as, for example, *large* or *not very small*. In this way, hard limits are not set, and a degree of truth is preferred[6]. Linguistic variables are combined using fuzzy logic, into fuzzy rules, in the form of if-then relations. In the context of our interest, a fuzzy system is used to describe and encode the expert traders' knowledge, which is qualitative and subjective by nature. By knowledge, we mean, for example, the interpretation of technical indicators.

6. After all, does it make sense to state that a car is moving 'fast' if its speed is greater than 80 km/h, and, accordingly, that the car is moving 'not fast' if its speed is, say, 79 km/h?

Stock Prices and Stock Indexes Kuo, Chen and Hwang (2001) develop a system that uses a fuzzy neural network (FNN) to interpret qualitative factors, such as political, economic, and international events, and quantify their impact on Taiwan stock prices. The fuzzy system is also used to integrate the signals from time effect, stock market tendency without special events, and events effect. In terms of generated return, the FNN system proposed by the authors is able to double the B&H return, during the period from January 1996 to April 1997, on the Taiwan Stock Exchange (TWSE) index. Similarly, Chang and Liu (2008) works on Taiwan stocks. In contrast with the previous study, though, a neural network is not used. Rather, the authors apply simulated annealing, a meta-heuristic to approximate global optimisation, to fit the parameters of a knowledge base of fuzzy rules. Based on stock prices and on a selection of technical indicators, in the out-of-sample period from July to December 2005, the system is able to predict changes in price direction with accuracy close to 98.00 % (based on MAPE) both for TWSE index and MediaTek stocks.

Atsalakis and Valavanis (2009a) propose a system that uses adaptive neuro-fuzzy inference system (ANFIS), to predict the next-day rate of change of a 3-day moving average of the prices[7] of various stocks from the US and Greek markets (the former being well-developed, the latter being regarded as an emerging market). ANFIS is a control system where an ANN is able to learn and adapt the membership functions of a fuzzy rule-based inference system. The authors adopt such composite method to leverage both the ability of neural networks to recognise patterns and adapt to changing environments, and the possibility of expressing human knowledge and expertise in a fuzzy inference system. Similarly to most other systems reviewed in this survey, the investor allocates assets to the stock (buys stocks) when there is a predicted uptrend for the next day, and sells when there is a predicted downtrend for the next day. The authors report good performance of the system, both on US and Greek stocks: with a long training phase over the period from January 1986 to March 2005, and out-of-sample testing on three successive periods between April 2005 and May 2006, the system is able to predict the sign of the change 62.32 % times, on average. For the

7. The 3-day moving average is chosen to smooth out excessive noise, with the result of focusing on a short-term trend of the stock price.

Study	Target	Transaction Costs	Risk Adjustment	Results
Kuo, Chen and Hwang (2001)	Stock prices and Stock indexes	No	No	Positive
Chang and Liu (2008)	Stock prices and Stock indexes	No	No	–
Atsalakis and Valavanis (2009a)	Stock prices	No	No	Positive

Table 3.4: Summary of fuzzy logic-based studies.

Greek stocks only, the authors report that the strategy significantly outperforms the B&H approach. In further research, Atsalakis, Protopapadakis and Valavanis (2016) focus on a portfolio of US stocks during particularly problematic periods of the history, which had clear impact on stock markets: Black Monday in 1987[8], Russian Ruble crisis in 1998[9], terror attacks on 11th September 2001, and the 2008 crisis. Data from such periods is regarded to be particularly complex, as violent fluctuations were occurring; nevertheless, the same system presented by the authors in 2009 is able to predict the sign of the change more than 66 % times, with consistently positive returns (while the B&H gives significantly negative results).

8. On 19th October 1987, financial markets around the world faced a significant and sudden drop, possibly caused by generalised investor panic, however unrelated to any interpretations of recent news events. https://www.bloomberg.com/news/features/2017-10-16/black-monday-at-30-wall-street-remembers-the-1987-stock-market-crash

9. The Russian Ruble crisis, started on 17th August 1998, is among the effects of a general situation of financial distress in international markets, and was prompted by the announcement of a set of measures by the Russian government, such as the devaluation of the Ruble to US dollar FX rate. https://economics.rabobank.com/publications/2013/september/the-russian-crisis-1998/

3.5 *Concluding Remarks*

While an outright comparison of the presented methods is difficult, not last because of different testing approaches and periods, it is still possible to discuss a few remarks on the success of the different methods. In light of what comes next, we decide to focus our review on the application of genetic programming (GP) to the prediction of financial markets.

We notice that the vast majority of research papers report successful results, but under specific and restrictive conditions, or on a limited set of targets. With a few exceptions, researchers test their methods on a few stocks or foreign exchange rates (FX rates), rarely exceeding a dozen data series. Moreover, transaction costs are not always considered in the evaluation of the performance of an approach, leading to overly optimistic, and sometimes altogether misleading results. Likewise, risk-adjustment is considered only sometimes: like in the case of transaction costs, results that are not adjusted to the risk of trading on financial markets can be misleading, as well as prone to overfitting (Bhattacharyya, Pictet and Zumbach, 2002).

3.5.1 *Methods*

If we were to select an approach to apply to financial data prediction, based solely on the presented results, the choice would be hard, as most studies fail to demonstrate solid and generalisable achievements in the forecasting activity. Rather, only an assessment of the strengths and weaknesses of each method can guide our choice, as an outstanding approach cannot be easily identified. As it was already mentioned, standard methods (section 3.1) are somewhat outdated, and their application would not exploit the full computing power available nowadays, leading to partial and limited results.

The solutions of genetic programming (section 3.2), being arrangements of financial state-of-the-art elements (i.e., technical indicators), are a composition of interpretable pieces (Bhattacharyya, Pictet and Zumbach, 2002).

This allows the human trader to perform manual analysis and tailoring: for example, a trader may be interested in testing the effectiveness and behaviour of a solution under different market conditions, in order to establish manual rules to constrain the application of the solution itself. A selected model may be found, for instance, to perform poorly near market closing hours, but exhibit good performance otherwise: the trader may want to modify the model or restrict its execution accordingly. On the downside, genetic programming is sensitive to the choice of the primitives with which the models are built: a poor choice of functions and technical indicators may limit the expressive power of the solutions. Nevertheless, a careful evaluation of the function set, as we will see in the next chapter, can greatly limit this risk.

The research in the artificial neural network (ANN) field (section 3.3), however promising, is wide, poorly structured, and oftentimes weakly motivated (Timmermann and Granger, 2004). The application of neural networks to financial prediction sometimes appears to be motivated by the desire to test a novel approach on yet another field of application, rather than by careful evaluation and domain knowledge. This is not to say that research in this field should be abandoned; rather, further research should be done, but taking a more scientific approach. On the application side, ANN models suffer from an unavoidable disadvantage: they are black boxes (Brenner, 2006). In a field where the human trader is supremely interested in monitoring what the 'machine' is doing, in order to avoid potentially damaging money losses, interpretability is a definite quality.

Fuzzy logic-based methods (section 3.4), in turn, have been gaining popularity only in recent years. Encouraging results are shown in our brief review, but maybe the research in the field is less mature than in other fields.

3.5.2 Targets

Although Park and Irwin (2007) summarise their survey of research papers published until 2004, by stating that technical trading strategies yielded economic profits in US stock markets through the late 1980s, they failed to do so thereafter. As to emerging stock markets, instead, several studies find economic profits regardless of the considered sample periods. For foreign exchange markets (FX markets), technical trading strategies generated economic profits over the last few decades, although some studies suggest that technical trading profits have declined or disappeared since the early 1990s.

Our review of more recent works, however, gives less clear insight on the predictive capabilities of technical trading methods, depending on the target financial data (we can notice, however, that most studies focus their attention on the prediction of returns, or even more often on the sign of returns). We cannot rule out that this is due to an increased tendency of researchers to publish only studies with positive results (the so-called 'file drawer bias'[10]), or the far more dangerous disposition to ignore details and factors that would invalidate the results (e.g., transaction costs, risk correction). Either way, we are taking a conservative position, and explore, in the next chapter, the predictability of all three categories of data reviewed so far, that are stock prices, stock indexes, and FX rates.

10. See Timmermann and Granger (2004) for a discussion.

3.5.3 *Time Granularity*

As we have already noticed, the focus of researchers on daily data prediction is overwhelming. We can attempt to propose a few reasons. The prediction on a daily basis has a more practical interest, rather than on longer spans of time. Trading at a lower frequency (e.g., monthly) does not allow to exploit small trends of market prices, thus missing the opportunity for profit; in other words, trading at a low frequency is similar to the trivial buy and hold (B&H) approach, the benchmark for all trading strategies. Predicting at higher frequency, instead, is generally harder, from a practical as well as technical standpoint. First, data availability: data providers usually charge for intraday data, and most commonly data is delayed by a small but not negligible amount of time. Real-time financial data is difficult to obtain, and usually is available only for the major actors of the exchange market.

Even when financial data can be obtained at the desired time granularity, technical difficulties arise. Whereas Müller et al. (1990) detect fractal behaviour in intraday FX rates (more specifically, their returns), Goodhart and O'Hara (1997) note that a fine understanding of the microstructure of the market (i.e., the fine details of the market data process) is needed, to properly use high-frequency data in financial analysis.

4
Genetic Programming for Financial Markets Prediction

In this work, we decided to focus our attention on the prediction of financial markets using genetic programming (GP). As it was previously discussed, one of the main advantages of GP is that the generated solutions are human-readable, and its very application allows to emulate the behaviour and choices a human trader would do, when defining her trading strategy in some financial market.

This chapter further develops and expands what was presented in section 3.2. We structured this chapter as follows. In section 4.1 we briefly present how GP works, at least in its most common fashion; this serves as an introduction, to outline the framework all the surveyed studies work in. Then, sections 4.2 to 4.5 review a number of studies, collected in different sections based on some common aspect. In particular, we discuss first the motivations that lead researchers to adopt GP (section 4.2), that are the interpretability of the generated models and the mitigation of data snooping. Then, we present a few relevant examples of original primitives sets (section 4.3), and subsequently how different studies deal with the dynamic behaviour of the market (section 4.4). To conclude, more complex systems, that integrate multiple predictive models, are introduced (section 4.5). The full set of studies that is reviewed in this chapter is presented in table 4.1, along with their target and time granularity.

4.1 Genetic Programming

Introduced by Koza (1992), genetic programming (GP) is a particular type of evolutionary optimisation technique, where the solutions are in the form of functions, or computer programs by extension, and their fitness is determined by their ability to solve a computational problem. Like other kinds of evolutionary algorithms, GP uses mechanisms inspired by biological evolution to evolve generations of candidate solutions, that play the role of individuals in the population. Similarly to natural selection in biology, the candidates with higher fitness measures are more likely to reproduce offspring and keep their genes alive, generation after generation.

More often than not, GP solution candidates are represented as tree-like structured computer programs[1], where each internal node is a function, and a leaf is a terminal (i.e., a constant value, a variable): GP assembles

1. As we will see later, different representations are possible: a grammar, a directed graph, a sequence of instructions from a programming language or machine language. The description that follows, which is focused on a tree-based GP, can be extended to the other representations with little effort.

