The Application of Intelligent Systems to Financial Time Series Analysis

Martin Victor Sewell

Thesis submitted for the degree of
Doctor of Philosophy
of
UCL

Department of Computer Science
University College London
University of London

3 May 2012
I, Martin Victor Sewell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.


Abstract

This thesis investigates the application of intelligent systems to financial time series analysis. The research is motivated by the following thesis question: ‘Can one improve upon the state of the art in financial time series analysis through the application of intelligent systems?’ The work is split according to the following time series trichotomy: 1) characterization — determine the fundamental properties of the time series; 2) modelling — find a description that accurately captures features of the long-term behaviour of the system; and 3) forecasting — accurately predict the short-term evolution of the system.

The research on characterization comprises of three experiments. In a test for long memory, my implementation of Hurst’s rescaled range ($R/S$) analysis (in C++) found little evidence of long memory in US stock market returns. In tests for dependence, runs tests are applied to US stock market returns, and some surprising dependence found. In a test of market efficiency, the performance of investment newsletters is analysed, evidencing weak-form efficiency. All three experiments test for market efficiency, and impart domain knowledge vis-à-vis financial time series for the work on forecasting.

The work on modelling utilizes behavioural finance to 1) model market action and 2) model investors’ risk preferences. For the former, the evolved heuristics and biases exhibited by fundamental analysts and technical analysts are used to build an agent-based artificial stock market (in Excel). The proportion of technical analysts is varied and the statistics of the time series generated by the artificial market analysed. In the second part, I devise and implement (in PHP and VB) an investment performance measurement metric, cumulative prospect theory certainty equivalent (CPTCE).

In the research on forecasting financial markets, a hidden Markov model (HMM) is trained on foreign exchange data to derive a Fisher kernel (which I implement in C++) for a support vector machine (SVM), and the DC (difference of convex functions) algorithm and the Bayes point machine are also used to create kernels. Furthermore, the DC algorithm is used to learn the parameters of the HMM in the Fisher kernel. I ported two implementations of SVMs to Windows and also added semi-automated parameter selection. SVMdark is written in C for Win32, and winSVM in C++ for Win32.

The thesis makes several novel contributions to science (in both computer science and finance). The work on characterization resulted in an improved characterization of financial time series, with implications for market efficiency. The research on modelling led to an insightful artificial stock market built from first principles, plus a novel investment performance measurement metric, CPTCE. The experiments on forecasting resulted in a full implementation of the Fisher kernel, plus the new DC algorithm-Fisher kernel hybrid algorithm.
Firstly, I wish to thank my supervisor Philip Treleaven for providing me with the opportunity to embark upon a PhD in the first place. Thanks also to David Corney for acting as a mentor. Thanks to John Campbell for sparking my interest in multiagent systems. For general help within the department I thank Chris Clack, and for friendship and collaboration Wei Yan. Thanks to Hurst Consulting and INLECOM for sponsorship. Thanks to Salus Alpha, Fortress Investment Group, Anova Fund, Global Advisors and eStats Capital for providing data. Thanks to Brian McIntosh and Paul Jeffrey of The School of Water Sciences at Cranfield University for employment, and thanks to Christopher Lucas for work with Cirrus Capital.

Thanks to my UCL drinking companions Robert, David, Emma B, Miles, Michael, Dirk, Gordon, Bill, Tim and Emma T for many great evenings, and thanks to Robert and Nancy for helping me stop. Thanks to Julie D., Philippa, Emma, Michael, Nancy, John & Zoë, Vanessa, Petya, Christophe, Valerie, Sauminh, Bill, Carol, Amy, Monica, Wei, Niklas, Lai Nee, Tristan, Tom, Jon, Vicki, Julie Z., Qi, Florence, Julie K., Kiri, Jas, Maya and my parents for feeding me at various times. Many thanks to Nancy for always being there. Thanks to Carol for accompanying me on several trips to the cinema. Thanks to Vicki for many pleasant days out. Eternal thanks to Zac Harland for sharing his unbridled enthusiasm and extensive knowledge of trading over numerous and regular coffees in the Gower Street bookshop. Thanks to my clubbing friends Dave, Amanda, Jim, Rose, Rosie, et al. for festivals, fun and keeping me young.

Thank you to Robert Dodier for convincing me of the superiority of probability right from the start. Thanks to David Wolpert for quick and helpful replies to my queries on the no free lunch theorems, to Andrew Lo for clarifying the EMH and to Eugene Fama for a historical correction. Thanks to our ex-Head of Department, Anthony Finkelstein, for having faith in me. For comments on drafts of this thesis, I thank David Corney, Monica Diekmann-Kotzee, Anthony Finkelstein, John Shawe-Taylor, Christophe Cuny, Niklas Traub, Bill Langdon, Zac Harland, Miles Hansard, Robin Hankin and Professor Treleaven. A special thanks are due to Professor Shawe-Taylor for algorithmic suggestions and thanks to Professor Treleaven for guidance on the home straight and helping me turn a book into a thesis. Thanks also to David Barber, Edward Tsang and anonymous reviewers for various corrections and suggestions. With a view to future work, thanks to Adam Johansen for help regarding particle filters and thanks for communications with Frank Lad, surely the purest subjectivist Bayesian since de Finetti. Thanks to Wei Si for spotting an error.
Acknowledgements

I would also like to thank Chris Neely for numerous emails patiently detailing the methodology used in his work with Paul Weller. In Cambridge, thanks to the directors of 4CMR Terry Barker and Doug Crawford-Brown for their generous and patient support, thanks to Robin Hankin for his rigorous approach to science. Thanks also to Mairéad, Annela, Svetlana, Paul, Mark, Jun Li, Yongfu and the rest of 4CMR for putting up with me. Thanks to Stephen Stretton, a work colleague and friend, for his intelligent insight. Thanks also to Richard Schuhmertl from Quant Capital for recognising my potential and providing me with a great opportunity. A special mention to Julie Kim, whose texts and company make me smile. Thanks to Jas for her eternal cheekiness. Thanks to Anouska and Freddie for being there. Naturally, eternal thanks go to my parents for their life-long support. Thanks to luck for my ability to select the unique model Talisa Qi, the only model that’s right. Finally, thank god I’ve finished.


Contents

Abstract 3

1 Introduction 12
  1.1 Description of the Area of Study .................................................. 12
  1.2 Problems Found in the Area .............................................................. 13
  1.3 Thesis Statement .............................................................................. 14
  1.4 Why the Thesis Question is Important .............................................. 17
  1.5 Research Undertaken in the Thesis .................................................. 18
  1.6 How the Thesis Question Relates to Future Work in the Area .......... 19
  1.7 What the Thesis Achieves .................................................................. 20
  1.8 Structure of the Thesis ..................................................................... 21

2 Background 23
  2.1 Characterization ................................................................................. 23
  2.2 Modelling .............................................................................................. 29
  2.3 Forecasting ............................................................................................. 40

3 Characterization 53
  3.1 Background .......................................................................................... 53
    3.1.1 Stylized Facts ................................................................................ 53
    3.1.2 Time Series .................................................................................... 56
    3.1.3 Efficient Market Hypothesis ......................................................... 56
    3.1.4 Data ................................................................................................. 60
  3.2 Autocorrelation ...................................................................................... 61
  3.3 Long Memory ........................................................................................ 61
  3.4 Runs Test ............................................................................................... 63
    3.4.1 First Runs Test ............................................................................. 63
    3.4.2 Second Runs Test ......................................................................... 64
  3.5 Investment Newsletters ....................................................................... 64
  3.6 Conclusion and Summary ..................................................................... 66
Contents

4 Modelling 67
  4.1 An Artificial Stock Market 67
    4.1.1 Design 67
    4.1.2 Implementation 68
    4.1.3 Testing 72
  4.2 Investment Performance Measurement 76
    4.2.1 Design 76
    4.2.2 Implementation 77
    4.2.3 Testing 80
  4.3 Conclusion and Summary 81

5 Forecasting 82
  5.1 Design 82
  5.2 Implementation 99
  5.3 Testing 102
  5.4 Conclusion and Summary 105

6 Assessment 107
  6.1 Hypothesis 107
  6.2 Precision 109
  6.3 Thoroughness 110
  6.4 Contributions 115
  6.5 Comparison with Similar Work of Others 117
  6.6 Conclusion and Summary 120

7 Conclusion and Future 122
  7.1 Conclusion 122
  7.2 Further Work 125

A ISO 4217 Currency Codes 132

B Exchanges and Stock Market Indices 133
  B.1 Exchanges 133
  B.2 Stock Market Indices 134

C Time Series Glossary 135

D Key Articles on the Efficient Market Hypothesis 137

E Runs Tests Source Code 138

F Runs Tests on DJIA Returns 146
Contents

G  Rescaled Range Analysis Source Code  151

H  Technical Analysis Taxonomy  154

I  Cumulative Prospect Theory Certainty Equivalent Source Code  156

J  Kernel Methods/Support Vector Machines  166
   J.1  Gram Matrix  .............................................. 166
   J.2  Hilbert Space  .............................................. 166

K  Fisher Kernel Source Code  167

L  Similar Publications  170

Bibliography  172
List of Figures

2.1 A hypothetical value function in prospect theory ........................................ 39
2.2 Weighting functions for gains and losses in cumulative prospect theory .......... 39

4.1 Mean log return (P&L) per analyst ............................................................... 74
4.2 Statistics of price log returns ................................................................. 74
4.3 Kurtosis of price log returns .................................................................. 75
4.4 Autocorrelations of price log returns ...................................................... 75
4.5 Maximal Sharpe ratio ............................................................................ 78

5.1 SVM_{dark} ......................................................................................... 88
5.2 winSVM ............................................................................................. 89

6.1 Net market exposure for various strategies in equities ............................... 114
List of Tables

1.1 Well-defined problem .......................................................... 16
3.1 Stochastic processes and their applicability to markets ................. 55
3.2 Autocorrelation of DJIA log returns ........................................ 61
3.3 Rescaled range analysis on detrended DJIA log returns .................. 62
3.4 Runs of DJIA log returns above and below the mean ..................... 63
3.5 ‘Up markets’ and ‘down markets’ as defined by the ‘Forbes/Hulbert investment letter survey’ ................................................ 65
3.6 Correlation analysis of ‘The Forbes/Hulbert investment letter survey’ ............................................................. 65
4.1 Statistics of daily stock log returns ............................................ 73
4.2 Statistics generated by the artificial stock market .......................... 73
4.3 Range of proportions of technical analysts in the artificial stock market that replicate stylized facts .............................................. 76
4.4 Sharpe ratio and CPTCE correlations ........................................ 80
5.1 Pseudocode for the fixed length HMM kernel ............................. 92
5.2 Fisher kernel test results ......................................................... 99
5.3 Fisher kernel symbol allocation ............................................... 101
5.4 Out of sample results, USD/DEM ............................................. 103
5.5 Out of sample results, USD/JPY ............................................... 103
5.6 Out of sample results, GBP/USD ............................................. 103
5.7 Out of sample results, USD/CHF ............................................. 104
5.8 Out of sample results, DEM/JPY ............................................. 104
5.9 Out of sample results, GBP/CHF ............................................. 104
5.10 Out of sample results, Mean .................................................. 105
5.11 Summary of results on forecasting ......................................... 106
D.1 Key articles on the EMH ....................................................... 137
F.1 DJIA daily returns: increasing ................................................. 146
F.2 DJIA daily returns: decreasing ................................................. 147
F.3 DJIA weekly returns: increasing ................................................. 147
List of Tables

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.4</td>
<td>DJIA weekly returns: decreasing</td>
<td>148</td>
</tr>
<tr>
<td>F.5</td>
<td>DJIA monthly returns: increasing</td>
<td>148</td>
</tr>
<tr>
<td>F.6</td>
<td>DJIA monthly returns: decreasing</td>
<td>149</td>
</tr>
<tr>
<td>F.7</td>
<td>DJIA annual returns: increasing</td>
<td>149</td>
</tr>
<tr>
<td>F.8</td>
<td>DJIA annual returns: decreasing</td>
<td>150</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This chapter sets the scene. It includes a description of the area of study, the problems found in the area, the particular problems addressed in the thesis (the ‘thesis statement’), why the thesis question is important, an outline of the research undertaken in the thesis, how the thesis question relates to future work in the area and what the thesis achieves with respect to the thesis statement and the area. The chapter finishes with an annotated guide to the rest of the thesis.

1.1 Description of the Area of Study

This thesis concerns the application of computer science to the financial domain. Specifically, the thesis concerns the application of artificial intelligence (AI) to financial markets.

After being pursued with excessive optimism in the 1960s and 70s, the meagre rate of progress meant that the term ‘AI’ became rather unfashionable and spawned various pseudonyms and sub-disciplines, often in an attempt to escape its reputation for failure (see Kassan (2006) and the references therein). I am not immune from history: this research is conducted from within the ‘Intelligent Systems Group’, and the thesis is titled accordingly. Indeed, as far as naming the area of research goes, one is spoilt for choice. Intelligent systems search for patterns and discover relationships in large data sets. They include neural networks, genetic algorithms and fuzzy systems. The work on forecasting in Chapter 5 is most commonly described as machine learning. Machine learning is the study of computer algorithms that improve automatically through experience. In practice, this involves creating programs that ‘learn’ through the analysis of data. Multidisciplinary research throws up the potential for terminology to confuse. In computer science, data mining is the practice of (automatically) searching (large) data sets for patterns. To do this, computational techniques from statistics, machine learning and pattern recognition are used. In contrast to the usual definition, the term ‘data mining’ in economics has negative connotations. It is not held to be a respectable mode of enquiry because it reverses the usual hypothetico-deductive model of scientific method and spurious relationships in the data may be found.

Whenever there are valuable commodities to be traded, there are incentives to develop a social arrangement that allows buyers and sellers to discover information and carry out a voluntary exchange more efficiently, i.e. develop a market. The largest and best organized markets in the world tend to be

\footnote{For introductory texts, see Langley (1996), Mitchell (1997), Bishop (2006), Marsland (2009) and Alpaydin (2010).}
the securities markets. Predicting financial markets is deceptively difficult. The good news is that, on
average, tossing a coin will predict the market direction correctly five times out of ten, and one only
needs to do marginally better to profit. However, the bad news is that a market price is generated by
a non-stationary, stochastic, discontinuous and probably non-linear dynamic process, and any useful
(i.e. profitable) signal is extremely noisy. The resulting time series approximates a martingale (i.e. the
expected return is zero), which makes prediction extremely difficult. Whilst it is usual for machine
learning to be applied to domains that are non-linear, dynamic and noisy, such as bioinformatics, financial
markets fit that description, but are even harder because they also have feedback and an asymptotically
low signal-to-noise ratio. An interesting thing about markets is their zero-sum nature. Traders make
and lose equal amounts, whilst the broker takes a cut, so there is actually a negative sum game embedded
in a zero sum game. This begs the question, why do people play? Speculation thrives on overconfidence.
Fortunately, financial markets are amenable to analysis. It was long ago realised that, in a market,
whenever buyers and sellers trade, it makes sense to record the agreed price at which the transaction
took place. This price record creates a time series.

The central paradigm in the model-driven domain of finance is the ‘efficient market hypothesis’
(EMH), and an efficient market is one in which prices always ‘fully reflect’ available information. A
random walk in a stock price is neither necessary nor sufficient for an efficient market, the important
concept is a martingale. The EMH, by itself, is not a well-defined and empirically refutable hypothesis;
one would need to test it with an auxiliary hypothesis which describes investors’ risk preferences. If
participants make money in a market, an economist might argue that this was not due to skill, but due to
changing risk preferences. Furthermore, because information is costly, a market price cannot perfectly
reflect the information which is available, since if it did, those who spent resources to obtain the infor-
mation would receive no compensation. Paradoxically, market efficiency relies on the fact that market
participants believe it to be false. For if investors believed that markets were efficient, they would not
speculate in the market, and it is speculation that drives the market towards efficiency as any anomalies
are traded out of the market. Contrary to popular belief, the EMH does not require that all market par-
ticipants are rational. Indeed, markets can be efficient even when a group of investors are irrational and
correlated, so long as there are some rational traders present together with arbitrage opportunities. In the-
ory, if not in practice, market efficiency can be tested. If information is revealed to market participants,
the reaction of security prices can be measured. If and only if prices do not move when the information
is revealed, the market is efficient with respect to that information set. There is little consensus between
the opinions held in academia and industry. Unsurprisingly, most of the support for the EMH comes
from the former. It is likely that markets have become increasingly efficient.

1.2 Problems Found in the Area

The task of forecasting financial markets is one of predicting a time series generated from a social sci-
ence, which in practice is purely an exercise in information processing. An attractive way of achieving

\footnote{The exception is the stock market, which is a positive-sum game for a rising market. Note also that government intervention in foreign exchange markets may provide a positive sum game for other participants in the short-term.}
this is to make minimal assumptions and use a data-driven, model-free, flexible and nonparametric approach. In other words, use machine learning, in the guise of supervised learning, which encompasses both theoretical soundness and experimental effectiveness. Multiagent systems would appear the most natural way of modelling a market when the market participants are both numerous and autonomous. However, it is possible that financial markets are not the best test bed for new algorithms in general, because the low signal-to-noise ratio of a financial time series makes it harder to distinguish between an algorithm that has identified a signal from one that has forecast noise due to luck. That is, computer science has a lot to offer finance, but the converse may be less true. Because computer science, at least when compared to the more traditional subjects such as mathematics, physics or economics, is a comparatively new area of research, and also in part due to Moore’s Law, computer science is a relatively fast-changing discipline (Hughes, 2002, p. 2). This can lead academics may race to publish or present their latest algorithm before others, which could lead to a lack of rigour relative to older disciplines, such as economics. My other main criticism is that a lot of the machine learning community tend to champion their favourite technique, which in practice results in a solution looking for a problem. Also, often simple solutions (such as linear regression) are overlooked. Other criticisms include the irrelevance of fuzzy logic (see Lindley (1987)) and the intellectual dishonesty of the evolutionary and supervised learning communities. The former often (implicitly) deny the no free lunch theorems, whilst the latter regularly fail to declare what events their probabilities are conditioned on (e.g. Vapnik (1982), Vapnik (1998) and Vapnik (1999) are guilty of this sin). There is also the temptation for researchers in academia to treat machine learning as an exercise in searching for data sets that show their algorithm in a favourable light. Despite these grievances, on balance, I recommend the machine learning paradigm above all others, but use with care and common sense.

Machine learning can be viewed as an attempt to automate ‘doing science’ [Hume (1739–1740)] pointed out that ‘even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience’. More recently, and with increasing rigour, Mitchell (1980), Schaffer (1994) and Wolpert (1996) showed that bias-free learning is futile, and that there can be no guarantee that any algorithm will generalize well. For this reason, it is the author’s belief that the key to developing successful machine learning algorithms is to carefully consider the assumptions being made, which requires extracting as much domain knowledge as possible. Putting theory into practice, the research on characterization in this thesis is motivated by the work on forecasting.

1.3 Thesis Statement

This section entails stating what is variously known as the ‘problem statement’, ‘thesis statement’, ‘thesis question’ or ‘research question’. The ‘thesis’, in the narrow sense of the word, is split into three hypotheses. A ‘well-defined problem’ is then introduced to give formalism and structure to both the central thesis and the three hypotheses.

[3]My Bayesian take on science was recently presented in Cambridge [Sewell (2012b)].
1.3. Thesis Statement

Problem Statement

The thesis research question is: ‘Can one improve upon the state of the art in financial time series analysis through the application of intelligent systems?’ The thesis is split according to the three goals of time series analysis: characterization, modelling and forecasting.

**characterization** Characterization attempts with little or no a priori knowledge to determine fundamental properties, such as the stationarity of a system or the amount of randomness.

**modelling** The goal of modelling is to find a description that accurately captures features of the long-term behaviour of the system.

**forecasting** The aim of forecasting (also called predicting) is to accurately predict the short-term evolution of the system.

In the forecasting third of the thesis the aims may be graduated thus: 1) to ‘beat the market’, 2) to improve standard algorithms, and 3) to beat the ‘state of the art’. Note that in a foreign exchange market ‘beating the market’ simply means earning a positive return, but it is not a useful benchmark as it does not incorporate risk. Whilst the most ambitious goal, 3), is ill-defined, as there is no consensus within academia and it is likely to be proprietary outside, but an algorithm that successfully forecast financial markets published in the academic literature shall be used as a proxy.

Well-Defined Problem

The concept of a well-defined problem (McCarthy 1956; Newell and Simon 1972; Nozick 1993) is employed in order to make the ‘thesis’ (in the narrow sense of the word) explicit:

**Goal** Improve upon the state of the art in financial time series analysis.

**Initial state** The current state of the art in financial time series analysis.

**Admissible operations** Intelligent systems.

**Constraints** Finite resources such as time, money, processing power, ability and support (‘bounded rationality’).

**Outcome** An improved ability to characterize, model and forecast financial markets.

A solution to the problem is a sequence of admissible operations (in this case, intelligent systems) that transforms the initial state (the current state of the art in financial time series analysis) into an outcome that meets the goal (improve upon the state of the art in financial time series analysis), without violating any constraints (finite resources such as time, money, ability and support) at any time along the way. The central thesis shall be decomposed into three related hypotheses, each of which will be tested. Each hypothesis is treated as a well-defined problem, as described in Table 1.1 (p. 16). The table enables one to compare the three strands of work. Note that it describes my personal approach to the problem, for example the initial states describe my own starting point, and the constraints are those I expect to find personally. Further, note that this exercise is wholly completed prior to any research taking place, so the table includes desired outcomes, and does not necessarily describe what was eventually achieved.
### 1.3. Thesis Statement

<table>
<thead>
<tr>
<th>Characterization</th>
<th>Modelling</th>
<th>Forecasting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td>The goal is to minimize the difference between empirical financial data and our statistical description of the data, the value of simplicity is also recognised.</td>
<td>To provide insight into the impact of technical analysts on a financial market.</td>
</tr>
<tr>
<td><strong>Initial state</strong></td>
<td>Current literature on empirical analysis.</td>
<td>Current best models described in the literature.</td>
</tr>
<tr>
<td><strong>Admissible operations</strong></td>
<td>An appraisal and summary of previous literature plus my own statistics.</td>
<td>Any behaviour which stems from evolved heuristics and biases, plus an implementation of a multiagent system.</td>
</tr>
<tr>
<td><strong>Constraints</strong></td>
<td>A finite amount of data (as with previous authors, whose literature is reviewed).</td>
<td>Must be a realistic abstraction of the real world with an appropriate constraint on complexity.</td>
</tr>
<tr>
<td><strong>Desired outcome</strong></td>
<td>A more accurate characterization of financial markets and more complete stylized facts.</td>
<td>A better understanding into the impact that technical analysts have on a financial market.</td>
</tr>
</tbody>
</table>

**Table 1.1: Well-defined problem**
1.4 Why the Thesis Question is Important

A greater understanding of financial markets is important, if only because of the vast amount of money involved. Due to the Wall Street Crash the Dow Jones Industrial Average (DJIA) lost 40 per cent of its value in September and October of 1929. More recently, and in the UK, the events that took place on Black Wednesday (16 September 1992) when sterling was forced out of the ERM are an example of when the markets were more powerful than the governments; the markets were right in the sense that the devaluation forced on sterling was justified by the country’s economic dilemma. A third example concerns the ‘dot-com bubble’, the DJIA tripled between 1994 and 1999, whilst, over the same period, basic economic indicators did not come close to tripling. Lastly, average daily foreign exchange market turnover rose to $4.0 trillion in April 2010 [Bank for International Settlements (2010)]. This is over 15 times greater than the gross domestic product (GDP) of the world economy and over 70 times greater than the dollar value of all international trade.

Financial markets are sensitive. When Alan Greenspan, chairman of the Federal Reserve Board in Washington, used the term ‘irrational exuberance’ to describe the behaviour of stock market investors in an otherwise staid speech on 5 December 1996, the FTSE 100 index was down 4 per cent at one point during the day.

Markets usually reflect the decisions of thousands or even millions of people going about their daily lives. The sheer size of the markets also makes the area research-worthy and non-trivial. The markets uniquely capture the psychology of individuals on a large scale.

The area of work this thesis covers serves a useful purpose. For large multinational firms that conduct substantial currency transfers in the course of business, being able to accurately forecast the movements of exchange rates can result in considerable improvement in the overall profitability of the firm. Even if trading is not profitable, speculators help keep markets liquid, better allowing hedgers to trade as and when they wish.

The work this thesis addresses is potentially extremely lucrative. Schwager (1989) contains some amazing stories, such as an electrical engineering graduate from MIT whose trading amassed an astounding 250,000 per cent return over a sixteen-year period. Taub (2006) reports that Renaissance Technologies made approximately a 34 per cent annualized net return since its 1988 inception. Both employed a scientific computer-based approach to trading. The philosophers’ stone was thought by alchemists to be the key ingredient in changing base metals into gold. Today, traders are seeking a modern-age philosophers’ stone: the ‘holy grail’ of trading is a successful trading system. However, most traders lose money, and some lose heavily: in 1998 Long-Term Capital Management lost $4.6 billion in less than four months.

The research domain is growing. As technology drives down transaction costs, markets are increasingly accessible to an increasing number of participants. Also, globally, the failure of communism has ensured that the market economy continues to grow.
1.5 Research Undertaken in the Thesis

The research on the characterization of financial markets comprises of three experiments which contribute to the ‘stylized facts’ of financial markets, and each has implications vis-à-vis market efficiency. My implementation of Hurst’s rescaled range (\(R/S\)) analysis (in C++) found little evidence of long memory in stock market returns. My implementation of \(R/S\) analysis is more accurate than commercially available software, but slower. A runs test (a non-parametric test of the mutual dependence of the elements of a sequence) performed on the DJIA showed that daily returns, in particular, exhibit significant dependence. I purchased ‘The Forbes/Hulbert investment letter survey’, the data encompasses performance from 31 May 1990 to 31 December 2001 and includes just those newsletters tracked that have a predominant US equity focus. The performance of the recommendations of the newsletters is analysed by means of correlation analysis on the quantitative data and the results evidenced weak-form market efficiency.

The work on modelling utilizes behavioural finance to 1) model market action and 2) model investors’ risk preferences. For the former, the evolved heuristics and biases exhibited by fundamental analysts and technical analysts, such as representativeness and conservatism, are used to build an agent-based artificial stock market (in Excel). The relative proportion of technical analysts and fundamental analysts was allowed to vary, leading to the following broad conclusions. Whether a fundamental analyst, or a technical analyst, it pays to be in a small majority of about 60 per cent, whilst being in a small minority is the least profitable position to be in. As the number of technical analysts increases, the standard deviation of returns decreases, whilst the skewness increases. Whilst kurtosis of market returns peaks with around 40 per cent technical analysts, and rapidly declines as the number of technical analysts exceeds 90 per cent. The autocorrelation of returns is close to zero with 100 per cent fundamental analysts, and rapidly declines as the proportion of technical analysts approaches 100 per cent. With a realistic proportion of technical analysts and fundamental analysts, the artificial stock market replicates mean returns, the standard deviation of returns, the absolute returns correlation and the squared returns correlation of a real stock market. However, the artificial stock market failed to accurately replicate the skewness, kurtosis and autocorrelation of returns. The number of free parameters was kept to a minimum, so there was little scope for tuning the model until it output the desired results. In the second part, I devised and implemented (in both PHP and Visual Basic) an investment performance measurement metric developed from prospect theory (Kahneman and Tversky [1979], Tversky and Kahneman [1992]) known as cumulative prospect theory certainty equivalent (CPTCE). The implementation of CPTCE makes up part of a more general performance measurement calculator which I wrote and is freely available online[^1]. It calculates mean return, standard deviation, skewness, kurtosis, beta, Jensen’s alpha, Sharpe ratio, Sortino ratio, Treynor’s measure, information ratio, Stutzer ratio, Omega, \(M^2\), \(T^2\) and maximum drawdown, and is in use by the financial industry.

The Fisher kernel, introduced by Jaakkola and Haussler [1999a] and named in honour of Sir Ronald Fisher, gives a ‘natural’ similarity measure that takes into account an underlying probability distribution.

[^1]: [http://www.performance-measurement.org](http://www.performance-measurement.org)
1.6. How the Thesis Question Relates to Future Work in the Area

It seems natural to compare two data points through the directions in which they ‘stretch’ the parameters of the model, that is by viewing the score function at the two points as a function of the parameters and comparing the two gradients. If the gradient vectors are similar it means that the two data items would adapt the model in the same way, that is from the point of view of the given parametric model at the current parameter setting they are similar in the sense that they would require similar adaptations to the parameters. In the research on forecasting financial markets, a left-to-right hidden Markov model (HMM) which uses the Baum-Welch (maximum likelihood) training algorithm is trained on foreign exchange data to derive a Fisher kernel (which I implement in C++) for a support vector machine (SVM), and the DC (difference of convex functions) algorithm (as implemented by Argyriou et al. (2006)) and a Bayes point machine (developed by Tom Minka) are also used to create kernels. Furthermore, and most novel of all, the DC algorithm is used to learn the parameters of the HMM in the Fisher kernel. The four methods, along with a vanilla SVM, are compared with the genetic programming approach used in Neely et al. (1997) and reported in Neely et al. (2009) (NWD/NWU). All four methods performed better than the vanilla SVM, but none better than NWD/NWU. I ported two implementations of SVMs to Windows and also added semi-automated parameter selection. SVM\textsuperscript{dark} is based on SVM\textsuperscript{light} (Joachims, 2004) and written in C for Win32, whilst winSVM\textsuperscript{6} is based on mySVM (Rüping, 2000) and written in C++ for Win32. My Windows SVM software has been used by the financial industry.

1.6 How the Thesis Question Relates to Future Work in the Area

As the quantity of historical data available and computing power both increase, the accuracy of the characterization of financial time series in the future can only increase. As psychologists identify more heuristics and biases, increasingly accurate models of human behaviour, and therefore markets, should be possible, via the field of behavioural finance. Work on forecasting time series is always ripe for future improvement as machine learning algorithms are constantly being enhanced and refined, whilst more data becomes available and computing power increases. However, it will always be an arms race, with only the best succeeding.

Algorithmic trading typically involves splitting up an order to buy or sell a fixed number of shares in an optimal manner over a period of time, this is extremely fertile territory for the application of machine learning. Intelligent techniques could be used to optimize a trading system based on cointegration. One could employ copulas to inspect the relationships between the inputs and the target to ensure that the inputs are combined in an optimal manner, then use stochastic programming for robust portfolio optimization. Ensemble learning could be used to attempt to combine individual predictive models in an optimal way. An equity trading system could be built using the knowledge gained from the characterization of equity markets. Evolutionary algorithms, such as genetic algorithms or genetic programming, have been fruitfully applied to financial time series prediction, and further work is always possible. Intelligent techniques could be employed to select funds and allocate capital. Quantitative techniques could be applied to global macro hedge fund strategies. Optimizing an automated market-making algorithm

\textsuperscript{5}Available from http://svmdark.martinsewell.com/
\textsuperscript{6}Available from http://winsvm.martinsewell.com/
using intelligent techniques could be hugely lucrative for the financial industry. In the area of mergers and acquisitions, machine learning could be applied to the potentially lucrative task of predicting takeover targets. Once a trader has found a positive expected return, they need to decide what proportion of their capital to bet per trade: money management and position sizing are a challenging optimization problem. Another potential source of enormous wealth would be the application of intelligent techniques to option pricing. Rather than just the price, or the bid and the ask, an intelligent trading system could utilize several levels of the order book. A probabilistic model such as a particle filter could be employed to track a financial time series. Further work could be done on investment portfolio performance measurement, such as incorporating tailored risk preferences. Stochastic programming could be used for robust portfolio optimization. With trading systems, one could seek to optimize profits directly, rather than using mean squared error. Finally, intelligent techniques could be employed to predict yield curves. Section 7.2 (p. 125) addresses potential ideas for further work and goes into more detail.

My interests in developing the work further mainly concern working with ultra high frequency financial data. Tick data, preferably showing the order book, must be ripe for exploitation by machine learning, but would require impressive processing power, a vast amount of storage and robust algorithms. Today, tick data generated by financial markets quite possibly represents a greater volume than any other source outside high energy physics. Futures markets provide the most data, followed by foreign exchange (FX) markets, followed by stock markets, although most of the literature relates to stock markets. It is likely that the type of modelling required would be similar across all three types of market. In practice, we are more likely to be concerned with futures markets or FX markets because transaction costs are vanishingly small and leverage is possible.

1.7 What the Thesis Achieves

The results of the work on the characterization, modelling and forecasting of financial time series each lend support to their respective hypotheses and therefore to the central thesis. The characterization used existing literature plus statistics, the modelling used behavioural biases and multiagent systems, and the forecasting used supervised learning. The central argument of the thesis is that one can improve upon the state of the art in financial time series analysis through the application of intelligent systems.

The major contributions made are listed below.

- Using tests for autocorrelation and the runs test I reconcile the fact that daily DJIA log returns pass linear statistical tests of efficiency, yet nonlinear forecasting methods can still make above-average risk-adjusted returns. See Chapter 3.

- The evolved heuristics and biases exhibited by fundamental analysts and technical analysts, inducing underreaction and overreaction, are used to build an agent-based artificial stock market. The resultant time series provides a novel insight into the effect of the proportion of technical analysts relative to fundamental analysts. See Chapter 4.

- A novel investment performance measurement metric, cumulative prospect theory certainty equivalent (CPTCE), is developed from Tversky and Kahneman’s cumulative prospect theory. The
statistic models investors’ empirically-observed risk preferences (people care about losses and gains rather than absolute wealth, evaluate probabilities incorrectly, are loss averse, risk averse for gains, risk seeking for losses and have non-linear preferences), whilst no other performance metric does this effectively. The financial industry have taken interest, with offers to commercialize the product. See Chapter 4.

- Two Windows implementations of SVMs with semi-automated parameter selection are built. SVM_dawn is based on SVM_light and written in C for Win32, whilst winSVM is based on mySVM and written in C++ for Win32. For some time the software was the only Windows application dedicated to support vector machines, is frequently downloaded and has been used by the financial industry. The source code is also freely available to download. See p. 87.

- A (generative) hidden Markov model is trained on market data to derive a Fisher kernel for a (discriminative) support vector machine, the DC algorithm and a Bayes point machine are also used to create kernels. Furthermore, the DC algorithm is used to learn the parameters of the hidden Markov model in the Fisher kernel. The four algorithms are compared with a vanilla SVM and the genetic programming approach used in Neely et al. (1997) and reported in Neely et al. (2009). See Chapter 5.

1.8 Structure of the Thesis

Abstract

1 Introduction The first chapter ‘sets the scene’. It includes a description of the area of study, the problems found in the area, the particular problems addressed in the thesis (the ‘thesis statement’), why the thesis question is important, an outline of the research undertaken in the thesis, how the thesis question relates to future work in the area and what the thesis achieves with respect to the thesis statement and the area. Finally, an annotated guide to the rest of the thesis is provided here.

2 Background A survey and critical assessment of other work and its relation to the research in this thesis. The literature in the following areas is reviewed: the EMH, dependence and long memory in market returns, investment newsletters, technical analysis, behavioural finance, multiagent systems, investment performance measurement, kernel methods and support vector machines with a particular focus on the application of SVMs to the financial domain.

3 Characterization The first of the three core chapters comprises of three experiments. Runs tests performed on the DJIA showed that daily returns, in particular, exhibit significant dependence. An implementations of Hurst’s rescaled range (R/S) analysis found little evidence of long memory in DJIA returns. The performance of investment newsletters is analysed, evidencing weak-form market efficiency.

4 Modelling The work on modelling utilizes behavioural finance. The evolved heuristics and biases exhibited by fundamental analysts and technical analysts, inducing underreaction and overreaction, are used to build an agent-based artificial stock market. The time series generated by the
artificial market provides insight into the effect of technical analysts. A novel investment performance measurement metric, CPTCE, is developed from prospect theory \cite{Kahneman1979, Tversky1992}.

5 **Forecasting** A hidden Markov model is trained on foreign exchange data to derive a Fisher kernel for an SVM, and the DC algorithm and Bayes point machine are also used to create kernels. Further, the DC algorithm was used to learn the parameters of the hidden Markov model in the Fisher kernel. An implementation of SVMs with semi-automated parameter selection is built.

6 **Critical assessment of own work** The hypothesis is stated; precision, thoroughness and the contributions are demonstrated, and a comparison with the closest rivals is given. The results of the work on the characterization, modelling and prediction of financial time series each lend support to the hypotheses and therefore to the central thesis.

7 **Conclusion and Future** The conclusion summarizes the thesis and highlights the contributions made. Finally, potential ideas for further work in the field are addressed, including the application of intelligent techniques to algorithmic trading, cointegration, copulas, ensemble learning, an equity trading system, evolutionary algorithms, funds of funds, global macro strategies, market-making, merger arbitrage, money management, option pricing, the order book, a particle filter, investment performance measurement, portfolio optimization, a profit-objective error function and yield curve analysis.

A **ISO 4217 Currency Codes**

B **Exchanges and Stock Market Indices**

C **Time Series Glossary**

D **Key Articles on the Efficient Market Hypothesis**

E **Runs Test on DJIA Returns**

F **Rescaled Range Analysis Source Code**

G **Technical Analysis Taxonomy**

H **Cumulative Prospect Theory Certainty Equivalent Source Code**

I **Kernel Methods/Support Vector Machines**

J **Fisher Kernel Source Code**

K **Similar Publications**

Bibliography
Chapter 2

Background

This chapter is a survey and critical assessment of related work. The following topics are covered: the efficient market hypothesis, dependence in market returns, long memory in market returns, investment newsletters, technical analysis, behavioural finance, multiagent systems, investment performance measurement, kernel methods and support vector machines (SVMs) with a focus on financial applications. For further background reading, see the Feature Article in the August 2004 edition of the IEEE Computational Intelligence Society Newsletter where Tsang and Martinez-Jaramillo (2004) define the scope and agenda of research in computational finance. Also, in February 2009, IEEE Transactions on Evolutionary Computation published a Special Section on ‘Computational finance and economics’, and the editorial (Tsang and Isasi (2009) reviews the current state of the field.

2.1 Characterization

This section is a review of the literature relevant to the experiments conducted in Chapter 3 on the characterization of financial time series.

Efficient Market Hypothesis

The efficient market hypothesis (EMH) is of central importance to this thesis because it has profound implications regarding financial time series analysis. Together with assumptions of investors’ risk preferences, if a market is efficient (and real markets very nearly are), this puts constraints on what is possible in the characterization, modelling and forecasting of financial markets. The EMH is covered in detail in Section 3.1.3 (pp. 56–60), a brief literature review is presented here. The earliest known historical cases of speculation were in ancient Rome during the Republic of the second century BC (Chancellor (2000). The random walk hypothesis was conceived in the 16th century as a model of games of chance. Bachelier (1900) modelled the path of stock prices as Brownian motion and showed that speculators should be unable to beat the market five years before Brownian motion was rediscovered by Einstein. Samuelson (1965) proved that properly anticipated prices fluctuate randomly, whilst Fama (1970) defined an efficient market as one in which prices always ‘fully reflect’ available information. However, Grossman and Stiglitz (1980) argued that because information is costly, a market price cannot perfectly reflect the information which is available, since if it did, those who spent resources to obtain the information would receive no compensation. A more detailed history of the EMH is given in Sewell (2011d).
Dependence in Market Returns

The third experiment on the characterization of financial markets concerns the application of the runs test (a non-parametric statistical test of the mutual dependence of the elements of a sequence) on a major US stock market index, so the relevant literature on the dependence of market returns is addressed here.

Fama (1970) found that 22 out of the 30 stocks of the DJIA exhibited positive daily serial correlation. Fama and French (1988) found that autocorrelations of stock return indices (they used portfolios) form a U-shaped pattern across increasing return horizons. The autocorrelations become negative for 2-year returns, reach minimum values for 3–5-year returns, and then move back toward 0.0 for longer return horizons. Lo and MacKinlay (1988) found significant positive serial correlation for weekly and monthly holding-period index returns, but negative autocorrelations for individual securities with weekly data. Ball and Kothari (1989) found negative serial correlation in five-year stock returns. Lo and MacKinlay (1990a) found negative autocorrelation in the weekly returns of individual stocks, whilst weekly portfolio returns were strongly positively autocorrelated. Jegadeesh (1990) found highly significant negative serial correlation in monthly individual stock returns and strong positive serial correlation at twelve months. Brock et al. (1992) found positive autocorrelation in DJIA daily returns. Boudoukh et al. (1994) found that for small-firm indices, the spot index’s autocorrelation is significantly higher than that of the futures. Zhou (1996) found that high-frequency FX returns exhibit extremely high negative first-order autocorrelation. Longin (1996) found positive autocorrelation for a daily index of stocks. The autocorrelation of weekly stock returns is weakly negative, whilst the autocorrelations of daily, weekly and monthly stock index returns are positive. Campbell et al. (1996). Lo and MacKinlay (1999) found a positive autocorrelation for weekly holding-period market indices returns, but a random walk for monthly. They also found negative serial correlation for individual stocks with weekly data. Cont (2001) found negative autocorrelation on a tick-by-tick basis for both foreign exchange (USD/JPY) and a stock (KLM shares traded on the New York Stock Exchange (NYSE)). He also claims that weekly and monthly autocorrelations exist. The autocorrelation of 1 minute FX returns is negative. Dacorogna et al. (2001). Ahn et al. (2002) looked at daily autocorrelations and found that the indices are positive even though the futures are close to zero. Lewellen (2002) found negative autocorrelation for stock portfolios after a year. Llorente et al. (2002) found that the first-order autocorrelation of daily returns is negative for stocks with large bid–ask spreads (-0.088) and small sizes (-0.076). It is positive but very small for large stocks (0.003) and stocks with small bid–ask spreads (0.01). Bianco and Reno (2006) found negative serial correlation of returns on the Italian stock index futures in periods smaller than 20 minutes. Cerrato and Sarantis (2006) looked at monthly data on black market exchange rates and found evidence of non-linear mean reversion in the real exchange rates of developing and emerging market economies. Lim et al. (2008) examined ten Asian emerging stock markets and discovered that all the returns series exhibit non-linear serial dependence. Serletis and Rosenberg (2009) used daily data on four US stock market indices and concluded that US stock market returns display mean reversion.

In summary, weekly and monthly stock returns are weakly negatively correlated, whilst daily, weekly and monthly index returns are positively correlated.
that this somewhat paradoxical result can mean only one thing: large positive cross-autocorrelations across individual securities across time. High frequency market returns exhibit negative autocorrelation.