Study	Journal/Proceedings	Target	Granularity
Neely, Weller and Dittmar (1997)	*The Journal of Financial and Quantitative Analysis*	FX rates: USD/DM, USD/JPY, USD/GBP, USD/CHF	1 day
Allen and Karjalainen (1999)	*Journal of Financial Economics*	Stock indexes: S&P 500	1 day
Wang (2000)	*Journal of Futures Markets*	Stock indexes: S&P 500	1 day
Dempster and Jones (2001)	*Quantitative Finance*	FX rates: USD/GBP	15 minutes
Bhattacharyya, Pictet and Zumbach (2002)	*IEEE Transactions on Evolutionary Computation*	FX rates: USD/DM	1 hour
Svangård et al. (2002)	*Proceedings of the 2002 Congress on Evolutionary Computation*	Stock prices: NOK	1 minute
Dempsey, O'Neill and Brabazon (2006)	*2006 IEEE International Conference on Evolutionary Computation*	Stock indexes: S&P 500, Nikkei 225	1 day
Izumi et al. (2006)	*2006 IEEE International Conference on Evolutionary Computation*	Stock prices: Japanese stocks	1 day
Lee and Moon (2010)	*Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*	Stock indexes: KOSPI	1 day
Hsu (2011)	*Expert Systems with Applications*	Stock indexes: TAIEX	1 day
Wilson, Leblanc and Banzhaf (2011)	*Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*	Stock prices: AAPL, BRK-B, RIMM, RY	1 minute
Loginov and Heywood (2013)	*Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation*	FX rates: USD/EUR	1 hour
Contreras et al. (2017)	*Genetic Programming and Evolvable Machines*	Stock prices: European stocks	1 day

Table 4.1: Summary of the studies considered in this chapter. The journal or conference proceedings they are published in is reported, along with the target and time granularity of the data under investigation.

program structures of variable length from those basic units. Functions perform operations on their inputs, which can be either terminals or output from other functions, or both. In other words, the functions and terminals are the primitives, or building blocks, with which a GP program is built.

GP can perform a parallel exploration of the search space, by processing a large amount of useful information contained in each solution candidate: the evaluation of the solution candidate generates information about the quality of the component building blocks. A fitness-proportionate reproduction allocates an exponentially increasing weight to better building blocks, which leads to an optimal search in the stochastic environment, balancing between a promising search direction and a less visited direction.

Since the search operators in traditional GP can establish arbitrary functions and terminals as arguments for any function node, the function set is required to be well defined and closed with respect to the various arguments that it can have. This closure property thus requires that all elements of a tree return the same data type, so as to allow arbitrary subtrees to be recombined by crossover and mutation operators. It goes without saying that this requirement is very limiting, as the analyst is forced to constrain her function set in an unnatural way, for example by carefully defining primitives so as not to introduce multiple data types. Montana (1995) proposes a relevant improvement to the original GP implementation by Koza (1992), called strongly typed genetic programming (STGP). What Montana (1995) suggests is to enforce type constraints at all stages of the genetic exploration. Each terminal, function argument, and function output is assigned a data type, thus establishing a semantics for the initialisation and evolution of the trees candidate solutions: in fact, the tree initialisation routines and search operators are required to generate only semantically correct tree structures. In summary, two criteria have to be met, for a GP tree to be sound:

1. the root node of the tree returns a value of the type required by the problem;

2. each non-root node returns a value of the type required by the parent node as an argument.

Initialisation Procedure The initialisation procedure is the initial creation of the individuals, that will be evolved. The two type constraints introduced above are enforced on the creation of this first set of individuals. Two different methods were initially proposed (Banzhaf et al., 1998, chapter 5): the *grow* method and the *full* method. Upon the selection of the maximum depth a tree should reach, the *grow* method generates completely random trees; in other words, the generated trees are sound, but no other restriction is imposed on their structure. The resulting population is highly irregular, because no minimum depth limit is enforced: once a branch contains a terminal node, that branch has ended, even if the maximum depth has not been reached. The alternative generation method, the so-called *full* method, creates trees such that terminals are put only at the maximum depth, and intermediate levels contain only non-terminals (i.e., functions).

Those two methods, however, are deemed to be insufficient to introduce an appropriate degree of diversity in the GP population (Banzhaf et al., 1998, chapter 5). To solve this problem, an intermediate approach can be taken. The *ramped-half-and-half* technique works as follows: once a minimum and a maximum depths are set, the population is divided equally among individuals to be initialised with all the intermediate depth levels. For each depth group, half of the trees are initialised with the full technique and half with the grow technique.

Genetic Operators The fitness of the individuals of an initialised population is generally low. Evolution is intended to introduce variations in the individuals, with the goal of improving their ability to solve the problem at hand. The three principal genetic operators are:

- crossover;
- mutation;
- reproduction.

Crossover (fig. 4.1) is the operation by which the genetic material (i.e., subtrees) of two parents are combined, by swapping a part of one parent

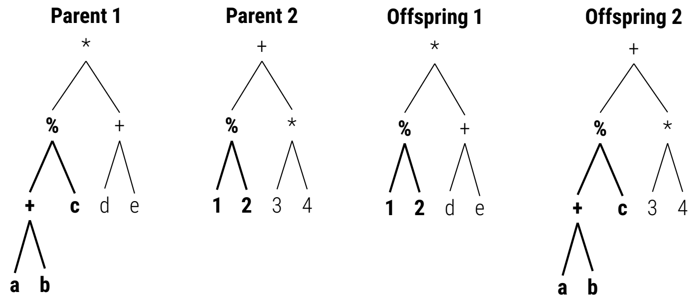

Figure 4.1: Crossover operation. The subtrees that are swapped are evidenced by darker lines and bold font.
Adapted from: Wolfgang Banzhaf et al. 1998. *Genetic Programming: An Introduction. On the Automatic Evolution of Computer Programs and its Applications.* San Francisco, CA, USA: Morgan Kaufmann, January, chapter 5

with a part of the other. To improve upon the exploration of complex solutions, the selection of subtrees can be biased so that terminals are selected as subtrees with lower probability than non-terminal functions.

Through *mutation*, single individuals can undergo a random change, usually with low probability. When an individual has been selected for mutation, the mutation operator randomly selects a point in the tree, and replaces the existing subtree at that point with a new, randomly generated, subtree. The new randomly generated subtree is created in the same way, and subject to the same limitations on depth, size, and type as programs in the initial population.

When an individual from the parent population is chosen for *reproduction*, it is copied, unmodified, in the offspring population.

Depending on the specific choice of the human programmer, crossover and mutation may be applied sequentially or exclusively: in the former case, the resulting individuals can be generated from crossover or mutation, or both; in the latter, an offspring will never result from both operations crossover and mutation.

Selection One of the key points of any GP, which greatly determines the quality of the results, is the selection policy, whereby the GP algorithm chooses which individuals of the population will be subject to the genetic operators of crossover, mutation, and reproduction. A common choice for the selection policy is (single) tournament selection: k individuals are randomly sampled with replacement from the current population of size N, and the one with the best fitness is selected, among those k. The tournament size allows researchers to adjust selection pressure: a large tournament size causes high pressure, resulting in a fast, although possibly suboptimal, convergence on a solution. In general, selection will be applied N times for each generation: this is because GP produces one offspring by applying mutation to one parent, and produces two offspring by applying crossover to two parents. Notice, however, that single tournament selection is not the only possible choice, and many other solutions have been proposed[2].

2. Interestingly enough, double tournament selection, an alternative to single tournament selection, has not been applied, in any of the studies we reviewed. First described by Luke and Panait (2002), double tournament selection can be used to reduce the risk of *bloat*, a common phenomenon that determines the uncontrolled and unbounded growth of individuals in the GP population, often without any significant improvement in the fitness.

Algorithm It is now possible to sketch a high-level view of the GP execution cycle. Notice that in GP, there generally exist well-defined and distinct generations, and each generation is represented by a complete population of individuals. The newer population is created from and then replaces the older population. The full GP algorithm is as follows:

1. initialise the population
2. evaluate the fitness of the population individuals
3. repeat the following, until a termination criterion is met:
   (a) repeat the following, until a new population is fully generated:
      - select an individual or two individuals in the population, using the selection algorithm
      - perform reproduction or, alternatively, either crossover or mutation on the selected individual(s)
      - insert the result(s) of the genetic operation into the new population

   (b) evaluate the fitness of the new population individuals

   (c) pass to the next generation the best individuals from the old population and the new
4. present the best individual in the population as the output from the algorithm

A few points are worth a discussion. At step 3a, each individual that is not selected for reproduction can undergo only one genetic operation between crossover and mutation, but not both. Alternatively, it is possible that mutation is applied also to those individuals generated by crossover operation. Moreover, at step 3c the so-called $(\mu + \lambda)$ strategy is applied. μ is the number of individuals to select for the next generation, and λ is the number of children to produce at each generation: at the end of each fitness evaluation, the best individuals in the whole population of parents and offspring is passed to the next generation. The common alternative strategy, identified as (μ, λ), provides that the best individuals among the offspring only make their way to the next generation. In the research literature, there is not full agreement on the better performance of $(\mu + \lambda)$ or (μ, λ) strategy.

4.2 Motivations

Researchers mainly propose two motivations for their choosing GP for financial markets prediction. The first one is the interpretability of the proposed solutions. From our discussion in the previous section, it is evident that GP solutions are compositions of operations and functions: should a mathematical relationship be detected between input data and the target variable (e.g., stock prices), such relationship can be read by the human. If unrestricted, however, the solution can grow large, which may jeopardise its interpretability. As we shall see, some researchers put in place measures to limit the growth of candidate models and enforce semantics in the construction of solutions.

A second proposed motivation to support the choice of GP is the mitigation of data snooping bias. Data snooping is a subtle issue typical of all data-driven applications. Most commonly, data snooping is due to a naive or careless management of the separation between in-sample and out-of-sample data, both within the context of the same experiment and among different experiments by separate researchers, and results in good but totally spurious out-of-sample performance of a predictive model. What some researchers maintain is that, through GP, it is easier to mitigate such risk.

4.2.1 *Interpretability of the Models*

Basic interpretability of models, besides being generally desirable, is particularly important in the financial field, where any sensible trader would like to check the behaviour of model components, possibly to implement domain-related knowledge in the solution. Bhattacharyya, Pictet and Zumbach (2002) report that financial models obtained through automated search are often subject to manual analysis and tailoring prior to implementation. It is common practice to test the behaviour and effectiveness of the selected models under different market conditions, in order to identify and establish manual rules that constrain their usage. A selected model may be found, for instance, to perform poorly near market closing hours, but exhibit good performance in the early trading hours; therefore, the human trader may want to implement manual rules on the whole model, or on single model components, to inhibit their effect, based on the previous testing. GP is particularly suitable for this manual testing and tailoring, as the generated solutions are intelligible compositions of functions and, possibly, state-of-the-art financial calculations, such as technical indicators. Net of the size of the model[3] (i.e., depth and number of nodes in the tree), the proposed GP solutions can be represented as trees and read by the human trader.

3. Which, when unconstrained, can grow huge.

Typically, the research papers where the adoption of GP is motivated by the interpretability of the solutions are those where semantic restrictions are imposed, and domain knowledge is more strongly and explicitly implemented in the genetic search process; this, in turn, usually limits the complexity (size) of the solutions. Bhattacharyya, Pictet and Zumbach (2002) develop a GP solution for foreign exchange trading (FX trading). The authors acknowledge that any FX trading model should meet the symmetry property: a trading model that is optimal for the USD/EUR rate, is expected to be optimal also for the inverted EUR/USD, since both rates correspond to the same market. Therefore, the recommended signals of a consistent and sensible FX trading model must be anti-symmetric with respect to price changes. This being acknowledged, the authors extend the idea of STGP (Montana, 1995) to allow a proper implementation of such symmetry-related semantic. In particular, the symmetry property is tracked at the individual nodes of the tree, by carefully assigning to each node both a data type (e.g., real, Boolean, ...) and a symmetry type: real values can be of antisymmetric type[4], symmetric type[5], or constant type. Arithmetic, comparison, and logical operators are redefined, in order to meet and enforce this semantics, noting that, for example, the sum of the results of an antisymmetric and of a symmetric operator is meaningless (e.g., is the sum of a moving average and of rolling variance symmetric or not? More so, is it meaningful at all?). These semantic restrictions are seen to ensure meaningful combinations of terms at the individual nodes, and to significantly limit the complexity of the solutions, although they are not a sufficient condition for straightforward interpretability, as some models are still found to be too

4. For example, a moving average function is antisymmetric. Let $A(\cdot)$ be a generic antisymmetric function (like the aforementioned moving average), and x be the logarithmic price; it is:

$$A(x) = -A(-x)$$

5. For example, the volatility (i.e., standard deviation) is symmetric. Let $S(\cdot)$ be a generic symmetric function (like the aforementioned volatility), and x be the logarithmic price; it is:

$$S(x) = S(-x)$$

Study	Mention Data Snooping	Validation	Multiple Test Periods
Neely, Weller and Dittmar (1997)	–	•	–
Allen and Karjalainen (1999)	•	•	•
Wang (2000)	•	•	•
Dempster and Jones (2001)	•	•	•
Bhattacharyya, Pictet and Zumbach (2002)	–	–	•
Svangård et al. (2002)	–	•	–
Dempsey, O'Neill and Brabazon (2006)	–	–	–
Izumi et al. (2006)	–	–	•
Lee and Moon (2010)	–	•	•
Hsu (2011)	–	•	•
Wilson, Leblanc and Banzhaf (2011)	–	–	–
Loginov and Heywood (2013)	–	•	–
Contreras et al. (2017)	–	–	–

Table 4.2: Application of measures to guard against data snooping bias and overfitting in the surveyed studies. The first column highlights studies that explicitly mention data snooping as a motivation for some of their implementation choices. The second column signals what studies make use of a validation period to select the best-performing models. The third column points out which studies test the performance of their models on multiple, non overlapping out-of-sample periods.

large or complex; either way, the number of nodes is nearly halved, with the introduction of semantic restrictions. Experimental runs on different test periods of length one year and a half, from early 1987 to mid-1994 show that the models exhibit more stable performance both in training and test periods, when semantics is introduced.

Dempster and Jones (2001) propose a slightly different motivation for their choice of adopting GP to forecast, once again, foreign exchange rates (FX rates). Their goal is to emulate the behaviour of a technical trader, who would build strategies from rules based on technical indicators, but would not use strategies based on very complex indicators, for a lack of transparency is detrimental to comprehensibility. The authors aim to discover whether or not existing indicator-based trading rules can be combined with Boolean operators to form profitable two-way trading strategies. In their work, the authors constrain the solutions to take the form of single if-then rules, with a sequence of individual conditions, connected by Boolean operators, that command a market action (i.e., buy or sell). For example, it is common practice to take a buy signal when a short-term moving average crosses upwards a long-term one: therefore, when a short-term moving average is above a long-term one, a buy signal is active. The system combines this and possibly other signals to issue a recommendation to the trader. Moreover, the authors limit the maximum depth of the tree at four (i.e., at most 16 nodes), half the maximum depth set in the work by Bhattacharyya, Pictet and Zumbach (2002).

4.2.2 *Mitigation of Data Snooping*

An issue of financial markets prediction, an inherently data-driven, non-experimental task, is that of data snooping bias. Data snooping is another term for the danger that the best forecasting model found in a given dataset by a certain specification search is just the result of luck, and not due to truly superior forecasting power (Lo and MacKinlay, 1990). Whenever the current research on the profitability of a specific trading strategy is motivated by, and based on, the successes and failures of past investigations, data snooping happens (this is also known as 'survivorship bias'): when researchers consider only popular trading rules, they are investigating only the subset of rules that may have produced abnormal positive returns by chance, possibly even without having any genuine forecasting power. Another source of data snooping arises when the properties of the data series under investigation influence the researcher's choice of model specification (data-driven investigation, as opposed to theoretically motivated models) (Sullivan, Timmermann and White, 2003). In other words, the reuse of a given set of data for purposes of model selection or inference is another source of data snooping. This may lead the researcher to completely mistake empirical correlation for causality, thus resulting in misjudged forecasting ability of the models under scrutiny and incorrect conclusions being drawn from the applied procedures. Timmermann and Granger (2004) argue that the mere application of a new search procedure to sample periods preceding the development of the procedure itself is inappropriate, and constitutes, in fact, yet another form of data snooping.

Notice that, in some sense, the concepts of data snooping bias and of overfitting are coupled, although different. Campbell, Lo and MacKinlay (1996, section 12.5) notice that while overfitting occurs when a model features an excellent in-sample fit but a poor out-of-sample performance, data snooping can lead to good but spurious out-of-sample performance. Overfitting happens when a model captures the random noise, along with the genuine non-linearities of the data under scrutiny. Data snooping bias, instead, arises when researchers ignore the fact that many specification searches have been conducted to obtain the final specification of a model they are fitting to the data, meaning that all the research has focused on the particular model that, on the peculiar stock and period, has produced positive results.

To a certain extent, the subtle and multifaceted issue of data snooping is inevitable, given the limited data availability and the specificity of the single markets. Some aspects of this issue can be addressed, so as to mitigate its overall impact and allow researchers to derive more robust results on the predictability of a data series and the forecasting power of a trading rule. Sullivan, Timmermann and White (1999, 2003) propose a few approaches to mitigate the problem, none of which can be expected to be always feasible, or sufficient to fully solve the issue: (i) utilising very long data series; (ii) emphasising the robustness of results across various, non-overlapping sub-periods for statistical inference; (iii) waiting for new data to become available, and (iv) using similar data from other sources.

Among the few researchers that explicitly mention the issue of data snooping in their studies, Allen and Karjalainen (1999) argue that, when a portion of the sample data is reserved for validation (sometimes referred to as selection data), GP is less subject to data snooping bias than other, more traditional approaches of studying technical trading rules: although the specification of GP trading rules is dependent upon their historical performance, their merit is further evaluated using validation data, not available to their construction. This way, the evolution of GP trading rules, that are generated *ex ante*, is (partially) decoupled from their *ex post* performance.

Building on the work by Neely, Weller and Dittmar (1997) and by Allen and Karjalainen (1999), Wang (2000) takes two different steps to detect and mitigate data snooping, and to ensure more robust results. First, a validation period is introduced, to select the best-performing programs and decide when to stop the training (i.e., the evolution process): after the population of individuals is initialised, the program with the highest in-sample fitness is stored as the best rule, and its performance is evaluated on a validation period, different from the in-sample data. At each generation, the fittest program is evaluated during the validation period, and retained as the best program only if it performs better than the best program so far. The GP evolution is stopped if a set number of generations is reached, or earlier, if the performance of the best program does not improve in a given number of generations. A further measure to reduce data snooping bias, applied by the author, is to train and test the system on different, non-overlapping periods, and to do so multiple times. This way, the presented results are less biased by the selection of the time periods, and more general conclusions can be drawn. In particular, for each of the 12 out-of-sample periods of one year, from 1987 to 1998, the author considered the previous year as the validation period, and the two years before as training period[6], and performed 10 different runs of the GP algorithm, thus running a total of 120 independent trials. Taking into account transaction fees, the results on daily Standard & Poor's 500 (S&P 500) index show that the performances of the trading rules in the out-of-sample periods varied over time: in some years, a majority of the rules beat the buy and hold (B&H) strategy, whereas in other years, most rules underperformed the market. The author concludes that, because an investor has to use the proposed procedure *ex ante*, she would not have very high confidence at the beginning of the new year that the trading strategies trained from the past would be consistently profitable. Nevertheless, the author presents evidence that the GP trading rules do have some predictive capability, based on statistical analyses of market-timing.

6. Therefore, when testing on 1987, the training period span 1984 to 1985, and the validation period is 1986. Likewise, when testing on 1998, the training period span 1995 to 1996, and the validation period is 1997.

The measures applied by Wang (2000) are present in the works of other authors, who take similar steps (table 4.2). However, seldom is data snooping bias mentioned as a motivation for their choice.

4.3 Terminals and Function Set

While many studies take a traditional and conservative approach to the definition of the inputs and the function set of the GP program (tables 4.3 and 4.4), a few stand out for their including a wider and original set of primitives. The goal of Svangård et al. (2002) is to evolve a trading strategy for an individual stock (Nokia) that outperform the simple B&H strategy. In order to do so, the authors expand the traditional set of input variables, which most commonly includes only daily prices and trading volume of the stock under investigation, in two ways. First, with the aim of capturing information on the conditions of the market, they include a selection of financial indicators, from the foreign exchange market (FX market) (e.g., various FX

Study	Prices	Volume	Moving Averages	Technical Indicators	Other
Neely, Weller and Dittmar (1997)	•	•	–	–	–
Allen and Karjalainen (1999)	•	•	–	–	–
Wang (2000)	•	•	–	–	–
Dempster and Jones (2001)	•	•	•	•	–
Bhattacharyya, Pictet and Zumbach (2002)	•	–	•	–	–
Svangård et al. (2002)	•	–	•	–	Feedback; Global FX rates, Stock indexes
Dempsey, O'Neill and Brabazon (2006)	•	–	–	–	–
Izumi et al. (2006)	•	–	–	–	–
Lee and Moon (2010)	•	•	–	–	–
Hsu (2011)	•	•	•	•	–
Wilson, Leblanc and Banzhaf (2011)	•	•	–	–	–
Loginov and Heywood (2013)	•	–	–	–	–
Contreras et al. (2017)	•	•	–	–	Fundamentals

Table 4.3: Inputs to the GP system in the surveyed studies. Moving averages and technical indicators are intended as statically computed, before the execution of the GP algorithm starts.

rates) and the stock market (e.g., the Dow Jones Industrial Average (DJIA) index), as well as short-term moving averages of Ericsson, whose changes are expected to be tightly linked to those of Nokia. Moreover, the authors notice that the ability to carry information over time is a prerequisite when searching for patterns in time series. In order to facilitate the ability to analyse lagged data, they introduce two feedback variables, namely the the previous output (thus making the agent aware of its action one minute ago) and the relative performance of the share price since the last transaction was made. This way, the the system is given an insight on the current gain or loss situation, possibly enabling it to decide whether it should capitalise now or in the next period. The system is tested on the second to third weeks of May 2001: the evolved agent is able to produce excess return, with respect to the B&H approach; the authors, however, are careful not to generalise the outcome, stating that the positive returns for the sample period say nothing about the long-run ability to outperform the passive (B&H) strategy. In particular, the GP agent is able to avoid trading in downward trends, although the timing of the purchase and sell actions appears to be slightly inaccurate, as they happen respectively when the share price is on its way down (but not at its lowest) and just before its peak.