Long Memory in Market Returns

In 1906, Harold Edwin Hurst, a young English civil servant, came to Cairo, Egypt, which was then under British rule. As a hydrological consultant, Hurst’s problem was to predict how much the Nile flooded from year to year. He developed a test for long-range dependence and found significant long-term correlations among fluctuations in the Nile’s outflows and described these correlations in terms of power laws. This statistic is known as the ‘rescaled range’, ‘range over standard deviation’ or ‘R/S’ statistic. From 1951 to 1956, Hurst, then in his seventies, published a series of papers describing his findings [Hurst 1951]. I run the world’s only website dedicated to long-range dependence [1] The definition of long memory below is taken from Beran (1994) (p. 42).

**Definition 1** If \( \rho(k) \) is the correlation at lag \( k \), let \( X_t \) be a stationary process for which the following holds. There exists a real number \( \alpha \in (0, 1) \) and a constant \( c_p > 0 \) such that

\[
\lim_{k \to \infty} \frac{\rho(k)}{c_p k^{-\alpha}} = 1. \tag{2.1.1}
\]

Then \( X_t \) is called a stationary process with long memory or long-range dependence or strong dependence, or a stationary process with slowly decaying or long-range correlations.

The parameter \( H = 1 - \frac{\alpha}{2} \) will also be used instead of \( \alpha \). In terms of this parameter, long memory occurs for \( \frac{1}{2} < H < 1 \). Knowing the covariances (or correlations and variance) is equivalent to knowing the spectral density \( f \). Therefore, long-range dependence can also be defined by imposing a condition on the spectral density.

**Definition 2** If \( f(\lambda) \) is the spectral density, let \( X_t \) be a stationary process for which the following holds: there exists a real number \( \beta \in (0, 1) \) and a constant \( c_f > 0 \) such that

\[
\lim_{\lambda \to 0} \frac{f(\lambda)}{c_f |\lambda|^\beta} = 1. \tag{2.1.2}
\]

Then \( X_t \) is called a stationary process with long memory or long-range dependence or strong dependence.

The second experiment on the characterization of financial time series concerns testing for long memory in the returns of financial markets, so the relevant literature on the dependence of market returns is addressed here.

Mandelbrot (1972) applied R/S analysis to financial returns. Greene and Fielitz (1977) claimed that many daily stock return series are characterized by long-term dependence. Aydogan and Booth (1988) concluded that there was no significant evidence for long-term memory in common stock returns.

2.1. Characterization

Lo (1991) modified the $R/S$ statistic to ensure that it is robust to short-range dependence and found little evidence of long-term memory in historical US stock market returns. Cheung (1993) found evidence of long memory in foreign exchange rates. Goetzmann (1993) considered three centuries of stock market prices. $R/S$ tests provided some evidence that the detrended London Stock Exchange and NYSE prices may exhibit long-term memory. Cheung and Lai (1993) examined the long memory behaviour in gold returns during the post-Bretton Woods period and found that the long memory behaviour in gold returns is rather unstable. They concluded, ‘[w]hen only few observations corresponding to major political events in the Middle East, together with the Hunts event, in late 1979 are omitted, little evidence of long memory can be found.’ Mills (1993) found little evidence of long memory in daily UK stock returns. Embrechts (1994) claims that the Hurst coefficient for JPY/USD returns indicates a memory effect. Embrechts et al. (1994) applied rescaled range analysis to US Fed Fund rates, US Treasury notes, CHF/USD exchange rates and the Japanese stock market (TOPIX) and claimed that it shows that most financial markets follow a biased random walk. Bhar (1994) tested for long-term memory in the JPY/USD exchange rate using Lo’s methodology and found no evidence of long-term memory. Moody and Wu (1995) performed rescaled range and Hurst exponent analysis on tick-by-tick interbank foreign exchange rates, and found that they are mean-reverting. Nawrocki (1995) considered the CRSP monthly value-weighted index and the S&P 500 daily index, and found that the Hurst exponent and the Lo-modified $R/S$ statistic indicate that there is persistent finite memory. Tschernig (1995) found evidence for weak long memory in the changes of DEM/USD spot rates and the CHF/USD spot rates; in contrast, there was no evidence for long memory in the DEM/CHF spot rate changes. Chow et al. (1996) found evidence that consistently revealed the absence of long-term dependence in 22 international equity market indices. Moody and Wu (1996) improve Lo’s $R/S$ statistic and conclude that the DEM/USD series is mildly trending on time scales of 10 to 100 ticks. Peters (1996) applied $R/S$ analysis and concluded that most of the capital markets are characterized by long memory processes. Lux (1996) analysed German stock market data and found no evidence for (positive or negative) long-term dependence in the returns series. Barkoulas and Baum (1996) applied the spectral regression method and found no evidence of long memory in either aggregate or sectoral stock indices, but evidence of long memory in 5, intermediate memory in 3 and no fractal structure in 22 of the 30 DJIA companies. Overall findings did not offer convincing evidence against the martingale model. Using the spectral regression method, Barkoulas and Baum (1997) found significant evidence of long memory in the 3- and 6-month returns (yield changes) on Euromoney deposits denominated by JPY (Euroyen). Hiemstra and Jones (1997) applied the modified rescaled range test to the return series of 1,952 common stocks and their results indicated that long memory is not a widespread characteristic of those stocks. Lobato and Savin (1998) found no evidence of long memory in daily stock returns. Willinger et al. (1999) found empirical evidence of long-range dependence in stock price returns, but the evidence was not absolutely conclusive. Huang and Yang (1999) applied the modified $R/S$ technique to intraday data and found the phenomenon of long-term memory in both NYSE and NASDAQ indices. Baum et al. (1999) reject the hypothesis of long memory in real exchange rates in the post-Bretton Woods era.
2.1. Characterization

Using the spectral regression method, Barkoulas et al. (2000) found significant and robust evidence of positive long-term persistence in the Greek stock market. Chen (2000) calculated the classical rescaled range statistic of Hurst for seven Asia-Pacific countries’ stock indices and concluded that all the index returns have long memory. Crato and Ray (2000) found no evidence for long memory in futures’ returns. Weron and Przybyłowicz (2000) found that electricity price returns are strongly mean-reverting. Zhuang et al. (2000) investigated British stock returns and found little or no evidence of long-range dependence. Sadique and Silvapulle (2001) examined the presence of long memory in weekly stock returns of seven countries, namely Japan, Korea, New Zealand, Malaysia, Singapore, the US and Australia. They found evidence for long-term dependence in four countries: Korea, Malaysia, Singapore and New Zealand. Cheung and Lai (2001) found long memory in JPY-based real exchange rates. Nath (2001) found indications of long-term memory in the Indian stock market using R/S analysis, but suggested that a more rigid analysis, such as Lo’s modified R/S statistic, should be used. Panas (2001) found long memory in the Athens Stock Exchange. Cavalcante and Assaf (2002) found little evidence of long memory in the returns of the Brazilian stock market. Nath and Reddy (2002) used R/S analysis and found long-term memory in the USD/INR exchange rate, although the variance ratio test clearly implied that there exists only short-term memory. Henry (2002) investigated long range dependence in nine international stock index returns. They found evidence of long memory in four of them, the German, Japanese, South Korean and Taiwanese markets, but not for the markets of the UK, US, Hong Kong, Singapore and Australia. Tolvi (2003a) found long memory in Finnish stock market return data. Using a monthly data set consisting of stock market indices of 16 OECD countries, Tolvi (2003b) found statistically significant long memory for three countries: Denmark, Finland and Ireland, which are all small markets. In a paper that examines and compares the behaviour of four tests for fractional integration in daily observations of silver prices, de Peretti (2003) concluded that one must use at least a bilateral bootstrap test to detect long-range dependence in time series, and deduced that silver prices do not exhibit long memory. Beine and Laurent (2003) investigated the major exchange rates and found no evidence of long memory in the conditional mean. Limam (2003) analysed stock index returns in 14 markets and concluded that long memory tends to be associated with thin markets. Sapio (2004) used spectral analysis and found long memory in day-ahead electricity prices. Cajueiro and Tabak (2004) found that the markets of Hong Kong, Singapore and China exhibit long-range dependence. Naively, Cajueiro and Tabak (2005) state that the presence of long-range dependence in asset returns seems to be a stylized fact. They studied the individual stocks in the Brazilian stock market and found evidence that firm-specific variables can explain, at least partially, the long-range dependence phenomena. Grau-Carles (2005) applied four tests for long memory to two major daily stock indices, the S&P 500 and the DJIA, two samples from each. There was no evidence of long memory in the returns. Oh et al. (2006) studied the long-term memory in various stock market indices and foreign exchange rates using detrended fluctuation analysis on daily and high-frequency (5-minute and 1-minute) data. For all daily and high-frequency market data studied, no significant long-term memory was detected in the return series. Elder and Serletis (2007) found no evidence of long memory in the DJIA. Serletis and Rosenberg (2009) used daily data on four US stock
market indices and concluded that US stock market returns display anti-persistence. \cite{Tan2010} found evidence of long memory in the Malaysian stock market before the 1997 financial crisis, but not afterwards. \cite{Kang2010} tested the daily closing prices of the KOSPI 50 index and its 50 constituent stock prices for long memory. Their broad conclusion was that there is no long memory in the return series of the Korean stock market. \cite{Rege2011} calculated the Hurst exponent for the Portuguese stock market and concluded that it exhibits both long-memory and short-memory depending on the scale of the time period used. \cite{Mishra2011} used R/S analysis on daily returns from the Indian stock market to reveal strong evidence of persistence or temporal dependencies. \cite{Mukherjee2011} found no evidence for long-memory in the Indian stock market. \cite{Anoruo2011} examined the daily closing prices of CASE 30 (Egypt), MASI (Morocco), TUNINDEX (Tunisia) and NSE All Share (Nigeria), and monthly data from SEM (Mauritius), NSE 20 (Kenya), JSE All Share (South Africa), ZSE Industrials (Zimbabwe), BSE (Botswana) and JSE All Share (Namibia), and found evidence of long memory in the returns in the cases of Egypt and Nigeria, and, to a lesser extent, for Tunisia, Morocco and Kenya. \cite{Boubaker2012} found strong evidence of long memory in North African stock market returns. \cite{Fouladi2012} examined twenty-two foreign exchange currencies vis-à-vis the Philippine peso and concluded that there was no convincing evidence of long-term memory in any of them.

To summarize, about 30 per cent of the articles analysing stock market returns concluded that they exhibit long memory, and about 50 per cent of those that analysed foreign exchange returns concluded that they exhibit long memory.

**Investment Newsletters**

The third and final experiment on the characterization of financial time series concerns an analysis of investment newsletters, and a literature review is given here.

\cite{Graham1996} analysed the advice contained in a sample of 237 investment newsletter strategies over 1980–1992 and found that there is little information in the investment newsletters’ opinions regarding stock market direction. However they did find that the degree of disagreement among letters predicts both realized and expected volatility as well as trading volume. \cite{Graham1997} examined the performance of 326 newsletter asset-allocation strategies for the 1983–95 period. They found that, as a group, newsletters do not appear to possess any special information about the future direction of the market. Nevertheless, they found that investment newsletters that are on a hot streak (have correctly anticipated the direction of the market in previous recommendations) may provide valuable information about future returns. The Value Line Investment Survey is the best known investment newsletter, it is well-respected and freely available. \cite{Graham1999} found that a newsletter analyst is likely to herd on Value Line’s recommendation if his reputation is high, if his ability is low or if the correlation across analysts’ signals is high. \cite{Jaffe1999} analysed the recommendations of common stocks made by the investment newsletters followed by the Hulbert Financial Digest. Taken as a whole, the securities that newsletters recommend did not outperform appropriate benchmarks and the performance of the newsletters did not exhibit persistence. They found little, if any, evidence
of herding. Newsletters tend to recommend securities that have performed well in the recent past and newsletters with poor past performance are more likely to go out of business. Metrick (1999) analysed the equity-portfolio recommendations made by 153 investment newsletters. Overall, there was no significant evidence of superior stock-picking ability and no evidence of abnormal short-run performance persistence (‘hot hands’).

Kumar and Pons (2002) analysed the behaviour and performance of 353 investment newsletters that made asset allocation recommendations during a period covering more than 21 years (June 1980–November 2001). On aggregate the newsletters failed to outperform a passive investment strategy, but active newsletters and contrarian newsletters exhibited market-timing ability. When they examined the recommendations of individual newsletters at a higher frequency (daily as opposed to monthly), they found considerable evidence of timing-ability. There was also evidence of persistence in newsletters’ performance and a trading strategy that followed the average recommendations of newsletters that have performed well in the past 10 months is capable of outperforming the market on a risk-adjusted basis (the annual over-performance is 2.56 per cent).

### 2.2 Modelling

This section is a review of the literature relevant to the experiments conducted in Chapter 4 on the modelling of financial time series.

#### Fundamental Analysis

The primary research in Chapter 4 concerns building an agent-based artificial stock market. One of the two main classes of ‘actors’ employed in the model consists of fundamental analysts. Fundamental analysis is a method of forecasting markets through the analysis of relevant news.

#### Technical Analysis

The second class of ‘actors’ employed in the model are technical analysts.

#### A Note on Terminology

The noun chartist and related verb charting are also used, sometimes referring to a subset of technical analysis. Also, technical analysts are often referred to as ‘noise traders’ in the academic literature (‘noise’ being anything other than news).

#### Definition

Let us define technical analysis. Formally, if $P$ is price, $D$ is data generated by the process of trading, $t$ is time, $E$ is expectation and $|$ the Bayesian probability conditioning bar, then technical analysis is the art of inferring $E(P_{t|t>0}|D_{t|t<0})$. In other words, the forecasting of future market prices by means of analysis of historical data generated by the process of trading.

#### Practitioners’ Definitions

- ‘Technical analysis is the process of analyzing a security’s historical prices in an effort to determine probable future prices.’ Achelis (2000), p. 4
2.2. Modelling

• ‘It refers to the study of the action of the market itself as opposed to the study of the goods in which the market deals. Technical Analysis is the science of recording, usually in graphic form, the actual history of trading (price changes, volume of transactions, etc.) in a certain stock or in “the Averages” and then deducing from that pictured history the probable future trend.’ Edwards et al. (2007), p. 4

• ‘Technical analysis is the study of market action, primarily through the use of charts, for the purpose of forecasting future price trends. The term “market action” includes the three principal sources of information available to the technician—price, volume, and open interest.’ Murphy (1999), pp. 1–2

• ‘The art of technical analysis, for it is an art, is to identify a trend reversal at a relatively early stage and ride on that trend until the weight of the evidence shows or proves that the trend has reversed. […] Therefore, technical analysis is based on the assumption that people will continue to make the same mistakes they have made in the past.’ Pring (2002), p. 3

The first three definitions are, in spirit, consistent with my own, but Pring’s definition is narrower and relies on the existence of trends and reversals.

Literature Review

Brown and Jennings (1989) showed that technical analysis has value in a model in which prices are not fully revealing and traders have rational conjectures about the relation between prices and signals. Frankel and Froot (1990) provided evidence for the increasing use of technical analysis in the foreign exchange markets between 1978 and 1988. Neftci (1991) showed that a few of the rules used in technical analysis generate well-defined techniques of forecasting, but even well-defined rules were shown to be useless in prediction if the economic time series is Gaussian. However, if the processes under consideration are non-linear, then the rules might capture some information. Tests showed that this may indeed be the case for the moving average rule. Taylor and Allen (1992) report the results of a survey among chief foreign exchange dealers based in London in November 1988 and found that at least 90 per cent of respondents placed some weight on technical analysis, and that there was a skew towards using technical, rather than fundamental, analysis at shorter time horizons. In a comprehensive and influential study Brock et al. (1992) analysed 26 technical trading rules using 90 years of daily stock prices from the DJIA up to 1987 and found that they all outperformed the market. Blume et al. (1994) showed that volume provides information on information quality that cannot be deduced from the price. They also show that traders who use information contained in market statistics do better than traders who do not. Neely (1997) explains and reviews technical analysis in the foreign exchange market. Neely et al. (1997) used genetic programming to find technical trading rules in foreign exchange markets. The rules generated economically significant out-of-sample excess returns for each of six exchange rates, over the period 1981–1995. Lui and Mole (1998) reported the results of a questionnaire survey conducted in February 1995 on the use by foreign exchange dealers in Hong Kong of fundamental and technical analyses. They found that over 85 per cent of respondents rely on both methods and, again, technical analysis
was more popular at shorter time horizons. Neely (1998) reconciled the fact that using technical trading rules to trade against US intervention in foreign exchange markets can be profitable, yet, long-term, the intervention tends to be profitable. LeBaron (1999) showed that, when using technical analysis in the foreign exchange market, after removing periods in which the Federal Reserve is active, exchange rate predictability is dramatically reduced.

Lo et al. (2000) examined the effectiveness of technical analysis on US stocks from 1962 to 1996 and finds that over the 31-year sample period, several technical indicators do provide incremental information and may have some practical value. Fernández-Rodríguez et al. (2000) applied an artificial neural network (ANN) to the Madrid Stock Market and find that, in the absence of trading costs, the technical trading rule is always superior to a buy-and-hold strategy for both ‘bear’ market and ‘stable’ market episodes, but not in a ‘bull’ market. One criticism I have is that beating the market in the absence of costs seems of little significance unless one is interested in finding a signal which will later be incorporated into a full system. Secondly, it is perhaps naive to work on the premise that ‘bull’ and ‘bear’ markets exist, statistically. Lee and Swaminathan (2000) demonstrated the importance of past trading volume. Neely and Weller (2001) used genetic programming to show that technical trading rules can be profitable during US foreign exchange intervention. Cesari and Cremonini (2003) made an extensive simulation comparison of popular dynamic strategies of asset allocation and found that technical analysis only performs well in Pacific markets. Cheol-Ho Park and Scott H. Irwin wrote ‘The profitability of technical analysis: A review’ (Park and Irwin 2004), an excellent review paper on technical analysis. Kavajecz and Odders-White (2004) showed that support and resistance levels coincide with peaks in depth on the limit order book and moving average forecasts reveal information about the relative position of depth on the book. They also show that these relationships stem from technical rules locating depth already in place on the limit order book. In their book, The Evolution of Technical Analysis, Lo and Hasan Hodzic (2010) provide a comprehensive history of the evolution of technical analysis from ancient times to the Internet age.

I host and run a website dedicated to technical analysis which, unusually, has an academic flavour.

**Taxonomy**

A taxonomy of the various methods of technical analysis applied by practitioners is given in Appendix H (pp. 154–155). Of the 26 techniques listed, according to the academic literature there is evidence for the efficacy of about 11, but no evidence for the efficacy of the remaining 15.

**Assumptions**

Technical analysts rely on the assumption that markets discount everything except information generated by market action, ergo, all you need is data generated by market action.

---

2 A limit order is an order to a broker to buy(sell) a security at or below(above) a specific price; whilst a limit order book is a record of unexecuted limit orders maintained by the specialist.

[http://www.technicalanalysis.org.uk](http://www.technicalanalysis.org.uk)
2.2. Modelling

Why is Technical Analysis so Popular?

If the weak form of the efficient market hypothesis holds, then technical analysis has no value. Conversely, for technical analysis to work requires that the weak (and therefore the semi-strong and strong) forms of the EMH are false. Also, if a market price follows a Markov process then technical analysis holds no value. Why, then, is technical analysis so popular? People often predict future uncertain events by taking a short history of data and asking what broader picture this history is representative of (independent of other information about its actual likelihood). This is a heuristic known as representativeness (Tversky and Kahneman, 1974). Technical analysis is representativeness. Below are some more psychological explanations of why a large number of people have a strong belief in technical analysis.

Communal reinforcement Communal reinforcement is a social construction in which a strong belief is formed when a claim is repeatedly asserted by members of a community, rather than due to the existence of empirical evidence for the validity of the claim.

Selective thinking Selective thinking is the process by which one focuses on favourable evidence in order to justify a belief, ignoring unfavourable evidence.

Confirmation bias Confirmation bias is a cognitive bias whereby one tends to notice and look for information that confirms one’s existing beliefs, whilst ignoring anything that contradicts those beliefs. It is a type of selective thinking.

Self-deception Self-deception is the process of misleading ourselves to accept as true or valid what we believe to be false or invalid by ignoring evidence of the contrary position.

Is Technical Analysis Self-Fulfilling?

Is technical analysis self-fulfilling or self-destructive? A priori, I hypothesize that if one conditions on price, then technical analysis is self-fulfilling; and if one conditions on time, then technical analysis is self-destructive. The evidence for the former includes the success of support and resistance (Brock et al., 1992; Osler, 2000), and the evidence for the latter includes the documented erosion of calendar effects (Schwert, 2003). Surprisingly, this is a contribution.

Conclusions

Publication bias (discussed on p. 109) should not adversely affect the relative performance of technical analysis, such as comparing different techniques, or their efficacy in different markets. The excellent review paper by Park and Irwin (2004) does precisely that. My literature review (pp. 29-31) together with Park and Irwin’s results give rise to the following conclusions:

- There is evidence in support of the usefulness of moving averages, momentum, support and resistance and some patterns; but no convincing evidence in support of Gann Theory or Elliott Wave Theory.