Lee and Moon (2010), for their system aimed at predicting the Korea Composite Stock Price Index (KOSPI) index, take a step further into the interpretability issue discussed in section 4.2.1, by explicitly introducing domain knowledge in the GP structure: what they suggest to do is to define some 'attractive technical patterns', that are functions of technical indicators (therefore, they are functions of functions of the prices). Essentially, such attractive patterns provide state-of-the-art interpretation of the common technical indicators, returning Boolean values that can be understood as buy and sell signals; the GP algorithm is used to optimise both the para-

Study	Arithmetic Functions	Comparison Functions	Logical Functions	Moving Averages	Technical Indicators	Other
Neely, Weller and Dittmar (1997)	•	•	•	•	–	–
Allen and Karjalainen (1999)	•	•	•	•	–	–
Wang (2000)	•	•	•	•	–	–
Dempster and Jones (2001)	–	–	•	–	–	–
Bhattacharyya, Pictet and Zumbach (2002)	•	•	•	–	–	–
Svangård et al. (2002)	•	•	•	–	–	Trigonometric, log/exp functions
Dempsey, O'Neill and Brabazon (2006)	•	•	•	•	–	–
Izumi et al. (2006)	–	–	–	–	–	Candlestick patterns
Lee and Moon (2010)	•	•	•	•	•	Technical patterns
Hsu (2011)	NA	NA	NA	NA	NA	NA
Wilson, Leblanc and Banzhaf (2011)	•	•	•	•	•	–
Loginov and Heywood (2013)	•	•	•	•	–	–
Contreras et al. (2017)	–	–	–	•	•	–

Table 4.4: Primitives to the GP system in the surveyed studies. Arithmetic operations include sum, difference, multiplication and division, along with lag and rolling maximum and minimum.

meters of technical indicators (e.g., the window size of a moving average) and the Boolean composition of the signals coming from technical patters. As the technical patterns are structured as pre-built subtrees (in this sense, the authors define their approach as modular GP), it is important to ensure that the genetic operations of mutation and crossover preserve the structure of such subtrees; only parameters are allowed to change. The technical patterns are introduced to improve the readability of the candidate solutions, by moving the interpretation effort from the single tree nodes to higher-level structures, composed of multiple nodes, whose meaning is well-understood and agreed upon. A simulation on the single test years from 2005 to 2009 (with training being performed, each time, on the previous three years) shows a general ability of the system to outperform the B&H strategy, even after transaction fees and taxes. Notably, the best-performing solutions for each of the test years consist of four to five patterns, joined with common Boolean operators, thus confirming the improvement on the interpretability of the solutions.

4.4 *Real-Time Adaptation*

A wide stream in the literature on the application of GP to financial prediction and trading is devoted to adapt the system to the dynamic market conditions. Most typically, this means identifying the correct moment to trigger a re-training of the GP population. In an effort to closely emulate the behaviour of a human trader in the market (section 4.2.1), Dempster and Jones (2001) allow the proposed GP solution to build trading rules from already established technical indicators, unlike in many other studies, where only arithmetic and logical functions are used. Their goal is to explore a number of different approaches to adapt the training, and ultimately the whole GP system, to the dynamics of the FX market. The simplest setting is a static evaluation, with a single training period (year 1993), validation period (Q1 1994), and test period (Q2 1994 through Q4 1997). Because the results show that the generated returns become increasingly volatile, during the long out-of-sample period, and the risk-adjusted fitness ratio[7] declines in value, the authors devise different experiments to re-optimise the GP system. In a first attempt, a new training is triggered by feedback from the system's performance (reactive approach), in the same way that a trader would be compelled to find new rules after losses. In particular, the new training is performed either when the system makes a quarterly loss greater than 1 % or when the system has experienced two losing quarters in the last 12 months. After the initial training, another four optimisations are triggered. This kind of adaptation (reaction, in fact) shows poor performance, with respect to the static approach: after the points of re-optimisation, the system generally makes a loss within two quarters. The only aspect that is shown to improve is the performance of the best-performing strategy, which remains constant over the whole out-of-sample test period (which, however, is expected, due to the retraining). The bad performance in the reactive setting is justified as an overreaction to the market, resulting in an overfitting of strategies to the specific market conditions that determined the loss, conditions that may not be typical of the market as a whole. As an alternative, a periodic re-training is performed every quarter, and the newly discovered trading rules are used for a quarter only, before the process is restarted. In this case, however, the results are significantly worse than in the two previous approaches. The authors try to explain this poor performance with a bad synchronisation between the re-training itself and market modes, or conditions, to which the periodically re-optimised GP system does not manage to adapt.

Dempsey, O'Neill and Brabazon (2006), who adopt a GP variation where the population of individuals is evolved based on the rules established with a Backus-Naur form grammar, take a further step to address adaptability to the dynamic conditions of the stock market. In order to embed both a memory of good past trading rules and an adaptive potential in the trading system, the authors devise the system so as not to discard the evolved population, but reuse it for the periodic training of the system. In particular, after an initial training where a certain number G of generations is evolved, the system starts trading with the best-performing rule. When a number x of days have passed, the moving trading window moves forward, and the evolution of the population restarts, for a number $g < G$ of generations. This way, the knowledge of good past trading rules is not completely lost, and rather it serves as a starting point for their subsequent adaptation. The values x and g can be seen as further parameters to optimise: a small value of g means that memory is emphasised over adaptation, as the new data has relatively less chance to influence the trading rules, which can be undesired in periods of rapid market change. Similarly, if the value of x is large, the trading rules are altered less frequently. In the the experiments conducted by the authors, they set: $x = 365$, $G = 100$, and $g = 10$. Their test on the American S&P 500 and the Japanese Nikkei Stock Average (Nikkei 225)

7. A modified Stirling ratio is used as risk-adjusted fitness function, as follows:

$$S = \frac{R}{1 + \max(D, 0.02)}$$

where the drawdown D is the largest loss throughout a given period, and is measured as a percentage of the traded asset, like the return R. The 2 % drawdown tolerance is said to be rather common in FX trading, and is introduced to fix the behaviour of the ratio when the denominator is small.

from 1993 to 1997 show a dependence between the results and the trend of the indexes. In particular, the S&P 500 index featured a steady growth over the whole period, and therefore the GP system underperforms the B&H strategy; the authors justify this behaviour by stating that the strong uptrend leaves little chance to the trading system to close long position to reinvest at a lower price. A comparison with a similar system, that does not adopt the proposed periodic evolution, but instead periodically performs a full retraining, shows better performance of the adaptive system. Unlike S&P 500, the Nikkei 225 index showed a great amount of volatility, in the period under test, with significant rises and drops in its value. Therefore, the system performs considerably better than the B&H strategy: while the benchmark B&H gives a negative return (−19 %), the system on average yields a return of 55 %.

Interestingly enough, Loginov and Heywood (2013) compare all the approaches proposed by Dempster and Jones (2001) and by Dempsey, O'Neill and Brabazon (2006), drawing not entirely similar conclusions. The four settings tested by the authors are: (i) static approach, where the training is performed only before the beginning of the test period (Dempster and Jones, 2001); (ii) continuous evolution, where the training is performed periodically, starting from the already evolved population (Dempsey, O'Neill and Brabazon, 2006); (iii) stepwise evolution, where the training is performed periodically, starting each time from a newly initialised population (Dempster and Jones, 2001; Dempsey, O'Neill and Brabazon, 2006), and (iv) adaptive evolution, where the retraining is triggered when some quality criteria are exceeded (Dempster and Jones, 2001). The quality criteria that are monitored and used to trigger a retraining from an entirely new population are the magnitude of a decline in the account value (drawdown), the number of consecutive losses, and the number of consecutive hours without trading activity. Testing the different methods on hourly FX rate data in the period from July 2009 to November 2012 (approximately 17 860 hours), and allowing 1000 hours for the initial training, gives the following results: (i) retaining the population content between market cycles (continuous and stepwise evolution) appears to hinder the the generation of new trading agents capable of reacting to the next market cycle; (ii) training once at the beginning of the three-year period (static approach) and then assuming that the resulting champion individual will be effective does not work, and (iii) an adaptive evolution is noticeably the best-performing approach and, when associated to the application of a validation period, it is the only method that provides solutions with even the first quartile investments being non-negative (i.e., at least 75 % runs returned profitable training strategies). Moreover, the authors notice that, in most cases, the implementation of a validation period (at most 500 hours) to confirm the quality of the candidate solutions has a positive impact on the results. It is possible that the results obtained by Loginov and Heywood (2013) regarding the re-training, which are quite different from those of Dempster and Jones (2001), may be due to a better detection of the appropriate instant to trigger retraining; however, given the different target (USD/GBP in the earlier study, USD/EUR in the later one) and testing periods, it is difficult to come to an ultimate conclusion.

The study by Wilson, Leblanc and Banzhaf (2011) addresses once again the problem of reacting to market changes, and does so by means of a periodic retraining, however with a few specificities. Focusing on intraday US stock market (1-minute time granularity), the authors propose to build different models, on different time frames, and to combine the buy or sell signals for the next minute, in the hope to capture reactivity both to the market and to longer trends. In particular, they examine the time frames of 5, 10, 15, and 30 minutes: a combination of signals for shorter and longer time frames creates a more conservative (safer) signal for buying and selling actions. The training is performed as follows: let us illustrate it with an example about the smallest time frame of five minutes. The GP trains on minute $\{1\}$ with minute $\{2\}$ unknown, then on minutes $\{1, 2\}$ with $\{3\}$ unknown, then on minutes $\{1, 2, 3\}$ with $\{4\}$ unknown, and finally on minutes $\{1, 2, 3, 4\}$ with $\{5\}$ unknown, giving a total of four fitness cases. Based on the mean value over all the trade signals of the best individuals after each fitness case, the overall signal for the next unknown minute is determined. A majority voting is applied to combine the signals of the different models. Results on four different stocks are shown to be promising: the combination of signals from models at 5-, 10-, and 15-minute time frames is the one that gives better performance, in general, allowing the system to beat the B&H strategy, in a few cases. However, the choice of the authors to test their system on a single, arbitrary day (18th October 2010) rings hollow, especially in light of the data snooping problem.