4For details of traditional technical analysis, see [Murphy, 1999], [Achelis, 2000], [Pring, 2002] and [Edwards et al., 2007]. There is money to be made selling books on technical analysis. Caveat emptor! The eminently more sensible [Aronson, 2006] offers a glimmer of hope.
• Technical analysis works best on currency markets, intermediate on futures markets, and worst on stock markets. An explanation is given on p. 113.

• Chart patterns work better on stock markets than currency markets.

• Non-linear methods work best overall. This is not at all surprising in light of the non-linearities found in markets [Hsieh 1989, Scheinkman and LeBaron 1989, Brock et al. 1991].

• Technical analysis doesn’t work as well as it used to. As transaction costs decrease, available computing power increases and the number of market participants increases, one would expect markets to become increasingly efficient and thus it is not surprising that the efficacy of technical analysis should diminish.

### Behavioural Finance

The algorithms employed by the artificial stock market are based on the behaviour of real market participants, rather than the actions of the rational but hypothetical Homo economicus. Behavioural finance is the study of the influence of psychology on the behaviour of financial practitioners and the subsequent effect on markets. Behavioural finance is of interest because it helps explain why and how markets might be inefficient. I host and run a popular behavioural finance website.

Back in 1896, Gustave le Bon wrote The Crowd: A Study of the Popular Mind, one of the greatest and most influential books of social psychology ever written [le Bon 1896]. Selden [1912] wrote Psychology of the Stock Market. He based the book ‘upon the belief that the movements of prices on the exchanges are dependent to a very considerable degree on the mental attitude of the investing and trading public’. In 1956 the US psychologist Leon Festinger introduced a new concept in social psychology: the theory of cognitive dissonance [Festinger et al. 1956]. When two simultaneously held cognitions are inconsistent, this will produce a state of cognitive dissonance. Because the experience of dissonance is unpleasant, the person will strive to reduce it by changing their beliefs. Herbert Simon first proposed the idea of bounded rationality, the notion that the rationality of individuals is limited by the information they have, their cognitive ability and the finite amount of time they have to make decisions [Simon 1955]. Pratt [1964] considered utility functions, risk aversion and also risks considered as a proportion of total assets.

Two psychologists, Amos Tversky and Daniel Kahneman, introduced the availability heuristic [Tversky and Kahneman 1973]: ‘a judgmental heuristic in which a person evaluates the frequency of classes or the probability of events by availability, i.e. by the ease with which relevant instances come to mind.’ The reliance on the availability heuristic leads to systematic biases. The following year, Tversky and Kahneman [1974] described three heuristics (including availability) that are employed when making judgments under uncertainty [Tversky and Kahneman 1974].

**representativeness** When people are asked to judge the probability that an object or event A belongs to class or process B, probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B.
availability  When people are asked to assess the frequency of a class or the probability of an event, they do so by the ease with which instances or occurrences can be brought to mind.

anchoring and adjustment  In numerical prediction, when a relevant value (an anchor) is available, people make estimates by starting from an initial value (the anchor) that is adjusted to yield the final answer. The anchor may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient.

From the field of economics, expected utility theory (also known as von Neumann-Morgenstern utility) (Bernoulli [1738], von Neumann and Morgenstern [1944], Bernoulli [1954]) is a normative model of decision making under risk. Expected utility theory states that when making decisions under risk people choose the option with the highest utility, where utility is the sum of the products of the utility of each potential outcome and the probability of occurrence of the outcome. The most cited paper ever to appear in Econometrica, the prestigious academic journal of economics, was written by the two psychologists Kahneman and Tversky [1979]. They present a critique of expected utility theory as a descriptive model of decision making under risk and develop an alternative model, which they call prospect theory. Prospect theory is considered in more detail on p. 38.

Thaler [1980] argues that there are circumstances when consumers act in a manner that is inconsistent with economic theory and he proposes that Kanneman and Tversky’s prospect theory be used as the basis for an alternative descriptive theory. Topics discussed are: underweighting of opportunity costs, failure to ignore sunk costs, search behaviour, choosing not to choose and regret, and precommitment and self-control. The paper introduced the notion of ‘mental accounting’ (described below). In another important paper Tversky and Kahneman [1981] introduced framing. They showed that the psychological principles that govern the perception of decision problems and the evaluation of probabilities and outcomes produce predictable shifts of preference when the same problem is framed in different ways.


In 1985 Werner F. M. De Bondt and Richard Thaler published ‘Does the stock market overreact?’ in the The Journal of Finance (De Bondt and Thaler [1985]), effectively forming the start of what has become known as behavioural finance. They discovered that people systematically overreacting to unexpected and dramatic news events results in substantial weak-form inefficiencies in the stock market. This was both surprising and profound. Mental accounting is the set of cognitive operations used by individuals and households to organize, evaluate and keep track of financial activities. Thaler [1985] developed a new model of consumer behaviour involving mental accounting. Tversky and Kahneman [1986] argue that, due to framing and prospect theory, the rational theory of choice does not provide an adequate foundation for a descriptive theory of decision making. In an interesting paper, Yaari [1987] proposed a modification to expected utility theory and obtains a so-called ‘dual theory’ of choice under risk. The dual theory separates an agent’s attitude towards risk from an agent’s attitude towards wealth, and has the property that utility is linear in wealth, but not in probabilities. De Bondt and Thaler [1987]
report additional evidence that supports the overreaction hypothesis. Samuelson and Zeckhauser (1988) perform a series of decision-making experiments and find evidence of a cognitive bias for the status quo, which they named the status quo bias. Poterba and Summers (1988) investigated transitory components in stock prices and found positive autocorrelation in returns over short horizons and negative autocorrelation over longer horizons, although random-walk price behaviour cannot be rejected at conventional statistical levels.

Kahneman et al. (1990) report several experiments that demonstrate that loss aversion and the endowment effect persist even in market settings with opportunities to learn and conclude that they are fundamental characteristics of preferences. Gilovich (1991) wrote How We Know What Isn’t So, a book about the fallibility of human reason in everyday life. Tversky and Kahneman (1991) present a reference-dependent model of riskless choice, the central assumption of the theory being loss aversion, i.e. losses and disadvantages have a greater impact on preferences than gains and advantages. Fernandez and Rodrik (1991) model an economy and show how uncertainty regarding the identities of gainers and losers can lead to status quo bias. Kahneman et al. (1991) discuss three anomalies: the endowment effect, loss aversion and status quo bias. Thaler (1992) published The Winner’s Curse: Paradoxes and Anomalies of Economic Life. Banerjee (1992) developed a simple model of herd behaviour. Tversky and Kahneman (1992) superseded their original implementation of prospect theory with cumulative prospect theory. Details are given on p. 38. Plous (1993) wrote The Psychology of Judgment and Decision Making which gives a comprehensive introduction to the field with a strong focus on the social aspects of decision making processes. Lakonishok et al. (1994) conjecture that value strategies yield higher returns because these strategies exploit the suboptimal behaviour of the typical investor. The equity premium puzzle refers to the empirical fact that stocks have outperformed bonds over the last century by a far greater degree than would be expected under the standard expected utility maximizing paradigm. Benartzi and Thaler (1995) offer an explanation based on behavioural concepts: loss aversion combined with a prudent tendency to frequently monitor one’s wealth. They dub this combination myopic loss aversion.

Grinblatt et al. (1995) analysed the behaviour of mutual funds and found evidence of momentum strategies and herding. Ghashghaie et al. (1996) claim that there is an information cascade in FX market dynamics that corresponds to the energy cascade in hydrodynamic turbulence. The study of heuristics and biases in judgment was criticized in several publications by G. Gigerenzer. Kahneman and Tversky (1996) reply and claim that contrary to the central criticism, judgments of frequency—not only subjective probabilities—are susceptible to large and systematic biases. Chan et al. (1996) found that both price and earnings momentum strategies were profitable, implying that the market responds only gradually to new information, i.e. there is underreaction. More recent developments in decision making under risk have improved upon cumulative prospect theory, such as the transfer of attention exchange model (Birnbaum and Chavez, 1997). In the accounting literature, Basu (1997) finds evidence for the conservatism principle, which he interprets as earnings reflecting ‘bad news’ more quickly than ‘good news’. Bikhchandani et al. (1998) argue that the theory of observational learning, and particularly of

---

6 A value strategy involves buying stocks that have low prices relative to earnings, dividends, book assets, or other measures of fundamental value.
informational cascades, can help explain phenomena such as stock market crashes. Motivated by a variety of psychological evidence, Barberis et al. (1998) present a model of investor sentiment that displays underreaction of stock prices to news such as earnings announcements and overreaction of stock prices to a series of good or bad news. In his third review paper Fama (1998) defends the efficient market hypothesis that he famously defined in his first, and claims that apparent overreaction of stock prices to information is about as common as underreaction. This argument is unconvincing, because under- and overreactions appear to occur under different circumstances and/or at different time intervals. Odean (1998) tested and found evidence for the disposition effect, the tendency of investors to sell winning investments too soon and hold losing investments for too long. Daniel et al. (1998) proposed a theory of security markets based on investor overconfidence (about the precision of private information) and biased self-attribution (which causes changes in investors’ confidence as a function of their investment outcomes) which leads to market under- and overreactions. Camerer and Lovallo (1999) found experimentally that overconfidence and optimism lead to excessive business entry. Wermers (1999) studied herding by mutual fund managers and he found the highest levels in trades of small stocks and in trading by growth-oriented funds. Thaler (1999) summarized the literature on mental accounting and concludes that mental accounting influences choice, that is, it matters. Gigerenzer et al. (1999) published Simple Heuristics That Make Us Smart, a book about fast and frugal heuristics. Odean (1999) demonstrated that overall trading volume in equity markets is excessive, and one possible explanation is overconfidence. He also found further evidence of the disposition effect which leads to profitable stocks being sold too soon and losing stocks being held for too long. Hong and Stein (1999) modelled a market populated by two groups of boundedly-rational agents: ‘newswatchers’ and ‘momentum traders’ which leads to underreaction at short horizons and overreaction at long horizons. Nofsinger and Sias (1999) found that institutional investors positive-feedback trade more than individual investors and institutional herding impacts prices more than herding by individual investors. Interestingly, Veronesi (1999) presented a dynamic, rational expectations equilibrium model of asset prices in which, among other features, prices overreact to bad news in good times and underreact to good news in bad times.

There is a commonly observed but unexpected negative correlation between perceived risk and perceived benefit. Finucane et al. (2000) concluded that this was due to the affect heuristic—people tend to derive both risk and benefit evaluations from a common source. Hong et al. (2000) proposed that firm-specific information, especially negative information, diffuses only gradually across the investing public, and this is responsible for momentum in stock returns. Shleifer (2000) published Inefficient Markets: An Introduction to Behavioral Finance, a quality book that considers behavioural finance vis-à-vis the EMH. In considering descriptive theories of choice under risk, Starmer (2000) reviews alternatives to expected utility theory. Shefrin (2000) wrote Beyond Greed and Fear, an excellent book on behavioural finance and the psychology of investing. In 2000, in his book Irrational Exuberance, Robert J. Shiller presented a persuasive case that the US stock market was significantly overvalued, citing structural factors, cultural factors and psychological factors. Shiller (2000), Kahneman and Tversky (2000) edited the book Choices, Values, and Frames, which presents a selection of the research that grew from their col-
laboration on prospect theory. Rabin (2000) provides a theorem showing that expected utility theory is an utterly implausible explanation for appreciable risk aversion over modest stakes. Lee and Swaminathan (2000) showed that past trading volume provides an important link between ‘momentum’ and ‘value’ strategies and these findings help to reconcile intermediate-horizon ‘underreaction’ and long-horizon ‘overreaction’ effects. Rabin and Thaler (2001) considered risk aversion and pronounce the expected utility hypothesis dead. Barber and Odean (2001) found that men trade 45 per cent more than women and thereby reduce their returns more so than do women and conclude that this is due to overconfidence. Barberis et al. (2001) incorporated prospect theory in a model of asset prices in an economy. Grinblatt and Keloharju (2001) identified the determinants of buying and selling activity and find evidence that past returns, reference price effects, tax-loss selling and the fact that investors are reluctant to realize losses are all determinants of trading. Barberis and Huang (2001) compared two forms of mental accounting by incorporating loss aversion and narrow framing into two asset-pricing frameworks: individual stock accounting and portfolio accounting. The former was the more successful. Gigerenzer and Selten (2001) edited Bounded Rationality: The Adaptive Toolbox, a collection of workshop papers which promote bounded rationality as the key to understanding how real people make decisions. The book uses the concept of an ‘adaptive toolbox,’ a repertoire of fast and frugal rules for decision making under uncertainty. Huberman (2001) provided compelling evidence that people have a propensity to invest in the familiar, while often ignoring the principles of portfolio theory. Gilovich et al. (2002) edited Heuristics and Biases: The Psychology of Intuitive Judgment, a book that compiles the most influential research in the heuristics and biases tradition since the initial collection in 1982 (Kahneman et al., 1982). In the Introduction Gilovich and Griffin (2002) identify six general purpose heuristics (affect, availability, causality, fluency, similarity and surprise) and six special purpose heuristics (attribution substitution, outrage, prototype, recognition, choosing by liking and choosing by default), whilst two heuristics have been superseded (representativeness (replaced by attribution-substitution (prototype heuristic and similarity heuristic)) and anchoring and adjustment (replaced by the affect heuristic)). Slovic et al. (2002) describe and discuss the affect heuristic: the specific quality of ‘goodness’ or ‘badness’. Daniel Kahneman won the 2002 Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel for his work on prospect theory, despite being a research psychologist and not an economist. If it were not for his untimely death, Amos Tversky, Kahneman’s collaborator, would have almost certainly shared the prize. Holt and Laury (2002) conducted a simple lottery-choice experiment and found differences in risk aversion between behaviour under hypothetical and real incentives. Barberis and Thaler (2003) published a survey of behavioural finance.

In Sewell (2009a) I explain that a responsible investment manager should seek a compromise between the normative expected utility theory and the prescriptive prospect theory, and call for a prescriptive model of risk preferences. Harrison and Rutström (2009) proposed a reconciliation of expected utility theory and prospect theory by using a mixture model. An economist would judge emotions as irrational, but in Sewell (2010) I explain that emotions likely evolved to maximize lifetime utility. In short, emotions help solve the prisoner’s dilemma by enabling us to cooperate. Wakker (2010) wrote the
2.2. Modelling

first book on prospect theory, it covers decision making under both known and unknown probabilities, and includes expected utility, rank-dependent utility and prospect theory. In [Sewell (2011g)] I speculate as to how the endowment effect, loss aversion, risk aversion, overconfidence, optimism, the representativeness heuristic, the availability heuristic and herding may have evolved, and focus on their effect on entrepreneurs and venture capitalists.

Important Heuristics

Affect  The affect heuristic concerns ‘goodness’ and ‘badness’. Affective responses to a stimulus occur rapidly and automatically: note how quickly you sense the feelings associated with the stimulus words treasure or hate.

Availability  Availability is a cognitive heuristic in which a decision maker relies upon knowledge that is readily available rather than examine other alternatives or procedures.

Similarity  The similarity heuristic leads us to believe that ‘like causes like’ and ‘appearance equals reality’. The heuristic is used to account for how people make judgments based on the similarity between current situations and other situations or prototypes of those situations.

Prospect Theory

Prospect theory is a descriptive model of decision making under risk, and a literature review was given on p. [34]. Kahneman and Tversky found empirically that people underweight outcomes that are merely probable in comparison with outcomes that are obtained with certainty; also that people generally discard components that are shared by all prospects under consideration. Under prospect theory, value is assigned to gains and losses rather than to final assets; also probabilities are replaced by decision weights. The value function is defined on deviations from a reference point and is normally concave for gains (implying risk aversion), commonly convex for losses (risk seeking) and is generally steeper for losses than for gains (loss aversion) (see Figure 2.1). Decision weights are generally lower than the corresponding probabilities, except in the range of low probabilities. The theory—which they confirmed by experiment—predicts a distinctive fourfold pattern of risk attitudes: risk aversion for gains of moderate to high probability and losses of low probability, and risk seeking for gains of low probability and losses of moderate to high probability.

As referenced on p. [35] Tversky and Kahneman [1992] superseded their original implementation of prospect theory with cumulative prospect theory. The new methodology employs cumulative rather than separable decision weights, applies to uncertain as well as to risky prospects with any number of outcomes, and it allows different weighting functions for gains and for losses (see Figure 2.2 (p. 39)). I have developed a cumulative prospect theory calculator, which is freely available online for the Web and Excel.

Note that there are two fundamental reasons why prospect theory (which calculates value) is inconsistent with expected utility theory. Firstly, whilst utility is necessarily linear in the probabilities, value is

http://prospect-theory.behaviouralfinance.net
2.2. Modelling

Figure 2.1: A hypothetical value function in prospect theory

Figure 2.2: Typical probability weighting functions for gains ($w^+$) and losses ($w^-$) in cumulative prospect theory
2.3. Forecasting

Multiagent Systems

The artificial stock market in Chapter 4 employs a multiagent system, which is defined and the concept criticised in Section 4.1.1 (p. 67). Two good books on multiagent system are Weiss (1999) and Wooldridge (2002). In a classic paper, Arthur et al. (1997) proposed a theory of asset pricing based on heterogeneous agents who continually adapt their expectations to the market that these expectations aggregatively create, thus creating an artificial stock market. LeBaron (2006) surveys research on agent-based models used in finance. Martinez-Jaramillo (2007) and Martinez-Jaramillo and Tsang (2009) developed an artificial financial market and modelled technical, fundamental and noise traders. They investigated the different conditions under which the statistical properties of an artificial stock market resemble those of a real financial market, and investigated the effects on the market when the agents learn. Railsback (2001) addresses the problem of getting ‘results’—general principles and conclusions—from multiagent systems and recommends a pattern-oriented approach.

Investment Performance Measurement

The secondary piece of research in Chapter 4 concerns investment performance measurement. Popular investment performance metrics include the Sharpe ratio (Sharpe, 1994), Sortino ratio (Sortino and van der Meer, 1991) and (less common) Omega (Shadwick and Keating, 2002). Omega has the advantage that it captures all of the moments of the returns distribution. Goetzmann et al. (2002) proved that an optimal (high) Sharpe ratio strategy would produce a distribution with a truncated right tail and a fat left tail.

2.3 Forecasting

This section is a review of the literature relevant to the experiments conducted in Chapter 5 on the forecasting of financial time series. Tsang (2009) reviews the work in computational intelligence on forecasting.

Kernel Methods

Central to the work on forecasting in Chapter 5 is the concept of a kernel. The technical aspects of kernels are dealt with in Section 5.1 (pp. 84–85), and the history is given here. The Fisher kernel is derived and implemented in Chapter 5 to save space, a thorough literature review is provided in Sewell (2011).

David Hilbert used the German word kern in his first paper on integral equations (Hilbert, 1904). The mathematical result underlying the kernel trick, Mercer’s theorem, is over a century old (Mercer, 1909). It tells us that any ‘reasonable’ kernel function corresponds to some feature space. The underlying mathematical results that allow us to determine which kernels can be used to compute distances in feature spaces was developed by Schoenberg (1938). The methods for representing kernels in linear spaces were first studied by Kolmogorov (1941) for a countable input domain. The method for repre-
senting kernels in linear spaces for the general case was developed by Aronszajn (1950). Dunford and Schwartz (1963) showed that Mercer’s theorem also holds true for general compact spaces. The use of Mercer’s theorem for interpreting kernels as inner products in a feature space was introduced into machine learning by Aizerman et al. (1964). The justification for a non-linear transformation followed by a linear transformation can be traced back to Cover (1965). The idea of polynomial kernels stems from Poggio (1975). Berg et al. (1984) published a good monograph on the theory of kernels. Micchelli (1986) discussed closure properties when making kernels. Saitoh (1988) showed the connection between positivity (a ‘positive matrix’ defined in Aronszajn (1950)) and the positive semi-definiteness of all finite set kernel matrices.

Reproducing kernels were extensively used in machine learning and ANNs by Tomaso Poggio and Federico Girosi, see for example Poggio and Girosi (1990), a paper on radial basis function networks. The theory of kernels was used in approximation and regularisation theory, and the first chapter of Spline Models for Observational Data (Wahba, 1990) gives a number of theoretical results on kernel functions. In a seminal paper, Boser et al. (1992) (re)introduced the notion of a kernel into the mainstream of the machine learning literature by combining kernel functions with large margin hyperplanes, leading to support vector machines (SVMs). They discussed Gaussian and polynomial kernels.

2.3. Forecasting


Support Vector Machines

Support vector machines (SVMs) are used extensively in the forecasting of financial time series and are covered in more detail in Section 5.1 (p. 86). Among other sources, the introductory paper (Hearst et al., 1998), the classic SVM tutorial (Burges, 1998), the excellent book (Cristianini and Shawe-Taylor, 2000) and the implementation details within Joachims (2002) have contributed to my own understanding.

The Application of Support Vector Machines to the Financial Domain

Gestel et al. (2000) developed the ‘volatility tube’ SVM and applied it to 1-day ahead prediction of the DAX 30 stock index, and significant positive out-of-sample results were obtained. Ahmed (2000)
used support vector regression for forecasting a foreign exchange rate time series. Guesde (2000) used SVMs for regression to predict the Canadian exchange rate (CAD), wisely recognized the problem of nonstationarity, dealt with it using experts and claimed that substantial profits were achieved. Fan and Palaniswami (2000) used SVMs to select bankruptcy predictors, and provided a comparative study. Trafalis and Ince (2000) compared two SVMs for regression (one implementing a primal-dual interior point quadratic programing (QP) algorithm and the other a standard QP algorithm) and two artificial neural networks (ANNs) (a backpropagation multilayer perceptron (MLP) and a radial basis function (RBF) network) by predicting IBM, Yahoo and America Online daily stock prices. Oddly, they forwent a validation set, and with the SVMs, set $\epsilon$ to zero, fixed $C$ and repeated the experiment for various fixed settings of the kernel parameter, $\sigma$, giving rise to several results. By considering either the best results or the average results for each of the four methods, the ranking was the same, from best to worst performance: 1st MLP, 2nd RBF, 3rd primal-dual SVM and 4th standard SVM.

Cao and Tay (2001b) dealt with the application of saliency analysis to feature selection for SVMs. Five futures contracts were examined and they concluded that saliency analysis is effective in SVMs for identifying important features. Tay and Cao (2001b) combined SVMs with a self-organizing feature map (SOM) and tested the model on Santa Fe Data Set C (high frequency USD/CHF data) and five real futures contracts. They showed that their proposed method achieves both significantly higher prediction performance and faster convergence speed in comparison with a single SVM model. Cao and Tay (2001a) found that SVMs forecast the S&P 500 daily price index better than a multilayer perceptron trained by the backpropagation algorithm. Gestel et al. (2001) applied the Bayesian evidence framework to least squares support vector machine (LS-SVM) regression to predict the weekly 90-day T-bill rate and the daily DAX 30 closing price returns. Tay and Cao (2001a) found that an SVM outperformed a multilayer backpropagation ANN on five real futures contracts, the S&P 500 stock index futures (CME-SP), US 30-year government bond (CBOT-US), US 10-year government bond (CBOT-BO), German 10-year government bond (EUREX-BUND) and French government stock index futures (MATIF-CAC40). Fan and Palaniswami (2001) used SVMs for classification for stock selection on the Australian Stock Exchange (ASX) and significantly outperformed the benchmark (the ‘market’).

Cao and Gu (2002) incorporate the prior knowledge that financial time series are non-stationary into their ‘dynamic support vector machines (DSVMs)’ and use an exponentially increasing regularization constant and an exponentially decreasing tube size to deal with structural changes in the data on the assumption that recent data points could provide more important information than distant data points. They conclude that DSVMs generalize better than standard SVMs in forecasting non-stationary time series, whilst they also use fewer support vectors, resulting in a sparser representation of the solution. In a similar paper, Tay and Cao (2002b) incorporated the problem domain knowledge of nonstationarity of financial time series into SVMs by using an adaptive tube in their so called ‘$\epsilon$-descending support vector machines ($\epsilon$-DSVMs)’. Experiment showed that $\epsilon$-DSVMs generalize better than standard SVMs in forecasting non-stationary financial time series and also converge to fewer support vectors, resulting in a sparser representation of the solution. Pang et al. (2002) found that, using film reviews as data,
standard machine learning techniques definitively outperformed human-produced baselines. However, they also found that the three machine learning methods they employed (naive Bayes, maximum entropy classification and SVMs) did not perform as well on sentiment classification as on traditional topic-based categorization. This paper is of interest because sentiment for stocks could be extracted from news articles or stock message boards on the Web. Abraham et al. (2002) analysed the performance of an ANN trained using the Levenberg-Marquardt algorithm, an SVM, a Takagi-Sugeno neuro-fuzzy model and a difference boosting neural network (DBNN) when predicting the NASDAQ-100 and the S&P CNX Nifty. There was no clear winner. Harland (2002a)/Harland (2002b) used an ensemble of SVMs for regression to trade the three month aluminium futures contract on the London Metal Exchange with positive results. Yang et al. (2002a) tried varying the margins in SVM regression in order to reflect the change in volatility of financial data and also analysed the effect of asymmetrical margins so as to allow for the reduction of the downside risk. The former approach produced the lowest total error when predicting the daily closing price of Hong Kong’s Hang Seng Index (HSI). Tay and Cao (2002a) developed ‘C-ascending’ SVMs, which penalise recent \( \epsilon \)-insensitive errors more heavily than distant \( \epsilon \)-insensitive errors, and found that they forecast better than standard SVMs on three real futures collected from the Chicago Mercantile Market. Cao and Chong (2002) considered the application of principal component analysis (PCA), kernel principal component analysis (KPCA) and independent component analysis (ICA) to SVMs for feature extraction. By examining sunspot data and one real futures contract, they showed that an SVM with feature extraction using PCA, KPCA or ICA can perform better than without feature extraction. Furthermore, they found that there is better generalization performance in KPCA and ICA feature extraction than PCA feature extraction. Yang et al. (2002b) used SVM regression with a non-fixed and asymmetrical margin, this time adapting the asymmetrical margins using momentum, and applied it to predicting the Hang Seng Index and the DJIA. Calvo and Williams (2002) compared the performance of ANNs, a naive Bayes classifier and SVMs for the automatic categorization of corporate announcements in the ASX Signal G data stream. The results were all good, but with the SVM underperforming the other two models.

Cao et al. (2003) applied kernel principal component analysis (KPCA) to an SVM for feature extraction. The authors examined sunspot data and one real futures contract, and found such feature extraction enhanced performance and also that KPCA was superior to PCA. Sansom et al. (2003b) evaluated utilizing ANNs and SVMs for wholesale (spot) electricity price forecasting. The SVM required less time to optimally train than the ANN, whilst the SVM and ANN forecasting accuracies were found to be very similar. Van Gestel et al. (2003a) compared four methodologies, ordinary least squares (OLS), ordinal logistic regression (OLR), the multilayer perceptron (MLP) and least squares support vector machines (LS-SVMs) when applied to credit scoring. The SVM methodology yielded significantly and consistently better results than the other rating methods. Van Gestel et al. (2003b) used least squares SVM classifiers for predicting bankruptcy of mid-cap firms in Belgium and the Netherlands. Cao (2003) showed that their method of ‘SVM experts’ achieved significant improvement above single SVM models when applied to Santa Fe Data Set C. Xu et al. (2003) applied wavelets and SVMs to short-term
2.3. Forecasting

electricity price forecasting, and achieved a reasonable forecasting accuracy. Pérez-Cruz et al. (2003) used SVMs for regression to estimate the parameters of a GARCH model for predicting the conditional volatility of stock market returns and showed that such estimates have a higher predicting ability than those obtained via common maximum likelihood methods. Again, Yang (2003) employed SVMs for regression and varied the width of the margin to reflect the change of volatility and controlled the symmetry of margins to reduce the downside risk. Results were positive. Similar to Abraham et al. (2002), Abraham et al. (2003) applied four different techniques, an ANN trained using the Levenberg-Marquardt algorithm, an SVM, a difference boosting neural network and a Takagi-Sugeno fuzzy inference system learned using an ANN algorithm (neuro-fuzzy model) to the prediction of the NASDAQ-100 and the S&P CNX Nifty. No one technique was clearly superior, but absurdly, they attempted to predict the absolute value of the indices, rather than use log returns. Sansom et al. (2003a) presented an analysis of the results of a study into wholesale (spot) electricity price forecasting with SVMs utilizing past price and demand data and Projected Assessment of System Adequacy (PASA) data. They concluded that the inclusion of PASA data showed little improvement in forecasting accuracy. Ongsritrakul and Soonthorn-phisan (2003) applied a decision tree algorithm for feature selection and then performed support vector regression to predict the gold price; their results were positive. Similar to Abraham et al. (2002) and Abraham et al. (2003), Abraham and AuYeung (2003) considered an ANN trained using the Levenberg-Marquardt algorithm, an SVM, a Takagi-Sugeno neuro-fuzzy model and a difference boosting neural network for predicting the NASDAQ-100 and the S&P CNX Nifty. They concluded that an ensemble of the intelligent paradigms performed better than the individual methods. The SVM outperformed the ANN. Gavrishchaka and Ganguli (2003) used SVMs for forecasting the volatility of foreign exchange data. Their preliminary benchmark tests indicated that SVMs can perform significantly better than or comparable to both naive and GARCH(1,1) models. Kim (2003) found that SVMs outperformed backpropagation ANNs and case-based reasoning when used to forecast the daily Korea Composite Stock Price Index (KOSPI). Cao and Tay (2003) used an SVM, a multilayer backpropagation (BP) ANN and a regularized radial basis function (RBF) ANN to predict five real futures contracts collated from the ‘Chicago Mercantile Market’ (sic). Results showed that the SVM and the regularized RBF ANN were comparable and both outperformed the BP ANN, with the SVM being best. Kamruzzaman et al. (2003) investigated the effect of different kernel functions and the regularization parameter when using SVMs to predict six different foreign currency exchange rates against the AUD.

Similar to Yang et al. (2002b), Yang et al. (2004b) used SVMs for regression with non-fixed and asymmetrical margin settings, this with momentum, to predict the Hang Seng Index and DJIA. Zhou et al. (2004) applied an accurate on-line support vector regression (AOSVR) to forecasting the prices of the electric-power markets, results showed that it was effective. Ince and Trafalis (2004) found that MLP ANNs outperform support vector regression when applied to stock price prediction. Huang et al. (2004) applied backpropagation ANNs and SVMs to corporate credit rating prediction for the US and Taiwan markets and found that the results were comparable (both were superior to logistic regression), with the SVM slightly better. Yang et al. (2004a) proposed a novel two-phase support vector regression training
procedure to detect and deflate the influence of outliers. The method was tested on the Hang Seng Index, NASDAQ and FSTE 100 index and results were positive. However, it’s not clear why the significance of outliers—such as market crashes—should be understated!

Shin et al. (2005) demonstrated that SVMs perform better than backpropagation ANNs when applied to corporate bankruptcy prediction. Van Gestel et al. (2005) apply linear and non-linear credit scoring by combining logistic regression and SVMs. Min and Lee (2005) found that, when applied to bankruptcy prediction, SVMs outperformed multiple discriminant analysis (MDA), logistic regression and three-layer fully connected backpropagation ANNs. Debnath and Giles (2005) devised a novel approach of extracting news metadata *HeadLines* using SVMs and using them to find story themes to get a sentence-based explanation for a stock price change. Gestel et al. (2005) developed credit rating systems by combining linear ordinal logistic regression and fixed-size least squares SVMs. Bao et al. (2005) used fuzzy support vector machines regression (FSVMR) to forecast a data set from the Shanghai Stock Exchange with positive results. Chen and Ho (2005) used an SVM for regression for forecasting the Taiwan Stock Exchange Capitalization Weighted Stock Index. Oddly, they considered price, rather than returns. The results demonstrated that the SVM outperformed the backpropagation ANN and random walk models. Hovsepian and Anselmo (2005) used SVMs for classification to predict relative volatility clusters and achieved accurate and robust results. Huang et al. (2005) compared the ability of SVMs, linear discriminant analysis, quadratic discriminant analysis and Elman backpropagation ANNs to forecast the weekly movement direction of the Nikkei 225 index and found that the SVM outperformed all of the other classification methods. Better still was a weighted combination of the models. Pasiouras et al. (2005) considered the prediction of acquisition targets within the EU banking industry acquired between 1998 and 2002 and compared and evaluated seven classification methodologies (discriminant analysis, logit analysis, UTilités Additives DIScriminantes (UTADIS), Multi-group Hierarchical Discrimination (MHDIS), classification and regression trees (CART), k-nearest neighbour (k-NN) and SVMs) and found that discriminant analysis and SVMs performed best. Pai and Lin (2005) proposed a hybrid ARIMA and SVM model for stock price forecasting, and results looked very promising.

Yu et al. (2006a) applied a random walk (RW) model, an autoregressive integrated moving average (ARIMA) model, an individual backpropagation ANN model, an individual SVM model and a genetic algorithm-based SVM (GASVM) to the task of predicting the direction of change in the daily S&P 500 stock price index and found that their proposed GASVM model performed the best, and the SVM second best. Chen et al. (2006b) compared SVMs and backpropagation (BP) ANNs when forecasting the six major Asian stock markets. Both models performed better than the benchmark AR(1) model in the deviation measurement criteria, whilst SVMs performed better than the BP model in four out of six markets. Chen and Shih (2006) used an SVM to classify Taiwan’s issuer credit ratings and found that it performed better than the backpropagation ANN model. Gavrishchaka and Banerjee (2006) used SVMs for forecasting stock market volatility with positive results. Hui and Sun (2006) applied an SVM, Fisher discriminant analysis, logistic regression and a backpropagation ANN to the early-warning of financial distress. The SVM obtained a better balance among fitting ability, generalization ability and
model stability than the other models. Hao and Yu (2006) applied support vector regression to stock composite index forecasting, and presented a preprocessing method for accelerating the training, and a method of modifying the regularized risk function. By using their methods, high prediction performance was achieved. Li et al. (2006) used SVMs for the evaluation of consumer loans, and also developed a visual decision-support tool. They found that the SVM surpasses traditional ANN models in generalization performance and visualization via the visual tool. Pai et al. (2006) developed a hybrid SVM model composed of a linear SVM and a non-linear SVM, furthermore the parameters of both were determined by genetic algorithms. Their approach outperformed an ANN, a chaotic model (vector-valued, local linear approximation) and a random walk model when predicting exchange rates. Chen et al. (2006a) used an SVM to estimate default probabilities of German firms. The SVM predicted a firm’s default risk and identified the insolvent firms more accurately than the benchmark logit model. Yu et al. (2006b) classified the impact of news on a stock price by incorporating time series analysis as domain knowledge to help preprocess the data for an SVM. Xie et al. (2006) compared SVMs with ARIMA and a backpropagation ANN for crude oil price prediction; the SVM outperformed the other two methods. Zhang et al. (2006) used SVMs for classification with analysts’ recommendations in an attempt to forecast whether a stock price would rise or fall the following day. Schumaker and Chen (2006) examined the role of financial news articles using three different textual representations, bag of words, noun phrases and named entities, and their ability to predict discrete number stock prices twenty minutes after an article release. An SVM derivative was shown to be superior to linear regression. Min et al. (2006) used a genetic algorithm to optimize both the feature subset and the parameters of an SVM simultaneously, and applied the technique to the prediction of bankruptcy in Korean companies. Their method outperformed logistic regression, an ANN and a standard SVM, whilst the standard SVM was the second best. Nalbantov et al. (2006) used SVMs for regression for equity style timing with positive results. Wu et al. (2007) used a real-valued genetic algorithm to optimize the parameters $C$ and $\sigma$ of an SVM for predicting bankruptcies in Taiwan. Their method achieved better predictive accuracy than a traditional SVM, discriminant analysis, logit analysis, probit regression and a feed-forward backpropagation ANN. The traditional SVM beat the ANN on the holdout sample. Chiu and Chen (2007) built a model integrating fuzzy theory, genetic algorithms and an SVM to classify the movements of the stock market in Taiwan. In an exercise in estimating the probability of a firm defaulting, Hardle et al. (2007) used SVMs, discriminant analysis and logit regression to rate firms. The SVM rating model dominated the other two. Sai et al. (2007) used rough sets for feature selection with an SVM for classification to predict the daily ‘H&S 300 stock price index’. Their model outperformed a random walk model, an ARIMA model, a backpropagation ANN and a standard SVM, with the standard SVM being the second best. Lee (2007) tackled the problem of corporate credit rating prediction, and found that SVMs outperformed multiple discriminant analysis, case-based reasoning and a three-layer fully connected backpropagation ANN. Gao et al. (2007) made the parameters of an SVM for regression flexible, and when forecasting the electricity market outperformed the classic SVM regression model. Hua et al. (2007) predicted corporate financial distress by integrating an SVM for classification with logistic regression, and their method...
outperformed a conventional SVM. Huang et al. (2007) applied SVMs to credit scoring and found that compared with ANNs, genetic programming and decision tree classifiers, the SVM classifier achieved an identical classificatory accuracy with relatively few input features. Martens et al. (2007) developed rule extraction for SVMs and applied the technique to credit scoring. Their technique lost only a small percentage in performance compared to a black box (standard) SVM, so they claim that it ranks ‘at the top of comprehensible classification techniques’.

Ding et al. (2008) applied multiple discriminate analysis, logistic regression, a three-layer backpropagation ANN and an SVM to forecasting financial distress of Chinese listed companies, and the SVM performed best. De Choudhury et al. (2008) successfully used an SVM for regression with communication dynamics in the blogosphere as inputs to predict the movement of stock prices over the following week. Yu et al. (2008) consider the application of SVMs and other computational intelligence techniques to credit risk analysis, and state that ‘[f]rom the investigation on 32 papers… support vector machines always have the best accuracy among the top methods.’ Powell et al. (2008) used $k$-means clustering and SVMs to classify S&P 500 stocks into increasing or decreasing prices on a weekly basis. The reported results appear too good to be true (e.g. the best result was an accuracy of 89.1 per cent for an SVM with a linear kernel), but on average the SVMs performed better than $k$-means clustering. Xin and Gu (2008) describe a method based on the least square SVM, which changes the inequality restriction in the traditional SVM into an equality restriction and uses the loss function of the quadratic sum of the errors as the empirical function for the training set. The quadratic programming problem is converted into one of solving linear equations, which significantly improves the training speed and the convergence accuracy. By way of example, the authors used their method to successfully predict a stock index, which resulted in faster training time and better accuracy than an ANN model. Pasiouras et al. (2008) investigated the relative performance of both linear and non-linear SVMs with a polynomial and an RBF kernel in the development of classification models for the prediction of acquisition targets in the EU banking sector. The differences between the models was marginal. Their results were positive, but they did not calculate excess returns. Chen and Hsiao (2008) used a genetic algorithm to determine the optimum parameters for an SVM for classification used to diagnose business crises in Taiwan. Their model achieved a high degree of accuracy. Ince and Trafalis (2008) used an SVM for regression and a multilayer perceptron ANN to predict daily stock prices of ten companies traded on the NASDAQ, and compared the results with ARIMA. On average, the SVM was the best technique.

Yu et al. (2008) used a genetic algorithm (GA) to select input features, and another GA for parameter optimization, for a least squares SVM applied to the classification of monthly S&P 500, DJIA and NYSE returns. Their method was superior to an autoregressive integrated moving average (ARIMA) model, a linear discriminant analysis (LDA) model, a backpropagation ANN and a standard SVM. The standard SVM beat the ANN. Bellotti and Crook (2009) apply an SVM, logistic regression, linear discriminant analysis and the k-nearest neighbour algorithm to credit scoring. The SVM with a linear or Gaussian kernel performed best, yielding the highest area under the receiver operating characteristic (ROC) curve. Boyacioglu et al. (2008) used four ANN architectures (multilayer perceptron (MLP), competitive learn-
2.3. Forecasting

(MLP) and self-organizing map (SOM), learning vector quantization (LVQ)), SVMs, multivariate discriminant analysis (MDA), k-means cluster analysis (CA) and logistic regression analysis (LRA) to predict bank financial failures. The best three, in order of performance on their (out of sample) validation set, were LVQ, MLP and SVMs. Chiu and Chen (2009) proposed a model integrating fuzzy theory, a genetic algorithm and an SVM to explore movements in the stock market in Taiwan. Huang and Tsai (2009) introduced a hybrid support vector regression/self-organizing feature map technique with filter-based feature selection to reduce training time and to improve prediction accuracies. They used their method to forecast Taiwan index futures (FITX), and it was shown to be an improvement over the traditional SVM for regression in average prediction accuracy and training time. Tang et al. (2009) constructed a multidimensional wavelet kernel function and applied it to forecasting the conditional volatility of stock market returns where their wavelet kernel outperformed a Gaussian kernel. Lu et al. (2009) proposed a two-stage forecasting model. First, independent component analysis (ICA) was used to remove the independent components (ICs) containing the noise, then the rest of the ICs were used to reconstruct the less noisy input variables for an SVM for regression model. Their method outperformed traditional support vector regression and a random walk model when forecasting the Nikkei 225 opening cash index and the TAIEX closing cash index. Sapankevych and Sankar (2009) present a general survey of SVM applications to time series prediction, and summarize 66 papers. Financial market prediction was the most studied application (21 papers). Grey systems theory is a theory dealing with those systems for which we lack information. Hui and Zhizhong (2009) proposed an approach for predicting financial time series based on a grey model integrated with support vector regression.