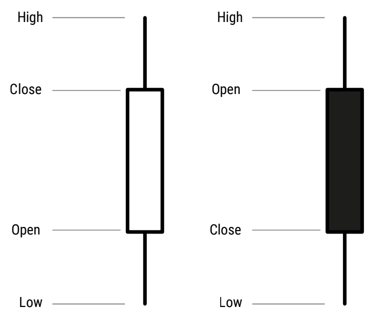

Figure 4.2: The opening, highest, lowest, and closing prices of each day are represented in a single vertical bar, consisting of a thick body and two thinner lines above and below it, called shadows. The area between the open and close prices is the body, and is coloured in white or black, depending on whether the open price is less or greater than the closing price, respectively. The two shadows represent the full price excursion between the highest and lowest price of the day.
From: https://en.wikipedia.org/wiki/Candlestick_chart

4.5 *Full Solutions*

Those studies that do not easily fit the previous sections, because of their proposing more sophisticated or alternative solutions to the problem of predicting financial markets, are reported here. Izumi et al. (2006), for example, apply genetic network programming (GNP), a variation of GP, where solutions are represented as networks (directed graphs). In GNP, nodes are connected by directed edges, which, in the current study, is the only element that is evolved, generation after generation. The goal is to obtain the network with optimised structure by the connections of nodes during evolution. The authors distinguish judgement nodes, where the information from the inputs is evaluated, and processing nodes, that determine the relevant trading action to take. The researchers develop stock trading strategies by combining GNP and the Japanese style of representation of stock prices as candlesticks charts (fig. 4.2). Judgement nodes are responsible for evaluating whether prices display relevant candlestick patterns: for example, it can be meaningful if yesterday's lowest price is higher than today's highest, or whether yesterday's body is black (i.e., the stock closed lower than the opening) and today's is white. At the processing node after judgement nodes, the trade is realised (i.e., buy or sell signals). Notice that the the GNP network structure represents a readable trading logic, consisting of patterns of candlestick charts. In the simulations using stock data of 10 Japanese companies for three different out-of-sample testing years (2002 to 2004), results appear to be promising; however, no comparison with the traditional B&H strategy is provided.

To capture the complex and dynamic behaviour of stock markets, and in the belief that market conditions tend to repeat themselves in patterns, Hsu (2011) proposes an integrated procedure for stock forecasting. In the proposed system, clusters in the sample data are identified, and then several GP models are evolved, one for each cluster. The purpose of clustering is to split the sample data, in such a way that the objects within each cluster are similar to each other and dissimilar to the objects in other clusters; this way, it should be easier for the GP algorithm to find relevant models, because each cluster can be expected to be more homogeneous. To perform clustering, the author uses a self-organizing map (SOM), an unsupervised type of artificial neural network (ANN), that is typically used to produce a low-dimensional, discretised representation of the input space of the training samples. To select the appropriate number of clusters to consider, unknown *a priori*, the common ANOVA test F-statistic is used. Focusing on Taiwan Capitalization Weighted Stock Index (TAIEX) index, the author computes a number of technical indicators, which constitute the input space along with the prices; seven clusters are identified to be optimal. Then, a commercial GP solution is employed to fit the models relative to each cluster, and and predict the next-day index value. Experimental runs over successive periods from 1997 to 2009 show an average mean absolute percentage error (MAPE) well below 2 %.

With the goal of identifying an optimal, low-risk portfolio of European stocks, Contreras et al. (2017) propose an automated trading system based on two, complementary GP-grammars, that combine technical, accounting, and macroeconomic information on companies and countries, in order to identify promising investment opportunities. In this study, companies are framed within industrial sectors and countries, because firms belonging to the same sector are expected feature similar market trends (which, in fact, was proven by the same authors in other studies). The proposed system is structured as follows. An external grammar, which serves as meta-grammar, is used to define candidate portfolios, in terms of country and sector, and secondly of set of companies, within such sector; the definition of portfolios is guided by the evaluation of macroeconomic and accounting variables, such as the valuation of the company (measured, among other ways, as the ratio between the market capitalisation and the net income), the gross domestic product (GDP) or the unemployment rate of the respective country. The internal grammar, instead, evolves the trading rules specific to each company of each portfolio identified by the external grammar. The fitness of the best trading strategy for each company, computed at the end of the internal evolution, contributes to the evaluation of the performance of each portfolio. To assess the performance of the proposed system, the authors collect 11 years of historical data regarding as many as 1000 publicly listed companies in the 28 European countries. The period from 2003 to 2010 is used to train the system, which is then tested on the year 2011; results show that the out-of-sample average return per company in the selected portfolios is 30.14 %, whereas a B&H strategy on the same portfolios had an average return per company of −13.36 %. It is important to notice that the fitness function adopted by the authors is risk-adjusted, and that no validation period is considered in the study.

5 Discussion

In this chapter, we conclude our presentation of genetic programming (GP) studies about the prediction of financial markets, by discussing the general accomplishments and results, as well as shortcomings of the current scientific research. In section 5.1 we summarise the results achieved in the reviewed studies, and briefly discuss them against the Efficient market hypothesis (EMH) presented in chapter 2. Sections 5.2 and 5.3 tackle the issue of discerning whether GP has been applied to the best of its potential, incorporating domain knowledge and fulfilling the motivations and expectations that lead researchers to adopt it as the search algorithm. Finally, section 5.4 presents the proposed future developments for GP-based trading systems.

5.1 Results

Given the wide spectrum of targets (stock prices, stock indexes, foreign exchange rates) and test periods (spanning roughly 25 years, from 1987 to 2012) considered in the studies reviewed in the previous chapter, hardly can general and unequivocal conclusions be drawn, about the actual success of GP applications in the financial markets prediction. Based on the presented results, the GP approach appears to be rather successful, even though no outstanding results are achieved. In most cases, the trading models generated through evolution of GP populations are characterised by some predictive ability, in their being able to correctly identify uptrend and downtrend periods; however, when trading fees and taxes are taken into account, profitability appears less consistently over testing periods. Wang (2000), for example, in his work on daily Standard & Poor's 500 (S&P 500) index, shows evidence that excess returns (i.e., the portion of investment returns that exceed those of the buy and hold strategy) are positive in some years, but negative in others. Likewise, Dempsey, O'Neill and Brabazon (2006) are able to achieve positive (exceptional, in fact) excessive returns, but only when markets showed high volatility and weak uptrend. Svangård et al. (2002) obtain positive results, but under the restrictive and questionable condition of test the profitability of the model on a single stock and arbitrary day. In contrast, the work by Contreras et al. (2017) reaches significant achievements in their selecting a portfolio of stocks and suggesting a trading strategy; although the test is performed on one year only, and not on a rolling basis, as it was suggested, among others, by Wang (2000), the results appear hard to dismiss. However, the work is too recent to have already been scrutinised by other researchers.

An important point should be made, about the benchmark against which trading strategies are evaluated. When profitability is compared with the returns of the buy and hold (B&H) strategy, more than one study evidence that, in periods of small B&H returns, the GP is more likely to produce positive excess returns. Yet, it is important to notice that a few researchers reject the comparison with B&H returns as a valid measure of the profitability of a strategy. Chen and Navet (2007), and Svangård et al. (2002) to some extent, point out that such result should be expected, because B&H can be regarded as the best strategy possible, in strong uptrend periods, and as far from optimal in volatile or downtrending markets. In this sense, benchmarking trading strategies against B&H appears to be limited and misleading; despite this point, most studies compare their performance to the B&H strategy, possibly due to the ease of calculation and immediate understandability.

We are not in the position to tell whether the mixed, but overall mediocre, results are due to the general efficiency of the markets (EMH, section 2.3), or rather of what Timmermann and Granger (2004) define as the 'reverse file drawer bias'. Their point is that a researcher who genuinely believes she has identified a reliable method for predicting the market has little incentive to publish it in an academic journal. Rather, she would presumably be tempted to sell it to an investment bank. Whether it is due to market efficiency or researchers' psychological bias, it is hard to tell, and probably the answer is an opinion rather than a scientific truth to discover. An attempt to formally settle the matter of whether financial time series are predictable using GP (i.e., understand whether the poor results obtained so far are due to market efficiency or inadequacy of the adopted models) was made by Kaboudan (2000) and extended, later, by Chen and Navet (2007). Kaboudan (2000) proposes the η statistic, to measure the probability that a time series (financial, in particular) is GP-predictable. In detail, η is defined as the average over several runs (for statistical consistency) over the ratio between the residual sum of squares (RSS) from the GP prediction of the series under scrutiny and the RSS from its randomly shuffled sequence[1]. The intuition is that if the observed data contained any signal, shuffling dismembers this signal and the two RSS representations will be significantly different; in contrast, they will be very similar if the observed data were pure noise. Kaboudan (2000) is able to conclude that GP can be successful and further research appropriate, by showing that stock prices are likely to be GP-predictable. Building on the work by Kaboudan (2000) and to test the profitability of GP solutions, in addition to the predictability of data series, Chen and Navet (2007) propose some different tests which, when put together, can help the researcher understand whether there are hidden patterns to discover and whether GP is properly designed to do the job. The essential idea underlying all proposed tests is to compare the performance of GP with random trading strategies or behaviour, although constraints are put in place to provide fair and informative comparison. Results confirm the findings of previous and subsequent studies: the efficiency of a market is variable over time, and therefore a trader making their investment decisions by running GP-generated models will not be able to consistently

1. The analytical expression of the statistic η is as follows:

$$\eta = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{RSS_Y}{RSS_S}\right)_i$$

where n is the number of GP runs, RSS_Y and RSS_S the unexplained variance (residual sum of squares) of the series under scrutiny before and after shuffling, respectively.

outperform the market. In light of the tests, the authors suggest that GP may only be able to take advantage of 'simple' regularities in the data.

5.2 Domain Knowledge and Semantics of Solutions

To ensure sensible and feasible models, the GP search algorithm should be theoretically justified, and possibly be enriched with explicit domain knowledge. Domain knowledge can be present at different degrees in the application of GP to the prediction of financial markets, from the simple inclusion of technical indicators in the GP function space, to systematic enforcement of a semantics. Surprisingly few researchers take this path. From tables 4.3 and 4.4, it is evident that only few studies consider technical indicators among the building blocks of the GP candidate solutions. Among those who do, Dempster and Jones (2001) and Izumi et al. (2006) use them with the explicit goal of combining the buy or sell signals generated by the indicators themselves into a profitable strategy. Lee and Moon (2010) take a step further, by implementing subtrees that already capture state-of-the-art interpretation of the indicators at hand; the identified patters are used because deemed to be relatively simple, profitable, and frequent. Bhattacharyya, Pictet and Zumbach (2002) are the only authors who carefully tailor the genetic search algorithm to the specific problem of predicting financial markets, namely foreign exchange rates (FX rates), by enforcing a semantics that could make all the solutions meaningful and relevant; another goal of the authors is to improve the interpretability of the candidate models. Although the adopted measures are not sufficient to guarantee straightforward interpretability of the solutions, their complexity is significantly reduced, their mean size is reduced to less than 20 nodes.

As for the matter of the data used as inputs to the prediction system, the situation is pretty similar. The vast majority of studies report using only prices and trading volume of the financial asset of interest, and possibly a handful of statically computed moving averages of prices themselves. Very few researchers diverge from this path. Svangård et al. (2002) consider feedback variables and macroeconomic indicators such as global FX rates and stock indexes, to predict the movements of Nokia stocks. However, it is only Contreras et al. (2017) who programmatically integrate technical and fundamental analysis in order to build a profitable, low-risk portfolio of European stocks, by means of a grammar-based GP search algorithm. Results appear to be very promising, but, as it was already mentioned, no scrutiny on their study by other researchers is available yet, due to its recency.

5.3 Fulfilment of Motivations and Expectations

In those studies where the choice of adopting GP as the search algorithm for a profitable trading strategy is properly motivated (section 4.2), the reasons are generally confirmed by results. Bhattacharyya, Pictet and Zumbach (2002), for example, are able to significantly reduce the mean complexity (i.e., size) of the evolved solutions, from nearly 30 to less than 20, by enforcing semantic restrictions so that all generated models are meaningful in the domain. Therefore, interpretability is generally, if not always, attained. Likewise, Lee and Moon (2010) are able to greatly limit the size of solutions, which result from the combination of no more than five technical patterns.

Whereas a net minority of studies explicitly mention data snooping bias as a problem to address in their approach to the problem of financial forecasting, still most of the reviewed research papers adopt measures that do guard against such risk, as well as detect overfitting situations (table 4.2). Validation data is adopted in most cases, in order to filter the solutions that are regarded as best, by validating them on an unseen portion of data before accepting them as new best solutions. Moreover, models are often tested on several, non-overlapping periods of time, in order to evaluate their performance over time and to more robustly assess the ability of GP to capture profitable market strategies.

The discussion about the adaptation of trading systems to real-time changes in the market behaviour (section 4.4) made it clear that the matter is not fully settled. Indeed, a number of authors provided different results and conclusions. Wang (2000) obtained inconsistent performance with a periodic retraining of the system, while Lee and Moon (2010) reached positive results with the same method; in a comparison of different approaches Dempster and Jones (2001) showed that a static training (i.e., a single training at the beginning of the test period) performs better than reactive or periodic retraining (but still the results are less profitable than the buy and hold strategy), whereas a reactive retraining is by far the best approach according to Loginov and Heywood (2013). Such mixed conclusions may well be the result of testing the different approaches on different years, but ultimately are evidence of the fact that further research is needed, to understand both the real capabilities of GP-evolved trading models and the intimate functioning of financial markets themselves.

5.4 Proposed Future Developments

As a general rule, researchers appear to be realistic about the limitations of their studies, and propose a number of different solutions. We summarise them, distinguishing between enhanced adaptation to real-life trading and technical improvements. To make their proposed forecasting or trading systems more suitable for a real-life adoption in financial contexts, researchers suggest the following steps:

- Fitness function is the mean to convey in the genetic search algorithm the concept of trading risk, thus making the evolved trading model risk-aware and inclined to suggest more conservative (and safer) strategies; a careful and deeper study of fitness functions, where to implement risk-adjustments, possibly in a multi-objective setting, is suggested (Bhattacharyya, Pictet and Zumbach, 2002; Contreras et al., 2017).

- In real-life settings, traders commonly adopt money management techniques that include a full disinvestment in case a loss is registered or when the stock trades below a certain price (stop-loss order); such strategies are able to greatly increase returns, by significantly limiting losses that would erode the obtained gains, but nevertheless seldom are such strategies considered in the evaluation of the performance of an automated strategy (Contreras et al., 2017).

- Rather than building trading models on single stocks or FX rates, it may be appropriate to and extend decision models to portfolios, in order to compensate small losses in single assets and therefore reduce transactions, which are costly due to fees (Svangård et al., 2002); alternatively, it may be relevant to seek well-performing models across multiple data series at a time, to reduce overfitting and improve robustness (Bhattacharyya, Pictet and Zumbach, 2002).

- To increase the amount of information available to the model, in an attempt to capture long-term and short-term trends and signals, data at various frequencies can be considered, for example by computing technical indicators at different frequencies (Dempster and Jones, 2001).

- To better capture information contained in the stock prices (or FX rates) data series, as well as improve the interpretability of the models, the set of considered technical indicators and patterns can be expanded (Lee and Moon, 2010; Contreras et al., 2017).

On the technical side, that is regarding the rationale for the implementation of the trading system, we can mention the following proposed future developments:

- Along with a needed data enrichment, with the possible introduction of multiple technical indicators and information at different time frequencies, some feature selection activity may be needed as well, to reduce noise and retain only relevant information (Hsu, 2011).

- Combining trading signals using standard ('hard') logic can result in inaccurate signalling, due to data noise and lack of signal weighting and strength; the definition of logic operators with smoother transitions, similar to fuzzy logic, can provide a more relevant and better-performing combination of trading signals (Bhattacharyya, Pictet and Zumbach, 2002).

- In light of the fact that market timing can be detected, but profitability does not appear consistently across time and assets, greater care should be placed in the selection of the financial assets to consider (Wang, 2000)

- Especially when discussing the issue of adapting the trading system to the evolving dynamics of stock and foreign exchange markets (FX markets), research appears to be far from conclusive (section 4.4); further investigation should be conducted, to fix and improve the tradeoff between the exploration of new solutions and exploitation of currently profitable models; unanswered questions regard, among others, the triggering of a new training phase and the reuse of previously evolved solutions (Loginov and Heywood, 2013).

6 Conclusion

In this survey work, we explored the possibility of predicting financial markets using machine learning techniques, with a focus on the application of GP to the estimation of future stock prices, stock indexes, and FX rates, or even the definition of profitable trading strategies. Financial markets are complex systems, the study of which has spanned over the years, both from a theoretical standpoint and from a more practical perspective. The theoretical discussion shows a contrast, sharp at times, between those who support EMH, a longstanding theory that provides an essential unpredictability of financial markets, and those who contrast it, by showing evidence of positive excess returns attained on the basis of analytical models developed over the years. Fundamental analysis, technical analysis, and behavioural analysis are the best-known such analytical models; respectively, they focus on the investigation of economic and company factors, on historical price prices, and on market agents' sentiment to gain a valuable insight on upcoming market changes, so as to decide when it is appropriate to take market actions.

Building mainly on technical analysis, the approach that is more suitable for a computer analysis, computer scientists over the years applied different machine learning techniques to capture market patterns and produce relevant predictions and strategies. Based on our research, the machine learning techniques that most commonly are applied to the daily forecast of stock prices, stock indexes, and FX rates are GP and artificial neural networks (ANNs), with fuzzy logic-based methods gaining popularity in recent years. Results of the different techniques are comparable, in that they are mixed and inadequate to establish a clear outperforming method. Rather, our choice of focusing our attention on and deepen the study of GP applications to the problem is guided by the strengths and weaknesses of each method, and also by the characteristics of the proposing literature, to a limited extent. GP solutions are human-readable, which is of paramount importance in the financial field, and the search process itself is a close simulation of the behaviour of a trader, who continuously fixes her trading strategy until some quality criteria (e.g., profit, possibly risk-adjusted) is met. ANNs, however promising, are black boxes; moreover, the research in the field of ANNs is so vast and unstructured[1] that surveying studies that apply them to financial prediction may well turn out to be a Sisyphean task. As for fuzzy logic-based methods, despite the increasing popularity, the scientific research on their application to financial prediction appears to

1. In terms, for example, of different architectures, topologies, but also of motivations.

be less mature than in the other two main cases.

In light of the above, we selected GP applications to focus on. In all the reviewed research papers, GP is used as the search algorithm to find profitable trading strategies in stock and foreign exchange markets, or, less frequently, to make an outright prediction of stock prices, stock indexes, or FX rates. We present and discuss the different applications from a number of different perspectives, starting from the motivations that guided researchers to adopt GP, then moving to the definition of inputs and building blocks of the GP algorithm, and concluding with the solutions adopted to deal with the issue of adapting the trading models to the dynamic conditions of financial markets. The motivations that guided the choice of GP as the search algorithm are essentially two: the interpretability of the candidate models, net of their size, and the mitigation of the data snooping bias, a subtle issue that is tightly linked to a naive reuse of data for inference, even among different studies. Even though the studies that mention either motivation are not many, the goals are generally fulfilled. As for the implementation of domain knowledge, of which the definition of inputs and GP primitives is a representative aspect, we notice that most studies strictly stick to a technical analysis approach, using as exclusive inputs past stock prices and trading volume. While arithmetic and logical functions are the most frequent primitives to the GP algorithm, technical indicators, common mathematical instruments in use among practitioners, are sometimes used; the implementation of technical indicators is usually proposed to improve the readability of the candidate models, as they have well-defined and agreed upon meaning in the financial field. Designing a prediction system that is able to adapt itself to the dynamic changes of market conditions is still an open issue, in that different approaches have been proposed, and different tests yielded significantly different results; the most common way of dealing with the problem is re-evolving the population: this may happen after some quality criteria are no longer met (reactive re-training), or periodically, either starting from a whole new population or continuing the evolution of the current population.

All things considered, genetic programming (GP) appears to be a relevant approach to the prediction of financial markets, with clear advantages over other methods. The research conducted so far is rather solid, and the open issues and proposed future developments well-defined.

Ringraziamenti

Alla fine di questo lavoro, che segna peraltro la fine del mio percorso accademico, è doveroso esprimere i miei debiti di riconoscenza, anche se difficilmente queste poche righe saranno sufficienti.

Partiamo dai ringraziamenti personali, a quelle persone che mi hanno accompagnato in questi cinque anni di crescita e scoperta del mondo. Camilla, Eleonora, Melinda e Deborah, voi siete sempre riuscite a ricordarmi che la vita non è composta solo di numeri e bit: la vostra presenza è, e sarà sempre, insostituibile. Andrea e Laura, Carolina e Cecilia, voi avete alleggerito i cinque anni politecnici che ci hanno visti insieme, lezione dopo lezione. Davide e Fabio, ciascuno nel suo ruolo, voi mi avete donato momenti bellissimi, e altri assolutamente dimenticabili: per entrambi vi ringrazio, perché mi avete aiutato a crescere. Ciascuno di voi, dieci indispensabili persone, mi ha aiutato a crescere, e a voi va la mia riconoscenza.

Un ringraziamento speciale, inoltre, va a Raffaella ed Emilio, genitori attenti e moderni, perché senza il vostro sostegno non avrei mai potuto arrivare dove sono ora.

Il capitolo dei ringraziamenti professionali, cioè delle persone che mi hanno accompagnato alla conclusione degli studi, si apre con il professor Pier Luca Lanzi, il quale ha svolto con giusta serietà il ruolo di relatore per questa tesi, dandomi al contempo l'impareggiabile opportunità di svolgere questo lavoro in un contesto aziendale, quale Business Integration Partners. Le tante lezioni imparate, prima e durante lo svolgimento di questo lavoro, mi accompagneranno a lungo. Segue, nella ruota dei riconoscimenti, Antonio, pacata voce autorevole anche in situazioni complesse. Ultimi in queste righe ma sicuramente non per importanza, sono Andrea e Danilo, per la fiducia che hanno risposto in me in questi mesi: con il loro sostegno e aiuto, ho potuto completare questo lavoro.