In an exercise in cross-sectional stock return analysis, Liu et al. (2010) analysed the relationship between stock returns and explanatory factors. They found that support vector regression outperformed ordinary least squares regression. Wang and Zhu (2010) proposed a two-step kernel learning method for SVM regression. Their method learns a sparse linear combination of a set of candidate kernels using the $L_1$-norm regularization approach. Since the regularization parameter must be carefully selected to facilitate parameter tuning, they develop an efficient solution path algorithm that solves the optimal solutions for all possible values of the regularization parameter. Their kernel learning method was applied to forecast the S&P 500 and NASDAQ market indices and showed promising results, but did not beat transaction costs. The stock ‘box theory’ apparently assumes that successful stock buying/selling generally occurs when the price effectively breaks out of an ‘oscillation box’ and into a new ‘box’. Wen et al. (2010) proposed a trading system based on oscillation box prediction by combining stock box theory and an SVM. In the system, two SVMs for regression are utilized to make forecasts of the upper bound and lower bound of the price oscillation box, then a trading strategy based on the two bound forecasts is constructed to make trading decisions. Their system significantly outperformed a buy-and-hold strategy on S&P 500 constituent stocks. Yu et al. (2010) proposed an SVM-based multiagent ensemble learning approach for credit risk evaluation. In an experiment on a British credit card application approval dataset their proposed ALNN-based multiagent ensemble learning model (ALNN) was compared with five popular classification models: linear discriminant analysis (LDA), quadratic discriminant analysis
2.3. Forecasting

(QDA), logit regression (LogR), a feed-forward neural network (FNN) and an SVM, as well as two other multiagent learning models: a majority voting based multiagent ensemble learning model (MV) and a TA-based weight averaging multiagent ensemble learning model (TA). The results were, from best to worst: ALNN, TA, MV, SVM, LogR, QDA, LDA, FNN. [Fu (2010)] forecasted the weekly EUR/CNY exchange rate using EMD-based support vector regression. His methodology first decomposes the original time series into a finite and often small number of sub-signals by empirical mode decomposition (EMD), then each sub-signal was modelled and forecast using an SVM for regression. The final forecast was obtained by aggregating all prediction results of sub-signals. His model outperformed a normal SVM for regression. [Pai et al. (2010)] combined rough set theory (RST) and directed acyclic graph support vector machines (DAGSVM) to predict the daily USD/JPY exchange rate. Their model greatly improved the prediction performance of either the RST model or the DAGSVM model on its own. [Zeng-min and Chong (2010)] used an SVM and a three-layer fully connected backpropagation neural network (BNN) to forecast the S&P 500 and the Nikkei 225, and found that the SVM outperformed the BNN. [Huang et al. (2010)] implemented a chaos-based SVM for regression applied to daily exchange rate forecasting. Firstly, the delay coordinate embedding was used to reconstruct the unobserved phase space (or state space) of the exchange rate dynamics, then an SVM was used for forecasting. Their proposed method performed better than traditional ANNs, traditional SVMs or chaos-based ANNs. The traditional SVM ranked second.

[Wang (2011)] applied a stochastic volatility model with jumps to support vector regression for pricing currency options. His method performed better than the Garman-Kohlhagen (GK) model and a hybrid ANN-stochastic volatility model.

Of the 38 articles above that compare SVMs with ANNs in the financial domain, SVMs outperformed ANNs in 32 cases, ANNs outperformed SVMs in 3 cases, and there was no significance difference in 3 cases. More specifically, of the 22 articles that concern the prediction of financial or commodity markets, 18 favoured SVMs, 2 favoured ANNs and 2 found no significant difference. This bodes well for SVMs, and as such, the research on forecasting shall employ them.

Evolutionary Computation

A genetic algorithm (GA) is a search technique that falls within the intersection of optimization algorithms and evolutionary algorithms. GAs became a widely recognized optimization method as a result of the work of John Holland in the early 1970s [Holland (1975)].

Genetic programming (GP) is an evolutionary algorithm that optimizes a population of computer programs according to a fitness landscape determined by a program’s ability to perform a user-defined task. The first experiments with GP were reported by [Smith (1980)] and [Cramer (1985)], and the seminal book is [Koza (1992)].

GP generalizes GAs, but [Woodward (2003)] argues that the two are effectively equivalent and it is the representation that is important. On average, GAs/GP are no better or worse than any other search/optimization algorithm [Wolpert and Macready (1997)].
The Application of Evolutionary Computation to Financial Time Series Prediction

Neely et al. (1997) used genetic programming to find technical trading rules for six exchange rates over the period 1981–1995 and found strong evidence of economically significant out-of-sample excess returns. Allen and Karjalainen (1999) used genetic algorithms to learn technical trading rules for the S&P 500, but failed to beat a buy-and-hold strategy after transaction costs. Kaboudan (2000) used GP to predict the price of six stocks. Oddly, he concluded that predicting price is easier than predicting returns and the out of sample period was a mere fifty trading days. However, he claims that his trading strategy ‘yielded relatively high returns on investment’. Potvin et al. (2004) employed genetic programming as a means to automatically generate short-term trading rules to trade the stocks of 14 Canadian companies listed on the Toronto Stock Exchange. stock markets and concluded that ‘results show that the trading rules generated by GP are generally beneficial when the market falls or when it is stable. On the other hand, these rules do not match the buy-and-hold approach when the market is rising.’ A meta-analysis by Park and Irwin (2004) found that genetic programming worked well on currency markets, but performed poorly on stock markets and futures markets.
Chapter 3

Characterization

The current plus the following two chapters (chapters 3, 4 and 5) make up the core of the thesis and contain the bulk of the contributions. The work in the present chapter seeks to extend the literature on the characterization of financial time series. In economics parlance, the statistics that make up the characterization of financial markets are known as ‘stylized facts’. This chapter first defines the term, then gives a list of stylized facts. The central paradigm in finance is the ‘efficient market hypothesis’ (EMH), and an efficient market (along with assumptions about investors’ risk preferences) puts constraints on what is possible vis-à-vis the characterization of financial markets. Conversely, the characterization of financial markets (again with assumptions about investors’ risk preferences) allows us to gauge the efficiency of financial markets. An understanding of the concept of a martingale is necessary for a thorough understanding of what an efficient market does and does not imply about the process generating a market price under risk neutrality, so the term is defined. A record of market price forms a time series, so a note on time series is included.

This chapter describes four experiments. The first and most straightforward is a measurement of the autocorrelation of stock market returns. For the following two experiments, two pieces of software are written, a program for performing the runs test and a program for testing for the existence of long memory, and both are used to analyse the dependence of stock market returns. The fourth experiment involves the analysis of the performance of investment newsletters. All four experiments (potentially) have implications apropos market efficiency. The characterization of financial time series provides us with the all-important domain knowledge that machine learning, employed in the chapter on forecasting, relies upon.

3.1 Background

3.1.1 Stylized Facts

A stylized fact is a term used in economics to refer to empirical findings that are so consistent (for example, across a wide range of instruments, markets and time periods) that they are accepted as truth. Due to their generality, they are often qualitative. The literature review in Section 2.1 on the characterization of financial time series included articles on dependence (pp. 24–25), long memory (pp. 25–28) and investment newsletters (p. 28). A thorough review of the literature on the characterization of financial markets
was conducted and is summarized below.

**Dependence** Autocorrelation in returns is largely insignificant, except at high frequencies when it becomes negative.

**Distribution** Approximately symmetric, increasingly positive kurtosis as the time interval decreases and a power-law or Pareto-like tail.

**Heterogeneity** Non-stationary (clustered volatility).

**Non-linearity** Non-linearities in mean and (especially) variance.

**Scaling** Markets exhibit non-trivial scaling properties.

**Volatility** Volatility exhibits positive autocorrelation, long-range dependence of autocorrelation, scaling, has a non-stationary log-normal distribution and exhibits non-linearities.

**Volume** Distribution decays as a power law, also calendar effects.

**Calendar effects** Intraday effects exist, the weekend effect seems to have all but disappeared, intramonth effects have been found in most countries, the January effect has halved, and holiday effects exist in some countries.

**Long memory** About 30 per cent of the articles analysing stock market returns concluded that they exhibit long memory, about 50 per cent of those that analysed foreign exchange returns concluded that they exhibit long memory, and about 80 per cent of those analysing market volatility concluded that it exhibits long memory.

**Chaos** There is little evidence of low-dimensional chaos in financial markets.

Early claims made for stable distributions, long memory in returns and chaos theory turned out to be largely unfounded as higher-frequency data became available. This evidences the importance of a data-driven approach.

In probability theory, a **martingale** is a model of a fair game where no knowledge of past events can help to predict future winnings. In particular, a martingale is a sequence of random variables (i.e., a stochastic process) for which, at a particular time in the realized sequence, the expectation of the next value in the sequence is equal to the present observed value even given knowledge of all prior observed values at a current time.

**Stochastic Processes in Financial Markets**

As bias-free learning is futile, domain knowledge is paramount. For this reason it is important to know as much as possible about the nature of financial markets. The concept of randomness is central to finance (and asset pricing models in particular) and this is formalized by the mathematics of stochastic processes. Table 3.1 (p. 55) gives a summary as to what extent the various random processes relate to financial markets.

---

1For the full review, see Sewell (2011a).
### Table 3.1: Stochastic processes and their applicability to markets. Note that it is the logarithm of the price of an asset (Osborne, 1959) that is of interest.

<table>
<thead>
<tr>
<th>Stochastic process</th>
<th>Description</th>
<th>Applicability to markets</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>diffusion process</td>
<td>satisfies the diffusion equation</td>
<td>poor</td>
<td>Regnault (1863) and Osborne (1959) discovered that price deviation is proportional to the square root of time, but the nonstationarity found by Kendall (1953), Houthakker (1961) and Osborne (1962) compromises the significance of the process.</td>
</tr>
<tr>
<td>Gaussian process</td>
<td>increments normally distributed</td>
<td>poor</td>
<td>Financial markets exhibit leptokurtosis (Mitchell, 1915, 1921; Olivier, 1926; Mills, 1927; Osborne, 1939; Larson, 1960; Alexander, 1961). For example, the kurtosis of daily returns of large cap stocks is of the order of 5 (Taylor, 2005, p. 53).</td>
</tr>
<tr>
<td>Lévy process</td>
<td>stationary independent increments</td>
<td>poor</td>
<td>Kendall (1953), Houthakker (1961) and Osborne (1962) found nonstationarities in markets in the form of positive autocorrelation in the variance of returns.</td>
</tr>
<tr>
<td>Markov process</td>
<td>memoryless</td>
<td>poor</td>
<td>Kendall (1953), Houthakker (1961) and Osborne (1962) found positive autocorrelation in the variance of returns.</td>
</tr>
<tr>
<td>martingale</td>
<td>zero expected return</td>
<td>submartingale: good for stock market</td>
<td>Bachelier (1900) and Samuelson (1965) recognised the importance of the martingale in relation to an efficient market. Whilst Cox and Ross (1976), Lucas (1978) and Harrison and Kreps (1979) pointed out that in practice investors are risk averse, so (presumably as compensation for the time value of money and systematic risk) they demand a positive expected return. In a long-only market like a stock market this implies that the price of a stock follows a submartingale (a martingale being a special case when investors are risk-neutral).</td>
</tr>
<tr>
<td>random walk</td>
<td>discrete version of Brownian motion</td>
<td>poor</td>
<td>LeRoy (1973) and (especially) Lucas (1978) pointed out that a random walk is neither necessary nor sufficient for an efficient market.</td>
</tr>
<tr>
<td>Wiener process (Brownian motion)</td>
<td>continuous-time, Gaussian independent increments</td>
<td>poor</td>
<td>Bachelier (1900) developed the mathematics of Brownian motion and used it to model financial markets. Note that Brownian motion is a diffusion process, a Gaussian process, a Lévy process, a Markov process and a martingale. On the one hand this makes it a very strong condition (and therefore the least realistic), on the other hand it makes it a very important 'generic' stochastic process and is therefore used extensively for modelling financial markets (for example, option pricing (Black and Scholes, 1973)).</td>
</tr>
</tbody>
</table>
3.1.2 Time Series

In a market, whenever buyers and sellers trade, it makes sense to record the agreed price at which the transaction took place. This price record creates a time series. For more information on time series analysis, the fourth edition of *Time Series Analysis: Forecasting and Control* (Box et al., 2008) is a revision of the classic 1970 book, Hamilton (1994)’s tome is the bible, whilst Weigend and Gershenfeld (1994) is the most relevant in terms of using advanced methods for time series prediction. For a time series glossary, see Appendix C (pp. 135–136).

3.1.3 Efficient Market Hypothesis

A literature review of the efficient market hypothesis was given in Section 2.1 (p. 23). The notion of an efficient market is central to this thesis and is dealt with in detail here. It employs the concept of a martingale. I host and run the world’s only website dedicated to the efficient market hypothesis.

Markets

Whenever there are valuable commodities to be traded, there are incentives to develop a social arrangement that allows buyers and sellers to discover information and carry out a voluntary exchange more efficiently, i.e. develop a market. The largest and best organized markets in the world tend to be the securities markets.

Efficiency in Economics

The concept of efficiency in economics is a general term for the value assigned to a situation by some measure designed to capture the amount of waste or ‘friction’ or other undesirable economic features present. Within this context, it has several quite distinct meanings. For example, allocative efficiency is concerned with the optimal distribution of scarce resources among individuals in the economy. An efficient portfolio is one with the highest expected return for a given level of risk. An efficient market is one in which information is rapidly disseminated and reflected in prices. It is this third definition that is of interest here.

Important

The EMH has been the central proposition of finance since the early 1970s and is one of the most controversial and well-studied propositions in all the social sciences.

Key Papers

The key papers that have contributed to the EMH are listed in Appendix D (page 137).

Definitions

As the efficient market hypothesis is of central importance to this thesis, it makes sense to review the literature in search of a definitive and rigorous definition of what it means for a market to be ‘efficient’. As the list below shows, it turns out that the definition of an efficient market has evolved over time, which some may interpret as economists moving the goalposts.

[http://www.e-m-h.org]
3.1. Background

**Fama (1965b)** ‘...an “efficient” market for securities, that is, a market where, given the available information, actual prices at every point in time represent very good estimates of intrinsic values.’

**Fama (1965a)** ‘An “efficient” market is defined as a market where there are large numbers of rational, profit-maximizers actively competing, with each trying to predict future market values of individual securities, and where important current information is almost freely available to all participants.’

**Fama et al. (1969)** ‘an “efficient” market, i.e. a market that adjusts rapidly to new information.’

**Fama (1970)** ‘A market in which prices always “fully reflect” available information is called “efficient.”’

**Jensen (1978)** ‘A market is efficient with respect to information set \( \theta_t \) if it is impossible to make economic profits by trading on the basis of information set \( \theta_t \).’

[‘By economic profits, we mean the risk adjusted returns net of all costs.’]

**Fama (1991)** ‘I take the market efficiency hypothesis to be the simple statement that security prices fully reflect all available information. A precondition for this strong version of the hypothesis is that information and trading costs, the costs of getting prices to reflect information, are always 0 (Grossman and Stiglitz (1980)). A weaker and economically more sensible version of the efficiency hypothesis says that prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed marginal costs (Jensen (1978)).’

**Malkiel (1992)** ‘A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set, \( \phi \), if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set, \( \phi \), implies that it is impossible to make economic profits by trading on the basis of \( \phi \).’

**Fama (1998)** ‘...market efficiency (the hypothesis that prices fully reflect available information)...’

‘...the simple market efficiency story; that is, the expected value of abnormal returns is zero, but chance generates deviations from zero (anomalies) in both directions.’

**Timmermann and Granger (2004)** ‘A market is efficient with respect to the information set, \( \Omega_t \), search technologies, \( S_t \), and forecasting models, \( M_t \), if it is impossible to make economic profits by trading on the basis of signals produced from a forecasting model in \( M_t \) defined over predictor variables in the information set \( \Omega_t \) and selected using a search technology in \( S_t \).’

**Taxonomy**

The classic taxonomy of information sets, due to **Roberts (1967)** and published by **Fama (1970)**, consists of the following:

**weak form efficiency** The information set includes only the history of prices.
3.1. Background

**semi-strong form efficiency** The information set includes all information known to all market participants (publicly available information).

**strong form efficiency** The information set includes all information known to any market participant (private information).

Note that the sets are nested, with each successive set being a superset of the preceding set. Later, the weak form was redefined by Fama (1991) to include variables like dividend yields and interest rates.

**Scope**
The term ‘efficient market’ was initially applied to the stock market, but the concept was soon generalized to other asset markets.

**Starting Point**
Regardless of whether or not one believes that markets are efficient, or even whether they are efficient, the efficient market hypothesis is almost certainly the right place to start when thinking about asset price formation. One can then consider relative efficiency.

**Two Informal Explanations for Market Efficiency**
The intrinsic value of an asset is implied by the cumulative impact of information received as news. Successive news items must be random because if an item of news were not random, that is, if it were dependent on an earlier item of news, then it wouldn’t be news at all. After all, news—by definition—is new. If the price rapidly reflects all information (i.e. the market is efficient), then the price will fluctuate randomly in accordance with a martingale.

If one could be sure that a price will rise tomorrow, the asset would be bought today, raising the price, so that it will not, in fact, rise tomorrow. Ergo, the price is unpredictable. More formally, this justification relies on the law of iterated expectations and leads to a martingale.

**Academics Versus Practitioners**
There is little consensus between the opinions held in academia and industry. Unsurprisingly, most of the support for the EMH comes from the former.

**Major Contributors Unrecognised**
The original unsung hero of the EMH, Louis Bachelier, has finally gained the recognition that he deserves. For example, the Bachelier Finance Society came into being in 1996. However, the American astrophysicist, M. F. Maury Osborne, was the first to publish the hypothesis that the logarithm of prices (and thus log returns) follows Brownian motion (Osborne, 1959), and was jointly responsible for the earliest literature identifying the fat tails (Osborne, 1959), that price deviation is proportional to the square root of time (Osborne, 1959) and nonstationarity (Osborne, 1962). However, the key concept is the martingale, so I reserve the most credit for Bachelier (1900) and Samuelson (1965).

**Joke**
“There is an old joke, widely told among economists, about an economist strolling down the street with a companion when they come upon a $100 bill lying on the ground. As the companion reaches down to
pick it up, the economist says ‘Don’t bother – if it were a real $100 bill, someone would have already picked it up’.” Lo (1997)

Rational?

Contrary to popular belief, the EMH does not require that all market participants are rational. Indeed, markets can be efficient even when a group of investors are irrational and correlated, so long as there are some rational traders present together with arbitrage opportunities. See Friedman (1953) and Shleifer (2000).

Paradoxical

EMH holds ⇒ prices always ‘fully reflect’ available information ⇒ many transactions ⇒ people believe EMH to be false ⇒ many transactions ⇒ EMH holds ⇒ etc.

If the EMH holds, it is because people believe it to be false. The paradox is resolved by allowing for irrational behaviour.

Not Possible

Beja (1977), (most importantly) Grossman and Stiglitz (1980) and Tirole (1982) argued that because information is costly, prices cannot perfectly reflect the information which is available, since if this was the case, those who spent resources to obtain it would receive no compensation, leading to the conclusion that an informationally-efficient market is impossible.

Not Refutable

An efficient market will always ‘fully reflect’ available information, but in order to determine how the market should ‘fully reflect’ this information, it is necessary to determine investors’ risk preferences. Therefore, any test of the EMH is a test of both market efficiency and investors’ risk preferences. For this reason, the EMH, by itself, is not a well-defined and empirically refutable hypothesis. This ‘joint hypothesis problem’ was first pointed out by Fama (1970).

Testing Market Efficiency

In theory, if not in practice, market efficiency can be tested. If information is revealed to market participants, the reaction of security prices can be measured. If and only if prices do not move when the information is revealed, the market is efficient with respect to that information set. See Campbell et al. (1996, pp. 21–22).

Increasing Efficiency?

Although only one paper published before 1960, Cowles and Jones (1937), found significant market inefficiencies; with decreasing transaction costs, an increasing number of market participants, increasing processing power and improving algorithms, one would expect markets to become increasingly efficient. The relative proportion of the papers summarized in Sewell (2011d) that reject the EMH peaked in the 1980s and 1990s, and Kim et al. (1991), Schwert (2003) and Tóth and Kertész (2006) suggest that markets are becoming increasingly efficient. It could be that markets in the 1980s and 1990s were less efficient because they were the decades of technological asymmetry: some market participants used
microcomputers, whilst others did not. It could also be that it took until the 1980s/1990s for data to be of sufficient quality and quantity to reject market efficiency with any degree of confidence. Analyses post-2000 tend to support market efficiency simply because markets have become increasingly efficient. The Red Queen effect ensures that one’s ability to make money in the markets is dependent on the ability of the other market participants: the game is relative and moving.

Conclusion

Just under half of the papers reviewed in Sewell (2011d) support market efficiency, with most of the attacks on the EMH coming in the 1980s and 1990s. Recall that a market is said to be efficient with respect to an information set if the price ‘fully reflects’ that information set (Fama, 1970). On the one hand, the definitional ‘fully’ is an exacting requirement, suggesting that no real market could ever be efficient, implying that the EMH is almost certainly false. On the other hand, economics is a social science, and a hypothesis that is asymptotically true puts the EMH in contention for one of the strongest hypothesis in the whole of the social sciences. Strictly speaking the EMH is false, but in spirit is profoundly true. Besides, science concerns seeking the best hypothesis, and until a flawed hypothesis is replaced by a better hypothesis, criticism is of limited value. I conclude matters on efficient markets with a few pertinent quotes.

- ‘I believe there is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Market Hypothesis.’ Jensen (1978)
- ‘If the efficient markets hypothesis was a publicly traded security, its price would be enormously volatile.’ Shleifer and Summers (1990)
- ‘It is disarmingly simple to state, has far-reaching consequences for academic pursuits and business practice, and yet is surprisingly resilient to empirical proof or refutation.’ Lo (1997)
- ‘Market efficiency survives the challenge from the literature on long-term return anomalies. Consistent with the market efficiency hypothesis that the anomalies are chance results, apparent over-reaction to information is about as common as underreaction, and post-event continuation of pre-event abnormal returns is about as frequent as post-event reversal. Most important, consistent with the market efficiency prediction that apparent anomalies can be due to methodology, most long-term return anomalies tend to disappear with reasonable changes in technique.’ Fama (1998)
- ‘What, then, can we conclude about market efficiency? Amazingly, there is still no consensus among financial economists. Despite the many advances in the statistical analysis, databases and theoretical models surrounding the efficient markets hypothesis, the main effect has been to harden the resolve of the proponents on each side of the debate.’ Lo (2000a)

3.1.4 Data

The first three experiments below (autocorrelation, long memory and the runs test) use data from the Dow Jones Industrial Average (DJIA). The Dow is the best-known US stock index, and the second-
oldest (after the Dow Jones Transportation Average). The index is a price-weighted average (rather than a market-value weighted or capitalization-weighted index) of 30 large, publicly-owned companies based in the US. The DJIA daily closing prices from 1 October 1928 to 23 March 2012 were downloaded from Yahoo! Finance. Although the Dow represents the average of its constituent stocks, care should be taken when extrapolating the characteristics of a stock index to the characteristics of individual stocks. For example, as pointed out on pp. 24–25, although weekly and monthly stock returns are weakly negatively correlated, daily, weekly and monthly index returns are positively correlated, due to large positive cross-autocorrelations across individual securities across time.

### 3.2 Autocorrelation

A necessary (but not sufficient) condition for the martingale hypothesis to hold is that the time series has zero autocorrelation. Note that due to the definition of autocorrelation, detrending the data is not necessary, the results will be the same. Table 3.2 shows the autocorrelation of DJIA returns. Autocorrelation is small but positive for all time periods. The autocorrelations for daily and weekly returns are the closest to zero, and thus an efficient market.

<table>
<thead>
<tr>
<th>Time interval</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>0.0138</td>
</tr>
<tr>
<td>Weekly</td>
<td>0.0117</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.0793</td>
</tr>
<tr>
<td>Annual</td>
<td>0.1194</td>
</tr>
</tbody>
</table>

**Table 3.2: Autocorrelation of DJIA log returns.**

### 3.3 Long Memory

Another necessary (but not sufficient) condition for the martingale hypothesis to hold is that the time series has no long memory. As defined on p. 25, a random process has long-memory when its autocorrelation function has hyperbolic decay. Hurst’s rescaled range \( R/S \) statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. If \( \{r_1, r_2, \ldots, r_n\} \) is a sample of continuously compounded asset returns and \( \bar{r}_n \) the sample mean \( \frac{1}{n} \sum_j r_j \), then the rescaled-range statistic, \( R/S \), is given by

\[
R/S \equiv \frac{1}{s_n} \left[ \max_{1 \leq k \leq n} \sum_{j=1}^k (r_j - \bar{r}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^k (r_j - \bar{r}_n) \right] 
\]

(3.3.1)

where \( s_n \) is the standard deviation,

\[
s_n \equiv \left[ \frac{1}{n} \sum_j (r_j - \bar{r}_n)^2 \right]^{1/2}.
\]

(3.3.2)

Each constituent makes up a fraction of the index that is proportional to its price.
The Hurst exponent, $H$, is defined by

$$R/S = cn^H$$  \hspace{1cm} (3.3.3)

(where $c$ is a constant) and estimated using the following regression

$$\log R/S = \log(c) + H \log n.$$  \hspace{1cm} (3.3.4)

Intuitively, the first term within the square brackets in (3.3.1) will be large and positive if there are many large successive positive returns, and the second term will be large and negative if there are many large successive negative returns, so $R/S$, and hence $H$, will be large if the returns show persistence. Further, (3.3.1) utilizes the sum of deviations from the mean over a sequence of returns, rather than merely comparing successive returns, so measures long-term persistence.

Two implementations of software for measuring Hurst’s rescaled range ($R/S$) statistic were written, one in Visual Basic for Excel, and one in C++. Both are available online, and the Visual Basic code is given in Appendix C (pp. 151–153). In both cases, the input sequence should be stationary, with mean zero. So if analysing financial data, the input data must be 1) returns (not price) and 2) detrended (zero mean). It should be noted that a given time series has a single value of $H$, but annual, monthly, weekly and daily data will give different approximations of $H$. The spreadsheet version also generates a graph of $\log(R/S)$ over $\log$ (time) and one can also identify cycles in the time series from kinks in the line.

I ran the C++ program on daily, weekly, monthly and annual detrended DJIA returns. In order to make the processing time reasonable, a maximum of 1000 data points were processed at a time. The daily data was processed in 21 batches and the weekly data was processed in 5 batches. In both cases the mean value of $H$ was calculated.

My implementation of $R/S$ analysis calculates $H$ and is extremely accurate, although suffers from long run times (one hour twenty minutes for 1000 data points on a PC with a 1.66 GHz Intel Core Duo Processor T2300 and 2GB of RAM). Table 3.3 shows the results of the analysis on detrended DJIA returns, which appear to show persistence. However, in light of the fact that $R/S$ analysis fails to distinguish between short-range dependence and long-range dependence (Lo, 1991), and the fact that DJIA returns showed positive autocorrelation (Section 3.2), I cannot conclude that there is significant evidence for the existence of long memory in stock returns, so the results are consistent with an efficient market.

<table>
<thead>
<tr>
<th>Time interval</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>0.5645</td>
</tr>
<tr>
<td>Weekly</td>
<td>0.5802</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.5571</td>
</tr>
<tr>
<td>Annual</td>
<td>0.6004</td>
</tr>
</tbody>
</table>

Table 3.3: Rescaled range analysis on detrended DJIA log returns
3.4 Runs Test

The runs test is a non-parametric statistical test that can be used to test for serial dependence and may be considered a test of market efficiency. A ‘run’ within a sequence is a maximal non-empty consecutive subsequence consisting of adjacent equal elements. For example, the sequence ‘+−−−+−+−−−−+’ consists of seven runs.

3.4.1 First Runs Test

Given a sequence of length \( n \) with \( n_+ \) occurrences of ‘+’ and \( n_- \) occurrences of ‘−’ (so \( n = n_+ + n_- \)), if each element in the sequence is independent, then the number of runs is a random variable with an approximately normal distribution, mean \( \mu \) and variance \( \sigma^2 \), where

\[
\mu = \frac{2n_+ n_-}{n} + 1 \tag{3.4.1}
\]

and

\[
\sigma^2 = \frac{2n_+ n_- (2n_+ n_- - n)}{n^2 (n - 1)} = \frac{(\mu - 1)(\mu - 2)}{n - 1}. \tag{3.4.2}
\]

There is no need to detrend the data. A run is a consecutive sequence of DJIA returns above (below) the mean return. The above runs test is performed on daily, weekly, monthly and annual returns, in chronological order. The statistics generated by the runs test are displayed in Table 3.4, and show the actual and expected total number of runs. The statistics clearly show that the null hypothesis of independence is strongly rejected for daily returns, but accepted for weekly, monthly and annual returns. Using the runs test loses information, when compared to, say, autocorrelation, because the magnitude of the returns is lost. However, the concept of an efficient market relates to a martingale, which concerns only the expected return, so the runs test turns out to be a direct measure of market efficiency, which is precisely what we want. The results show that daily returns are the least consistent with an efficient market, whilst annual returns approximate an efficient market. Fama (1965b) performed a similar runs test on the price changes of stocks, but only considered the expected number of runs, so no statistical tests were performed.

<table>
<thead>
<tr>
<th></th>
<th>Actual number</th>
<th>Expected number</th>
<th>Standard deviation</th>
<th>Z-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>10064</td>
<td>10476.4512</td>
<td>72.3478</td>
<td>-5.7010</td>
<td>0.0000****</td>
</tr>
<tr>
<td>Weekly</td>
<td>2140</td>
<td>2164.6095</td>
<td>32.7894</td>
<td>-0.7505</td>
<td>0.2265</td>
</tr>
<tr>
<td>Monthly</td>
<td>487</td>
<td>496.7680</td>
<td>15.6696</td>
<td>-0.6234</td>
<td>0.2665</td>
</tr>
<tr>
<td>Annual</td>
<td>42</td>
<td>41.7711</td>
<td>4.4469</td>
<td>0.0515</td>
<td>0.4795</td>
</tr>
</tbody>
</table>

Table 3.4: The actual and expected total number of runs, where a run is a consecutive sequence of DJIA log returns above (below) the mean return. * indicates statistical significance at the 10% level, ** 5%, *** 1%, **** 0.5% and ***** 0.1%.
3.4.2 Second Runs Test

However, we can go further, and consider the number of increasing runs, and the number of decreasing runs, for runs of length \( i \), and compare this with a random walk. Let \( Y_i \) be the number of increasing (decreasing) runs of length \( i \) in a sequence of \( n \) numbers. Then the expected value for \( Y_i \) runs is given by

\[
E(Y_i) = \frac{2}{(i + 3)!} [n(i^2 + 3i + 1) - (i^3 + 3i^2 - i - 4)] \quad \text{for} \quad i \leq n - 2
\]

and

\[
E(Y_i) = \frac{2}{n!} \quad \text{for} \quad i = n - 1
\]

The algorithm for the standard deviation of the number of runs is better explicated by means of computer code. I have programmed both of the above runs tests in Visual Basic for Excel, the code is given in Appendix [E](pp. 138–145), and the spreadsheet is available online. As before, the above runs test is performed on daily, weekly, monthly and annual DJIA returns, in chronological order. Here, in contrast to the first runs test, we classify subsequences according to whether returns are increasing or decreasing. For example, the sequence -0.2, -0.1, 0, 0.1, 0.2, 0.1, 0, -0.1, -0.2 would be classified as one increasing run of length four and one decreasing run of length four. ∗ indicates statistical significance at the 10% level, ∗∗ 5%, ∗∗∗ 1%, ∗∗∗∗ 0.5% and ∗∗∗∗∗ 0.1%. Again, there is no need to detrend the data. Note that in neither of the two tests is it necessary to assume that the ‘+’s and ‘−’s have equal probabilities, they only assume that the elements are independent and identically distributed. If there are too many or two few runs, the hypothesis of statistical independence of the elements may be rejected. The results of this runs test are given in Appendix [F](pp. 146–146). The tables show, for each run length, the actual and expected number of increasing (decreasing) runs, the z-value, the p-value and the degree of any statistical significance. When compared to a random walk, returns are significantly less likely to increase or decrease for just one day, and far more likely to deteriorate for 2–5 days in a row. Returns are more likely to increase for just one week, or deteriorate for three or more weeks, relative to a random walk. The returns deteriorated for two successive months more frequently than expected. The market returns deteriorated for three successive years more frequently than would be expected from a random walk, and were relatively unlikely to increase for just one year. The only run of increasing annual returns that was over-represented was of length one. The results for annual returns are consistent with a business cycle. Overall, the results show that daily, weekly and decreasing returns are the least consistent with an efficient market.

3.5 Investment Newsletters

‘The Forbes/Hulbert investment letter survey’ ([Hulbert](2002)) was purchased. The data encompasses performance from 31 May 1990 to 31 December 2001 and includes just those newsletters tracked that have a predominant US equity focus. I contacted the editors of the listed newsletters to determine whether each newsletter was based on technical analysis, fundamental analysis or a combination of the two (in
which case I asked to what degree each type of analysis was used). Hulbert (2002) split the nearly dozen-year span into ‘up market’ and ‘down market’ periods, as shown in Table 3.5 (page 65). The data was analysed by performing correlation analysis on the quantitative data. The raw data is proprietary, so is omitted. The results of the analysis of financial newsletters are given in Table 3.6. The results showed a strongly negative correlation between returns in a bull market and returns in a bear market and a strongly negative correlation between risk (standard deviation) and returns in a bear market. Also, technical analysts underperformed the market, and their results were particularly poor during bull markets. Of eight purely fundamental newsletters, two beat the market, of nine purely technical newsletters, none beat the market. In particular, the risk-adjusted performance of the technical newsletters was derisory. If the market had a Sharpe ratio of 1.0, the average purely technical newsletter had a Sharpe ratio of 0.24. It can be concluded that technical analysis—as used by practitioners—fails to outperform the market.

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 May 1990</td>
<td>31 October 1990</td>
<td>down</td>
</tr>
<tr>
<td>1 November 1990</td>
<td>29 June 1990</td>
<td>up</td>
</tr>
<tr>
<td>30 June 1998</td>
<td>31 August 1998</td>
<td>down</td>
</tr>
<tr>
<td>1 September 1998</td>
<td>29 June 1999</td>
<td>up</td>
</tr>
<tr>
<td>30 June 1999</td>
<td>30 September 1999</td>
<td>down</td>
</tr>
<tr>
<td>1 October 1999</td>
<td>30 March 2000</td>
<td>up</td>
</tr>
<tr>
<td>31 March 2000</td>
<td>31 December 2001</td>
<td>down</td>
</tr>
</tbody>
</table>

**Table 3.5:** ‘Up markets’ and ‘down markets’ as defined by Hulbert (2002)

<table>
<thead>
<tr>
<th>Bull return</th>
<th>Bear return</th>
<th>Price of newsletter</th>
<th>SD</th>
<th>Return</th>
<th>Sharpe</th>
<th>% Fundamental</th>
<th>% Technical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bull return</td>
<td>1.00</td>
<td>-0.71</td>
<td>0.16</td>
<td>0.53</td>
<td>0.53</td>
<td>0.51</td>
<td>-0.51</td>
</tr>
<tr>
<td>Bear return</td>
<td>-0.71</td>
<td>1.00</td>
<td>-0.09</td>
<td>-0.71</td>
<td>-0.21</td>
<td>-0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Price of newsletter</td>
<td>0.16</td>
<td>-0.09</td>
<td>1.00</td>
<td>0.28</td>
<td>-0.16</td>
<td>-0.03</td>
<td>0.40</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.53</td>
<td>-0.71</td>
<td>0.28</td>
<td>1.00</td>
<td>-0.19</td>
<td>-0.24</td>
<td>0.13</td>
</tr>
<tr>
<td>Return</td>
<td>0.53</td>
<td>-0.21</td>
<td>-0.16</td>
<td>1.00</td>
<td>-0.19</td>
<td>1.00</td>
<td>0.42</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>0.53</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.24</td>
<td>0.84</td>
<td>1.00</td>
<td>0.53</td>
</tr>
<tr>
<td>% Fundamental</td>
<td>0.51</td>
<td>-0.12</td>
<td>0.40</td>
<td>0.13</td>
<td>0.42</td>
<td>0.53</td>
<td>1.00</td>
</tr>
<tr>
<td>% Technical</td>
<td>-0.51</td>
<td>0.12</td>
<td>-0.40</td>
<td>-0.13</td>
<td>-0.42</td>
<td>-0.53</td>
<td>-1.00</td>
</tr>
</tbody>
</table>

**Table 3.6:** Correlation analysis of ‘The Forbes/Hulbert investment letter survey’. ‘Bull return’ refers to the newsletter returns in ‘up markets’, and ‘bear return’ refers to the newsletter returns in ‘down markets’ defined in Table 3.5 (p. 65). Standard deviation (SD) refers to the standard deviation of returns, and is a proxy for risk. The Sharpe ratio is defined on p. 77.
3.6 Conclusion and Summary

In this chapter it was first explained what is meant by a stylized fact, then stochastic processes, time series and the efficient market hypothesis were covered in sufficient detail to characterize a financial time series. The investigation into autocorrelation and long memory found that detrended DJIA log returns exhibit persistence, but not necessarily long memory, when measured at daily, weekly, monthly and (especially) annual intervals. The runs test uncovered highly significant patterns in DJIA daily returns that are inconsistent with an efficient market. For example, a run of just one decreasing return is relatively unusual. This means that if returns improve on day one, then deteriorate the following day, they are more likely to deteriorate on the third day, than improve. Considering annual returns, relative to a random walk, the most common run of improved returns is one, and the most common run of deteriorating returns is three, totalling four years, which is consistent with a business cycle. The results of the analysis of investment newsletters were consistent with weak-form efficiency. The tests of autocorrelation and long memory show annual returns to be the least consistent with a martingale, which makes sense, as markets may be less efficient in the longer term because in practice investors have finite time horizons. In contrast, the runs tests showed the daily returns to be the least consistent with a martingale. Autocorrelation is sensitive only to a linear relationships, not nonlinear relationships, whilst the runs test has no such restriction. There is ample empirical evidence that a non-linear process contributes to the dynamics of market returns [Hsieh, 1989, Scheinkman and LeBaron, 1989, Brock et al., 1991]. This gives support for the efficacy of technical analysis, which relies on non-linearities being present [Neftci, 1991]. In their review paper, Park and Irwin (2004) found that, on average, non-linear methods outperformed genetic programming in all three types of market considered: stock markets, futures markets and currency markets. In sum, this chapter reconciles the apparent efficiency of markets according to linear measures, and the potential for nonlinear methods to make above-average risk-adjusted returns. The results of the investment newsletter analysis implied that technical analysis, as applied by the newsletter writers, holds no value. This is not surprising, as most such practitioners take a naive discretionary approach to technical analysis.
Chapter 4

Modelling

The focus of this chapter—the second part of the time series trilogy—is modelling. The primary aim is to build an agent-based artificial stock market and explore the effect of the ratio of fundamental analysts to technical analysts, and whether and when the resultant time series displays the statistical properties exhibited by a real market, i.e. reproduces the stylized facts described in the previous chapter. I am not suggesting that finding the right ratio alone would allow the artificial market to generate time series that exhibit stylized facts in the market, my model is not that flexible. For example, in the artificial stock market created by Martinez-Jaramillo and Tsang (2009), in addition to technical and fundamental traders, they employed noise traders, in addition to a risky asset their traders could hold cash, and in addition to market orders, their technical traders could place limit orders, plus their technical traders were equipped with up to 12 different technical and momentum indicators, and were able to learn. Thus they created a more flexible model with ‘a substantial number of parameters’ which was better placed to exhibit the stylized facts. Instead, I have focussed on simplicity and realism. This experiment is published as Sewell (2012a). Whilst Sewell (2011g) describes the evolution of the heuristics and biases used in the artificial stock market, work that was removed from the thesis to save space. The second experiment models investors’ risk preferences and develops a novel investment performance measurement metric, cumulative prospect theory certainty equivalent (CPTCE). [Sewell (2009a)] relates to this experiment.

4.1 An Artificial Stock Market

4.1.1 Design

Multiagent Systems

A multiagent system is a system in which several interacting, autonomous, intelligent agents pursue some set of goals or perform some set of tasks. A literature review was given on p. 40. Let’s consider some valid criticisms of the approach. Agent-based modelling can stand accused of being poor science. To do science, one needs ways to test hypotheses and reach general conclusions. Some of the problems with multiagent systems:

- Too many free parameters.
- In common with all empirical research, one can always find evidence to support what one seeks to
4.1. An Artificial Stock Market

prove. Too many possible explanations of the results leads to the opportunity for story telling.

- No general theoretical way to know whether a given simulation configuration is the only way to get from some set of initial conditions to a result or one of a family of hundreds or millions of ways to get to a result.

- Model validation can be complicated.

- Difficult to verify that the models are consistent enough to be useful.

Daniel Kahneman shared the Nobel Prize in Economics in 2002 with Vernon Smith. Economists once thought of their science as inherently non-experimental, but Smith pioneered laboratory experimental economics, and spearheaded ‘wind tunnel tests’, where trials of new markets could be tried out in the lab before being implemented in the real world, giving policy makers a better understanding of how a new market is likely to work in practice. Going one step further, from the laboratory to the computer, on balance I consider agent-based modelling to be an effective way of studying behavioural finance, because empirical results derived from the laboratory can be aggregated and modelled flexibly and at low cost.

4.1.2 Implementation

The literature on behavioural finance was reviewed in Section 2.2 (page 33). Consider some common heuristics and biases. *Availability* ([Tversky and Kahneman, 1973](#)) is a cognitive heuristic in which a decision maker relies upon knowledge that is readily available rather than examine other alternatives or procedures. *Representativeness* ([Tversky and Kahneman, 1974](#)) leads people to predict future events by looking for familiar patterns and taking a short history of data and assuming that future patterns will resemble past ones. The *status quo bias* ([Samuelson and Zeckhauser, 1988](#)) is a cognitive bias for the status quo; in other words, people tend to be biased towards doing nothing or maintaining their current or previous decision. The status quo bias can lead to another cognitive heuristic, known as *anchoring* ([Tversky and Kahneman, 1974](#)), which describes the common human tendency to make decisions based on an initial ‘anchor’. We prefer relative thinking to absolute thinking. Other observed consistent but irrational behaviour includes *overconfidence, optimism* and *herding*.

From my work on the evolutionary foundations of heuristics and biases ([Sewell, 2011g](#)), I identified the following heuristics and biases in the modern day investor.

- Overconfidence is likely to lead investors to trade too much, generally preferring actively managed funds. Excess overconfidence among males in particular explains the popularity of trading among men.

- Optimism naturally creates a ‘bullish’ tendency and can create asymmetry in the behaviour of markets.

- Availability could, for example, cause us to purchase shares in a company simply because it comes to mind more readily.
• Herding can lead investors to focus only on a subset of securities, whilst neglecting other securities with near identical exogenous characteristics.

• Representativeness leads analysts to believe that observed trends are likely to continue. Representativeness causes trend following by technical analysts and overreaction among fundamental analysts.

• Anchoring is likely to cause fundamental analysts to underreact, for example to earnings announcements.

Overconfidence leads to excess trading and helps create a liquid market in the first place, optimism likely increases market participation in general, whilst availability and herding will generally only effect a subset of stocks so their impact would be diluted when aggregated across stocks in general. So I only implement the final two heuristics/biases above, which are the most relevant regarding market impact. In summary, following Barberis et al. (1998) we expect underreaction to news but an overreaction to a series of good or bad news from fundamental analysts, and trend following from technical analysts. We do not have sufficient news data to test this hypothesis directly, but would expect it to generate kurtosis and non-linearities in market data, which are indeed found in real markets (Cont, 2001).