Acronyms

JPY Japanese yen. 35, 36, 42

KOSPI Korea Composite Stock Price Index. 34, 42, 50

MAPE mean absolute percentage error. 35, 37, 55

Nikkei 225 Nikkei Stock Average. 42, 52, 53

PCA principal component analysis. 35

PSO particle swarm optimization. 36

RBFN radial basis function network. 35, 36

RSS residual sum of squares. 58

S&P 500 Standard & Poor's 500. 18, 24, 29, 31, 42, 49, 52, 53, 57

SOM self-organizing map. 55

STGP strongly typed genetic programming. 32, 43, 46

STNN stochastic time effective neural network. 35

TAIEX Taiwan Capitalization Weighted Stock Index. 42, 55

TDNN time delay neural network. 35

TWSE Taiwan Stock Exchange. 37

UK United Kingdom. 28

US United States. 19, 21, 27, 28, 30, 34, 35, 37–39, 54

USD US dollar. 32, 35, 36, 42, 46, 53

VIX Volatility Index. 24

Bibliography

Primary References

Allen, Franklin, and Risto Karjalainen. 1999. 'Using genetic algorithms to find technical trading rules'. *Journal of Financial Economics* 51, no. 2 (February): 245–271. DOI: 10.1016/S0304-405X(98)00052-X.

Atsalakis, George S., Eftychios E. Protopapadakis and Kimon P. Valavanis. 2016. 'Stock trend forecasting in turbulent market periods using neuro-fuzzy systems'. *Operational Research* 16, no. 2 (July): 245–269. DOI: 10.1007/s12351-015-0197-6.

Atsalakis, George S., and Kimon P. Valavanis. 2009a. 'Forecasting stock market short-term trends using a neuro-fuzzy based methodology'. *Expert Systems with Applications* 36, no. 7 (September): 10696–10707. DOI: 10.1016/j.eswa.2009.02.043.

———. 2009b. 'Surveying stock market forecasting techniques – Part II: Soft computing methods'. *Expert Systems with Applications* 36, no. 3 (April): 5932–5941. DOI: 10.1016/j.eswa.2008.07.006.

Baker, Malcolm, and Jeffrey Wurgler. 2007. 'Investor Sentiment in the Stock Market'. *Journal of Economic Perspectives* 21, no. 2 (April): 129–151. DOI: 10.1257/jep.21.2.129.

Balance of payments and international investment position manual. 2008. Sixth edition. International Monetary Fund (IMF), December.

Barberis, Nicholas, and Richard H. Thaler. 2002. 'A Survey of Behavioral Finance'. Chap. 18 in *Handbook of the Economics of Finance,* 1st ed., edited by G. M. Constantinides, M. Harris and R. M. Stulz, 1:1053–1128. Cambridge, MA: Elsevier, September. DOI: 10.3386/w9222.

Bhattacharyya, S., Olivier V. Pictet and G. Zumbach. 2002. 'Knowledge-intensive genetic discovery in foreign exchange markets'. *IEEE Transactions on Evolutionary Computation* 6, no. 2 (April): 169–181. DOI: 10.1109/4235.996016.

Brock, William, Josef Lakonishok and Blake LeBaron. 1992. 'Simple Technical Trading Rules and the Stochastic Properties of Stock Returns'. *The Journal of Finance* 47, no. 5 (December): 1731–1764.
DOI: 10.1111/j.1540-6261.1992.tb04681.x.

Cavalcante, Rodolfo C., Rodrigo C. Brasileiro, Victor L. F. Souza, Jarley P. Nobrega and Adriano L. I. Oliveira. 2016. 'Computational Intelligence and Financial Markets: A Survey and Future Directions'. *Expert Systems with Applications* 55 (August): 194–211.
DOI: 10.1016/j.eswa.2016.02.006.

Chang, Pei-Chann, and Chen-Hao Liu. 2008. 'A TSK type fuzzy rule based system for stock price prediction'. *Expert Systems with Applications* 34, no. 1 (January): 135–144.
DOI: 10.1016/j.eswa.2006.08.020.

Contreras, Iván, J. Ignacio Hidalgo, Laura Nuñez-Letamendía and J. Manuel Velasco. 2017. 'A meta-grammatical evolutionary process for portfolio selection and trading'. *Genetic Programming and Evolvable Machines* 18, no. 4 (December): 411–431.
DOI: 10.1007/s10710-017-9304-1.

Dempsey, Ian, Michael O'Neill and Anthony Brabazon. 2006. 'Adaptive Trading With Grammatical Evolution'. In *2006 IEEE International Conference on Evolutionary Computation,* 2587–2592. CEC06. IEEE.
DOI: 10.1109/CEC.2006.1688631.

Dempster, M. A. H., and C. M. Jones. 2001. 'A real-time adaptive trading system using genetic programming'. *Quantitative Finance* 1, no. 4 (April): 397–413.
DOI: 10.1088/1469-7688/1/4/301.

Diebold, Francis X., and James A. Nason. 1990. 'Nonparametric exchange rate prediction?' *Journal of International Economics* 28, nos. 3-4 (May): 315–332.
DOI: 10.1016/0022-1996(90)90006-8.

Fama, Eugene F. 1970. 'Efficient Capital Markets: A Review of Theory and Empirical Work'. *The Journal of Finance* 25 (2): 383–417.

Gençay, Ramazan. 1998a. 'Optimization of technical trading strategies and the profitability in security markets'. *Economics Letters* 59, no. 2 (May): 249–254.
DOI: 10.1016/S0165-1765(98)00051-2.

———. 1998b. 'The predictability of security returns with simple technical trading rules'. *Journal of Empirical Finance* 5, no. 4 (October): 347–359.
DOI: 10.1016/S0927-5398(97)00022-4.

Gençay, Ramazan, and Thanasis Stengos. 1998. 'Moving average rules, volume and the predictability of security returns with feedforward networks'. *Journal of Forecasting* 17, nos. 5-6 (September): 401–414.
DOI: 10.1002/(SICI)1099-131X(1998090)17:5/6<401::AID-FOR704>3.0.CO;2-C.

Hsu, Chih-Ming. 2011. 'A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming'. *Expert Systems with Applications* (May).
DOI: 10.1016/j.eswa.2011.04.210.

Izumi, Yoshihiro, Tokiyo Yamaguchi, Shingo Mabu, Kotaro Hirasawa and Jingle Hu. 2006. 'Trading Rules on the Stock Markets using Genetic Network Programming with Candlestick Chart'. In *2006 IEEE International Conference on Evolutionary Computation,* 2362–2367. CEC06. IEEE.
DOI: 10.1109/CEC.2006.1688600.

Kilian, Lutz, and Cheolbeom Park. 2009. 'The Impact of Oil Price Shocks on the U.S. Stock Market'. *International Economic Review* 50, no. 4 (November): 1267–1287.
DOI: 10.1111/j.1468-2354.2009.00568.x.

Kim, Kyoung-jae, and Ingoo Han. 2000. 'Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index'. *Expert Systems with Applications* 19, no. 2 (August): 125–132.
DOI: 10.1016/S0957-4174(00)00027-0.

Kuo, R. J., C. H. Chen and Y. C. Hwang. 2001. 'An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network'. *Fuzzy Sets and Systems* 118, no. 1 (February): 21–45.
DOI: 10.1016/S0165-0114(98)00399-6.

Kwon, Yung-Keun, and Byung-Ro Moon. 2007. 'A Hybrid Neurogenetic Approach for Stock Forecasting'. *IEEE Transactions on Neural Networks* 18, no. 3 (May): 851–864.
DOI: 10.1109/TNN.2007.891629.

LeBaron, Blake. 1999. 'Technical trading rule profitability and foreign exchange intervention'. *Journal of International Economics* 49, no. 1 (October): 125–143.
DOI: 10.1016/S0022-1996(98)00061-0.

Lee, Seung-Kyu, and Byung-Ro Moon. 2010. 'A new modular genetic programming for finding attractive technical patterns in stock markets'. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation,* 1219–1226. GECCO '10. New York, New York, USA: ACM Press.
DOI: 10.1145/1830483.1830704.

Lo, Andrew W., Harry Mamaysky and Jiang Wang. 2000. 'Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation'. *The Journal of Finance* 55, no. 4 (August): 1705–1765.
DOI: 10.1111/0022-1082.00265.

Loginov, Alexander, and Malcolm I. Heywood. 2013. 'On the impact of streaming interface heuristics on GP trading agents'. In *Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation,* 1341–1348. GECCO '13. New York, New York, USA: ACM Press. DOI: 10.1145/2463372.2463522.

Malkiel, Burton G. 2003. 'The Efficient Market Hypothesis and Its Critics'. *Journal of Economic Perspectives* 17, no. 1 (March): 59–82. DOI: 10.1257/089533003321164958.

———. 2012. *A Random Walk Down Wall Street.* 10th ed. W. W. Norton & Company.

Martinez-Jaramillo, Serafin, and Edward P. K. Tsang. 2009. 'An Heterogeneous, Endogenous and Coevolutionary GP-Based Financial Market'. *IEEE Transactions on Evolutionary Computation* 13, no. 1 (February): 33–55. DOI: 10.1109/TEVC.2008.2011401.

Montana, David J. 1995. 'Strongly Typed Genetic Programming'. *Evolutionary Computation* 3, no. 2 (June): 199–230. DOI: 10.1162/evco.1995.3.2.199.

Murphy, John J. 1999. *Technical Analysis of the Financial Markets.* 2nd ed. New York Institute of Finance.

Neely, Christopher J. 2002. 'The temporal pattern of trading rule returns and exchange rate intervention: intervention does not generate technical trading profits'. *Journal of International Economics* 58, no. 1 (October): 211–232. DOI: 10.1016/S0022-1996(01)00163-5.

———. 2003. 'Risk-adjusted, ex ante, optimal technical trading rules in equity markets'. *International Review of Economics and Finance* 12, no. 1 (March): 69–87. DOI: 10.1016/S1059-0560(02)00129-6.

Neely, Christopher J., Paul A. Weller and Rob Dittmar. 1997. 'Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach'. *The Journal of Financial and Quantitative Analysis* 32, no. 4 (December): 405–426. DOI: 10.2307/2331231.

Olson, Dennis. 2004. 'Have trading rule profits in the currency markets declined over time?' *Journal of Banking and Finance* 28, no. 1 (January): 85–105. DOI: 10.1016/S0378-4266(02)00399-0.

Ou, Jane A., and Stephen H. Penman. 1989. 'Financial statement analysis and the prediction of stock returns'. *Journal of Accounting and Economics* 11 (4): 295–329. DOI: 10.1016/0165-4101(89)90017-7.

Park, Cheol-Ho, and Scott H. Irwin. 2007. 'What do we know about the profitability of technical analysis?' *Journal of Economic Surveys* 21, no. 4 (September): 786–826.
DOI: 10.1111/j.1467-6419.2007.00519.x.

Potvin, Jean-Yves, Patrick Soriano and Maxime Vallée. 2004. 'Generating trading rules on the stock markets with genetic programming'. *Computers & Operations Research* 31, no. 7 (June): 1033–1047.
DOI: 10.1016/S0305-0548(03)00063-7.

Saad, Emad W., Danil V. Prokhorov and Donald C. Wunsch. 1998. 'Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks'. *IEEE Transactions on Neural Networks* 9 (6): 1456–1470.
DOI: 10.1109/72.728395.

Sermpinis, Georgios, Konstantinos Theofilatos, Andreas Karathanasopoulos, Efstratios F. Georgopoulos and Christian Dunis. 2013. 'Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization'. *European Journal of Operational Research* 225, no. 3 (March): 528–540.
DOI: 10.1016/j.ejor.2012.10.020.

Shiller, Robert J. 2003. 'From Efficient Markets Theory to Behavioral Finance'. *Journal of Economic Perspectives* 17, no. 1 (March): 83–104.
DOI: 10.1257/089533003321164967.

———. 2015. *Irrational Exuberance.* 3rd ed. Princeton University Press.

Svangård, Nils, Peter Nordin, Stefan Lloyd and Clas Wihlborg. 2002. 'Evolving short-term trading strategies using genetic programming'. In *Proceedings of the 2002 Congress on Evolutionary Computation,* 2:2006–2010. CEC02. IEEE.
DOI: 10.1109/CEC.2002.1004551.

Taylor, Stephen J. 2000. 'Stock index and price dynamics in the UK and the US: new evidence from a trading rule and statistical analysis'. *The European Journal of Finance* 6, no. 1 (March): 39–69.
DOI: 10.1080/135184700336955.

Thaler, Richard H. 1999. 'The End of Behavioral Finance'. *Financial Analysts Journal* 55, no. 6 (November): 12–17.
DOI: 10.2469/faj.v55.n6.2310.

Timmermann, Allan, and Clive W. J. Granger. 2004. 'Efficient market hypothesis and forecasting'. *International Journal of Forecasting* 20 (1): 15–27.
DOI: 10.1016/S0169-2070(03)00012-8.

Wang, Jie, and Jun Wang. 2015. 'Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks'. *Neurocomputing* 156 (May): 68–78.
DOI: 10.1016/j.neucom.2014.12.084.

Wang, Jun. 2000. 'Trading and hedging in S&P 500 spot and futures markets using genetic programming'. *Journal of Futures Markets* 20, no. 10 (November): 911–942.
DOI: 10.1002/1096-9934(200011)20:10<911::AID-FUT3>3.0.CO;2-K.

White, Halbert. 1988. 'Economic prediction using neural networks: the case of IBM daily stock returns'. In *IEEE International Conference on Neural Networks,* 2:451–458. IEEE.
DOI: 10.1109/ICNN.1988.23959.

Wilson, Garnett, Derek Leblanc and Wolfgang Banzhaf. 2011. 'Stock trading using linear genetic programming with multiple time frames'. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation,* 1667. GECCO '11. New York, New York, USA: ACM Press.
DOI: 10.1145/2001576.2001801.

Yao, JingTao, and Chew Lim Tan. 2000. 'A case study on using neural networks to perform technical forecasting of forex'. *Neurocomputing* 34, nos. 1-4 (September): 79–98.
DOI: 10.1016/S0925-2312(00)00300-3.

Yu, Lean, Kin Keung Lai and Shouyang Wang. 2008. 'Multistage RBF neural network ensemble learning for exchange rates forecasting'. *Neurocomputing* 71, nos. 16-18 (October): 3295–3302.
DOI: 10.1016/j.neucom.2008.04.029.

Other References

Banzhaf, Wolfgang, Peter Nordin, Robert E. Keller and Frank D. Francone. 1998. *Genetic Programming: An Introduction. On the Automatic Evolution of Computer Programs and its Applications.* San Francisco, CA, USA: Morgan Kaufmann, January.

Bollen, Johan, Huina Mao and Xiaojun Zeng. 2011. 'Twitter mood predicts the stock market'. *Journal of Computational Science* 2, no. 1 (February): 1–8.
DOI: 10.1016/j.jocs.2010.12.007.

Brenner, Thomas. 2006. 'Agent Learning Representation: Advice on Modelling Economic Learning'. Chap. 18 in *Handbooks in Economics,* edited by Leigh Tesfatsion and Kenneth L. Judd, 13:895–947. 2. North-Holland.
DOI: 10.1016/S1574-0021(05)02018-6.

Brown, Gregory W., and Michael T. Cliff. 2004. 'Investor sentiment and the near-term stock market'. *Journal of Empirical Finance* 11, no. 1 (January): 1–27.
DOI: 10.1016/j.jempfin.2002.12.001.

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay. 1996. *The Econometrics of Financial Markets.* Princeton University Press.

Cao, Melanie, and Jason Wei. 2005. 'Stock market returns: A note on temperature anomaly'. *Journal of Banking and Finance* 29, no. 6 (June): 1559–1573.
DOI: 10.1016/j.jbankfin.2004.06.028.

Chen, Shu-Heng, and Nicolas Navet. 2007. 'Failure of Genetic-Programming Induced Trading Strategies: Distinguishing between Efficient Markets and Inefficient Algorithms'. In *Computational Intelligence in Economics and Finance,* edited by Shu-Heng Chen, Paul P. Wang and Tzu-Wen Kuo, 2:169–182. Berlin, Heidelberg: Springer Berlin Heidelberg.
DOI: 10.1007/978-3-540-72821-4_11.

Day, Theodore E., and Pingying Wang. 2002. 'Dividends, nonsynchronous prices, and the returns from trading the Dow Jones Industrial Average'. *Journal of Empirical Finance* 9, no. 4 (November): 431–454.
DOI: 10.1016/S0927-5398(02)00004-X.

Dougal, Casey, Joseph Engelberg, Diego García and Christopher A. Parsons. 2012. 'Journalists and the Stock Market'. *Review of Financial Studies* 25, no. 3 (March): 639–679.
DOI: 10.1093/rfs/hhr133.

Edmans, Alex, Diego García and Øyvind Norli. 2007. 'Sports Sentiment and Stock Returns'. *The Journal of Finance* 62, no. 4 (August): 1967–1998.
DOI: 10.1111/j.1540-6261.2007.01262.x.

Gençay, Ramazan. 1999. 'Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules'. *Journal of International Economics* 47, no. 1 (February): 91–107.
DOI: 10.1016/S0022-1996(98)00017-8.

Gervais, Simon, Ron Kaniel and Dan H. Mingelgrin. 2001. 'The High-Volume Return Premium'. *The Journal of Finance* 56, no. 3 (June): 877–919.
DOI: 10.1111/0022-1082.00349.

Goodhart, Charles A. E., and Maureen O'Hara. 1997. 'High frequency data in financial markets: Issues and applications'. *Journal of Empirical Finance* 4, nos. 2-3 (June): 73–114.
DOI: 10.1016/S0927-5398(97)00003-0.

Hirshleifer, David, and Tyler Shumway. 2003. 'Good Day Sunshine: Stock Returns and the Weather'. *The Journal of Finance* 58, no. 3 (June): 1009–1032.
DOI: 10.1111/1540-6261.00556.

IFRS 9: Financial Instruments. 2014. International Accounting Standards Board (IASB), July.

Kaboudan, M. A. 2000. 'Genetic programming prediction of stock prices'. *Computational Economics* 16, no. 3 (December): 207–236.
DOI: 10.1023/A:1008768404046.

Kaplanski, Guy, and Haim Levy. 2010. 'Sentiment and stock prices: The case of aviation disasters'. *Journal of Financial Economics* 95, no. 2 (February): 174–201.
DOI: 10.1016/j.jfineco.2009.10.002.

Kilian, Lutz, and Mark P. Taylor. 2003. 'Why is it so difficult to beat the random walk forecast of exchange rates?' *Journal of International Economics* 60, no. 1 (May): 85–107.
DOI: 10.1016/S0022-1996(02)00060-0.

Koza, John R. 1992. *Genetic Programming - On the Programming of Computers by Means of Natural Selection.* MIT Press.

Lo, Andrew W., and A. Craig MacKinlay. 1990. 'Data-Snooping Biases in Tests of Financial Asset Pricing Models'. *Review of Financial Studies* 3, no. 3 (July): 431–467.
DOI: 10.1093/rfs/3.3.431.

———. 1999. *A Non-Random Walk Down Wall Street.* Princeton University Press.

Lukac, Louis P., B. Wade Brorsen and Scott H. Irwin. 1988. 'A test of futures market disequilibrium using twelve different technical trading systems'. *Applied Economics* 20, no. 5 (May): 623–639.
DOI: 10.1080/00036848800000113.

Luke, Sean, and Liviu Panait. 2002. 'Fighting Bloat with Nonparametric Parsimony Pressure'. In *Parallel Problem Solving from Nature — PPSN VII: 7th International Conference Granada, Spain, September 7–11, 2002 Proceedings,* edited by Juan Julián Merelo Guervós, Panagiotis Adamidis, Hans-Georg Beyer, Hans-Paul Schwefel and José-Luis Fernández-Villacañas, 411–421. Berlin, Heidelberg: Springer Berlin Heidelberg.
DOI: 10.1007/3-540-45712-7_40.

Maver, W. 1892. *American Telegraphy: Systems, Apparatus, Operation.*

Meese, Richard A., and Kenneth Rogoff. 1983a. 'Empirical exchange rate models of the Seventies. Do they fit out of sample?' *Journal of International Economics* 14, nos. 1-2 (February): 3–24.
DOI: 10.1016/0022-1996(83)90017-X.

———. 1983b. 'The out-of-sample failure of empirical exchange rate models. Sampling error or misspecification?' In *Exchange Rates and International Macroeconomics,* edited by Jacob A. Frenkel, 67–112. University of Chicago Press.

Müller, Ulrich A., Michel M. Dacorogna, Richard B. Olsen, Olivier V. Pictet, Matthias Schwarz and Claude Morgenegg. 1990. 'Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis'. *Journal of Banking and Finance* 14, no. 6 (December): 1189–1208.
DOI: 10.1016/0378-4266(90)90009-Q.

Neely, Christopher J., and Paul A. Weller. 2001. 'Technical analysis and central bank intervention'. *Journal of International Money and Finance* 20, no. 7 (December): 949–970. DOI: 10.1016/S0261-5606(01)00033-X.

Preis, Tobias, Helen Susannah Moat and H. Eugene Stanley. 2013. 'Quantifying Trading Behavior in Financial Markets Using Google Trends'. *Scientific Reports* 3, no. 1 (December): 1684. DOI: 10.1038/srep01684.

Ratner, Mitchell, and Ricardo P. C. Leal. 1999. 'Tests of technical trading strategies in the emerging equity markets of Latin America and Asia'. *Journal of Banking and Finance* 23, no. 12 (December): 1887–1905. DOI: 10.1016/S0378-4266(99)00042-4.

Simon, Herbert A. 1955. 'A Behavioral Model of Rational Choice'. *The Quarterly Journal of Economics* 69, no. 1 (February): 99. DOI: 10.2307/1884852.

Sullivan, Ryan, Allan Timmermann and Halbert White. 1999. 'Data-Snooping, Technical Trading Rule Performance, and the Bootstrap'. *The Journal of Finance* 54, no. 5 (October): 1647–1691. DOI: 10.1111/0022-1082.00163.

———. 2003. 'Forecast evaluation with shared data sets'. *International Journal of Forecasting* 19, no. 2 (April): 217–227. DOI: 10.1016/S0169-2070(01)00140-6.

Zadeh, Lotfi A. 1973. 'Outline of a New Approach to the Analysis of Complex Systems and Decision Processes'. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (1): 28–44. DOI: 10.1109/TSMC.1973.5408575.