Theoretical Model of Market Action

Introduction

First, a reminder of three definitions.

Fundamental analysis A method of forecasting markets through the analysis of relevant news.

Technical analysis A method of forecasting markets through the analysis of data generated from the activity of trading itself. This was covered in detail on pp. 29–33.

Multiagent system A system in which several interacting, autonomous, intelligent agents pursue some set of goals or perform some set of tasks. See p. 40.

The objective is to model a stock market using a multiagent system. The implementation uses Microsoft Excel. The main criteria is to be as realistic as possible; that is, the problem domain is mapped onto the model. The only other criteria is to keep the model as simple as possible (which is often at odds with the quest for realism). In practice, traders are essentially divided into two groups, fundamental analysts (who tend to be longer term) and technical analysts (who tend to be shorter term); the distribution of agents in our model shall mirror this dichotomy (Lux, 1995; Hong and Stein, 1999) took a similar approach. Reviewing the existing literature, at one extreme, some artificial markets employ agents with zero intelligence (Gode and Sunder, 1993; Farmer et al., 2005). Whilst in some implementations agents are able to swap between technical analysis and fundamental analysis depending on their profits (they have the ability to learn) (Lux, 1998; Lux and Marchesi, 1999, 2000). I reject the application of zero intelligence agents, as in practice most traders have a reasonably consistent strategy (which may or may not work). I also reject the idea of agents swapping between technical analysis
and fundamental analysis, because in practice technical analysts and fundamental analysts tend to be somewhat antagonistic towards each other. Finally, I reject the notion of agents learning. Due to a combination of overconfidence, a limited exposure to markets (at most one working life) and noise, real traders do not learn how to predict market\(^{1}\) (even if they did, as new traders replaced the old, they would not improve ‘on average’); this stasis is trivially mirrored. Indeed, Martinez-Jaramillo (2007) developed an artificial financial market and investigated the effects on the market when the agents learn, and, on average, their model without learning replicated the stylized facts most accurately (though not by much). In my model the technical analysts simply follow the technician’s number one rule: they follow the trend, so the model fails to replicate some of the more complex strategies that chartists follow. The artificial market operates such that each time step represents one trading day, and the stock price may be interpreted as a daily closing price.

Below is a taxonomy of five groups of market participants.

<table>
<thead>
<tr>
<th><strong>FUNDAMENTAL ANALYSTS</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Poor</strong></td>
<td>Trade randomly—fundamental analysts lacking sufficient skills or experience to analyse a company will make mistakes at random.</td>
</tr>
<tr>
<td><strong>Real</strong></td>
<td>Consistent, correlated and irrational—<em>Homo sapiens</em> employed as fundamental analysts will be susceptible to behavioural biases and make systematic errors.</td>
</tr>
<tr>
<td><strong>Good</strong></td>
<td>Rational—Skilled fundamental analysts (<em>Homo economicus</em>) with the ability to accurately analyse a company, and thus evaluate the value of its stock.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TECHNICAL ANALYSTS</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Poor</strong></td>
<td>Trade randomly—those employed as technical analysts but lacking the ability or experience to follow the rules of technical analysis.</td>
</tr>
<tr>
<td><strong>Good</strong></td>
<td>Consistent, correlated and irrational—experienced technical analysts able to trade in accordance with the rules of technical analysis.</td>
</tr>
</tbody>
</table>

Assuming that all five types of market participant exist (they do), with imperfect arbitrage opportunities and no 100 per cent rational traders, the resultant effect on the market is the aggregate effect of real fundamental analysts trading against good technical analysts. A multiagent system with technical and fundamental agents is used to model price action. This work employs a bottom-up approach and has been developed from first principles.

---

1. There are two forms of analysis and the practitioners of each tend to be somewhat antagonistic. Fundamental analysts have referred to Technical analysts as indulging in voodoo and shamanism and a technician once described the former’s efforts as “fundamentally a waste of time” (Society of Technical Analysts, 1999, p. 2).
2. Indeed, there is a negative relationship between the tenure of a hedge fund manager and hedge fund returns (Boyson, 2003).
3. Technical analysis is a behavioural bias (representativeness), here a ‘good’ technical analyst is one who accurately and consistently trades according to the rules of technical analysis.
Fundamental Analysis

News, by definition, is unpredictable (otherwise it would have been reported yesterday), so let us assume that the cumulative impact of relevant news on a stock follows a geometric random walk. Fundamental analysts calculate the intrinsic value of a stock by the analysis of relevant news. Let the exogenous variable \( V_t \) be the perceived fundamental value at time \( t \), where \( \log V \) follows a random walk. Note that \( V \) is not directly observable, but changes in the variable are observable in the form of news, and the model assumes that \( V \) may be calculated. If \( V \) increases, this corresponds to good news, if it decreases, this corresponds to bad news. The fundamental analysts trade on the basis of this perceived fundamental value alone (they do not consider historical prices). At each time step, if the price of a stock is below (above) the perceived fundamental value of the stock, fundamental analysts will take a long (short) position in proportion to the logarithm of the perceived fundamental value over the price. In other words, the fundamental analysts trade in such a way that they always move the price towards the fundamental value. Formally,

\[
\log \frac{V_t}{V_{t-1}} > 0 \text{ represents good news, and } \\
\log \frac{V_t}{V_{t-1}} < 0 \text{ represents bad news.}
\]

Let \( n_f \) be the proportion of the total number of trades made by fundamental analysts and \( P_t \) the price at time \( t \). The idea is to model an underreaction to news, but an overreaction to a series of good or bad news. Therefore, the fundamental agents overreact to three or more successive good (or bad) news items, are neutral towards exactly two successive good (or bad) news items and underreact otherwise. In a market populated entirely by fundamental analysts, the log return of the price between time \( t \) and time \( t + 1 \) would be \( F_t \). The values for the reaction variable, \( r \), below, are chosen with reference to Theobald and Yallup’s direct measures of the degrees of overreaction and underreactions in financial markets, but the figures used here are subject to significant uncertainty.

\[
F_t = r \log \frac{V_t}{P_t} 
\]  

(4.1.1)

where

\[
r = \begin{cases} 
1.1 & \text{if } V_t > V_{t-1} > V_{t-2} > V_{t-3} \text{ or } V_t < V_{t-1} < V_{t-2} < V_{t-3}; \\
1 & \text{if } V_t > V_{t-1} > V_{t-2} \text{ or } V_t < V_{t-1} < V_{t-2}; \\
0.9 & \text{else}
\end{cases}
\]

(4.1.2)

Technical Analysis

The technician’s number one rule is that they follow the trend. Quoting a best-selling practitioner’s book on technical analysis (Murphy [1999] p. 49), ‘The concept of trend is absolutely essential to the technical approach to market analysis. All of the tools used by the chartist—support and resistance levels, price
patterns, moving averages, trendlines, etc.—have the sole purpose of helping to measure the trend of the market for the purpose of participating in the trend. We often hear such familiar expressions as “always trade in the direction of the trend,” never buck the trend,” or “the trend is your friend.” So, in this model, technical analysts follow the trend, i.e. display momentum; they consider the historical price of a stock, and nothing else. At each time step, they exhibit persistence by trading in such a way that the price is biased towards continuing in the same direction as the recent past. Let \( n_t \) be the proportion of trades made by technical analysts. The technical analysts’ trend-following strategy looks back three days and weights the price changes by recency. In this model if the market were populated entirely by technical analysts, the log return of the price between time \( t \) and time \( t + 1 \) would be \( T_t \).

\[
T_t = c^3 \log \frac{P_t - 2}{P_t - 3} + c^2 \log \frac{P_t - 1}{P_t - 2} + c \log \frac{P_t}{P_t - 1},
\]

(4.1.3)

where the coefficients \( c^3, c^2 \) and \( c \) form an increasing geometric sequence so that more recent price changes have a greater impact on \( T_t \), and sum to one. Solving \( c^3 + c^2 + c = 1 \), which has one real root, gives us \( c = 0.544 \).

**Stock Price Returns**

Changes in price are determined by the following equation:

\[
\log \frac{P_{t+1}}{P_t} = n_t F_t + n_t T_t.
\]

(4.1.4)

By way of example, if \( P_t > V_t \), the fundamental analyst believes that the stock is overvalued. Those who hold the stock may sell it, those who don’t may either do nothing or short the stock. Or the fundamental analyst may publish a recommendation that the stock is a sell. The point is that on aggregate the actions of the fundamental analysts will put pressure on the stock price to fall. If, however, the technical analysts put even greater selling pressure on the stock, the fundamental analysts will become net buyers.

Taylor (2005) includes various statistics on stocks, repeated in Table 4.1 (p. 73). In order to determine the mean and standard deviation of the Gaussian random variable \( \log \frac{V_t}{V_{t-1}} \), first, a realistic ratio of 50% fundamental trades and 50% technical trades (\( n_f = 0.5 \) and \( n_t = 0.5 \)) was chosen. Then the mean and standard deviation space was discretised, an exhaustive enumeration of return sequences generated, one for each discrete parameter setting pair, and the pair for which the mean and standard deviation of the simulated stock returns most closely matched those of the empirical data in Table 4.1 was chosen. This resulted in a mean of 0.0013 and a standard deviation of 0.023 for the Gaussian random variable \( \log \frac{V_t}{V_{t-1}} \). The model was run over 50,000 days twenty times, and averages of various statistics calculated.

### 4.1.3 Testing

Recall that \( P_t \) is the price of a stock at time \( t \), and \( V_t \) is the perceived fundamental value of the stock at time \( t \). Note that Shiller (1981) calculated that stock market volatility is five to thirteen times too high to be attributed to new information, so we should not expect the standard deviation of \( P \) log returns to equal the standard deviation of \( V \) log returns (although perhaps surprisingly, in this model, the latter is slightly greater). Table 4.2 (p. 73) lists various statistics of the returns generated by the model as the proportion of technical analysts to fundamental analysts varies. Figure 4.1 (page 74) shows the mean return per analyst,
4.1. An Artificial Stock Market

<table>
<thead>
<tr>
<th></th>
<th>Coca Cola</th>
<th>General Electric</th>
<th>General Motors</th>
<th>Glaxo</th>
<th>Marks &amp; Spencer</th>
<th>Shell</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.001167</td>
<td>0.000742</td>
<td>0.000558</td>
<td>0.001473</td>
<td>0.000723</td>
<td>0.000763</td>
<td>0.000905</td>
<td>0.000344</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.0169</td>
<td>0.0151</td>
<td>0.0176</td>
<td>0.0179</td>
<td>0.0166</td>
<td>0.0130</td>
<td>0.0162</td>
<td>0.0018</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.08</td>
<td>0.03</td>
<td>0.13</td>
<td>0.33</td>
<td>0.03</td>
<td>0.23</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.68</td>
<td>5.43</td>
<td>4.56</td>
<td>6.93</td>
<td>4.40</td>
<td>5.18</td>
<td>5.36</td>
<td>0.91</td>
</tr>
<tr>
<td>Returns autocorrelation</td>
<td>-0.035</td>
<td>-0.023</td>
<td>-0.003</td>
<td>0.08</td>
<td>0.034</td>
<td>0.045</td>
<td>0.016</td>
<td>0.044</td>
</tr>
<tr>
<td>Absolute returns autocorrelation</td>
<td>0.329</td>
<td>0.224</td>
<td>0.204</td>
<td>0.247</td>
<td>0.155</td>
<td>0.196</td>
<td>0.226</td>
<td>0.059</td>
</tr>
<tr>
<td>Squared returns autocorrelation</td>
<td>0.545</td>
<td>0.303</td>
<td>0.398</td>
<td>0.414</td>
<td>0.288</td>
<td>0.293</td>
<td>0.374</td>
<td>0.100</td>
</tr>
</tbody>
</table>

**Table 4.1:** Statistics of daily stock log returns (Taylor, 2005)

as the proportion technical analysts/fundamental analysts varies. Figure 4.2 shows the mean, standard deviation and skewness of market log returns as the proportion technical analysts/fundamental analysts varies. Figure 4.3 shows the kurtosis of market log returns as the proportion technical analysts/fundamental analysts varies. Figure 4.4 shows the autocorrelations of returns, absolute returns and squared returns as the proportion technical analysts/fundamental analysts varies. Table 4.3 shows that with a realistic proportion of technical and fundamental trades, the artificial stock market replicates mean returns, the standard deviation of returns, the absolute returns correlation and the squared returns correlation of a real stock market. However, the artificial stock market failed to accurately replicate the skewness, kurtosis and autocorrelation of returns.

<table>
<thead>
<tr>
<th>Fundamental analysts (%)</th>
<th>100</th>
<th>90</th>
<th>80</th>
<th>70</th>
<th>60</th>
<th>50</th>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean fundamental analyst return</td>
<td>0.0011</td>
<td>0.0012</td>
<td>0.0015</td>
<td>0.0017</td>
<td>0.0000</td>
<td>-0.0026</td>
<td>-0.0033</td>
<td>-0.0053</td>
<td>-0.0103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean technical analyst return</td>
<td>-0.0107</td>
<td>-0.0049</td>
<td>-0.0034</td>
<td>-0.0026</td>
<td>0.0000</td>
<td>0.0017</td>
<td>0.0014</td>
<td>0.0013</td>
<td>0.0011</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td>Mean return</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td>Returns standard deviation</td>
<td>0.0226</td>
<td>0.0208</td>
<td>0.0194</td>
<td>0.0182</td>
<td>0.0172</td>
<td>0.0163</td>
<td>0.0155</td>
<td>0.0149</td>
<td>0.0143</td>
<td>0.0138</td>
<td></td>
</tr>
<tr>
<td>Returns skewness</td>
<td>-0.0552</td>
<td>-0.0533</td>
<td>-0.0503</td>
<td>-0.0434</td>
<td>-0.0348</td>
<td>-0.0393</td>
<td>-0.0201</td>
<td>-0.0136</td>
<td>-0.0178</td>
<td>-0.0043</td>
<td>0.0025</td>
</tr>
<tr>
<td>Returns kurtosis</td>
<td>0.0822</td>
<td>0.1512</td>
<td>0.2010</td>
<td>0.2350</td>
<td>0.2476</td>
<td>0.2371</td>
<td>0.2073</td>
<td>0.1348</td>
<td>0.1100</td>
<td>0.0394</td>
<td>-1.4268</td>
</tr>
<tr>
<td>Returns autocorrelation</td>
<td>0.0658</td>
<td>0.2038</td>
<td>0.3338</td>
<td>0.4566</td>
<td>0.5690</td>
<td>0.6710</td>
<td>0.7627</td>
<td>0.8423</td>
<td>0.9088</td>
<td>0.9617</td>
<td>1.0000</td>
</tr>
<tr>
<td>Absolute returns autocorrelation</td>
<td>0.0093</td>
<td>0.0364</td>
<td>0.0931</td>
<td>0.1750</td>
<td>0.2803</td>
<td>0.4045</td>
<td>0.5403</td>
<td>0.6730</td>
<td>0.7984</td>
<td>0.9083</td>
<td>1.0000</td>
</tr>
<tr>
<td>Squared returns autocorrelation</td>
<td>0.0088</td>
<td>0.0401</td>
<td>0.1029</td>
<td>0.1899</td>
<td>0.3021</td>
<td>0.4259</td>
<td>0.5650</td>
<td>0.6974</td>
<td>0.8226</td>
<td>0.9244</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**Table 4.2:** Statistics generated by the artificial stock market

Conclusion

Results showed that whether a fundamental analyst, or a technical analyst, it pays to be among the majority, ideally of about 60 per cent, whilst being in a small minority is the least profitable position to be in. Mean stock returns are low and positive regardless of the relative proportions of analysts, this is consistent with a real market.

As the number of technical analysts increases, the standard deviation of returns decreases, whilst remaining realistic, whilst the skewness increases. The model exhibited slight negative skewness, whilst real markets exhibit significant positive skewness. Whilst the kurtosis of market returns peaks at around 0.25 with around 40 per cent technical analysts, and rapidly declines as the number of technical analysts
4.1. An Artificial Stock Market

Figure 4.1: Mean log return (P&L) per analyst

Figure 4.2: Statistics of price log returns
4.1. An Artificial Stock Market

Figure 4.3: Kurtosis of price log returns

Figure 4.4: Autocorrelations of price log returns
4.2. Investment Performance Measurement

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Proportion of technical analysts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean return</td>
<td>0–100%</td>
</tr>
<tr>
<td>Returns standard deviation</td>
<td>40–70%</td>
</tr>
<tr>
<td>Returns skewness</td>
<td>none</td>
</tr>
<tr>
<td>Returns kurtosis</td>
<td>none</td>
</tr>
<tr>
<td>Returns autocorrelation</td>
<td>none</td>
</tr>
<tr>
<td>Absolute returns autocorrelation</td>
<td>30–40%</td>
</tr>
<tr>
<td>Squared returns autocorrelation</td>
<td>40–50%</td>
</tr>
</tbody>
</table>

Table 4.3: Range of proportions of technical analysts in the artificial stock market that replicate stylized facts

exceeds 90 per cent. In contrast, the kurtosis of daily stock returns in real markets is around 5.

The autocorrelation of returns is close to zero with 100 per cent fundamental analysts, and approaches one as the proportion of technical analysts approaches 100 per cent. Unsurprisingly, the trend-following technical analysts created positive autocorrelations in returns in the model, but autocorrelations of returns are close to zero in real markets. The autocorrelation of absolute and squared returns is realistic only around the region of 30%–50% technical analysts. How has the model fared in light of the criticisms of multiagent systems that were highlighted in Section 4.1.1 (p. 67)? The main concern, that one can vary any free parameter until one obtains the result that one desires, i.e. high kurtosis, was mitigated by keeping the number of varying parameters to a minimum, by using realistic assumptions. Martinez-Jaramillo (2007); Martinez-Jaramillo and Tsang (2009) investigated the different conditions under which the statistical properties of an artificial stock market resemble those of a real financial market. Their approach replicated the stylized facts of a financial market far more accurately than my own; this was possible by including and adjusting a much larger number of parameters.

4.2 Investment Performance Measurement

4.2.1 Design

A brief literature review on performance measurement was given in Section 2.2 (p. 40). In this section, the Sharpe ratio, Sortino ratio and Omega are improved upon by developing a new performance metric, cumulative prospect theory certainty equivalent (CPTCE).

A 4 per cent return on a savings account will always be preferable to a 3 per cent return. The choice is not as clear cut when there is an element of chance. If you are offered a gamble, what would be a fair value for you to pay (or be paid) for the opportunity to take it? Consider a 50 per cent chance of losing £100 and a 50 per cent chance of winning £100. The expected return is £0. But what about risk? There is no principled way of measuring risk.

Kelly (1956) and Breiman (1961) showed that in order to achieve maximum growth of wealth, one should maximize the expected value of the logarithm of wealth after each period. However, most investors are unwilling to endure the volatility of wealth that such a strategy entails, and as John Maynard
Keynes reminded us, in the long run, we’re all dead. For this reason, various risk-adjusted performance metrics have been developed. Any risk-adjusted measure of performance makes assumptions about investors’ risk preferences. It is my contention that one could measure risk-adjusted performance in terms of the way people actually behave. Empirical research tells us that, in practice, people care about losses and gains rather than absolute wealth, evaluate probabilities incorrectly, are loss averse, risk averse for gains, risk seeking for losses and have non-linear preferences.

The Sharpe ratio is the most popular performance metric. Where $R$ is the asset return, $R_f$ is the return on a benchmark asset, such as the risk free rate of return, $E[R - R_f]$ is the expected value of the excess of the asset return over the benchmark return, and $\sigma = \sqrt{\text{Var}(R - R_f)}$ is the standard deviation of the excess return,

$$\text{Sharpe ratio} = \frac{E[R - R_f]}{\sigma}.$$  \hspace{1cm} (4.2.1)

The Sharpe ratio makes implicit assumptions which stem from the capital asset pricing model (CAPM) \cite{Treynor1962, Sharpe1964, Lintner1965, Mossin1966} it assumes either 1) normally distributed returns or 2) mean-variance preferences.

Both assumptions are suspect:

1. The returns generated by most hedge funds exhibit negative skewness \cite{Kat2002}.
2. In addition to the mean and variance, people also care about skewness (they like it positive) and kurtosis (they don’t like it), and higher moments matter too \cite{Scott1980}.

Because the Sharpe ratio is oblivious of all moments higher than the variance, it is prone to manipulation by strategies that can change the shape of the probability distribution of returns. Mathematically, maximizing the Sharpe ratio is a standard quadratic programming optimization problem with the constraint that the mean excess return is fixed. \cite{Goetzmann2002} proved that the solution produces a reversed lognormal distribution with a truncated right tail and a fat left tail leading to extreme negative skewness, as shown in Figure 4.5. The optimal strategy involves selling out-of-the-money calls (to remove the right tail of the distribution) and selling out-of-the-money puts (to enhance the left tail) in an uneven ratio. Such a strategy would generate a regular return from writing options, but would have a large exposure to extreme negative events. In other words, a manager with no special information can improve his Sharpe ratio in such a way that the distribution of returns exhibits negative skewness. As mentioned above, most investors prefer positive skewness, therefore, although a high Sharpe ratio is good thing, a high Sharpe ratio strategy is a bad thing.

### 4.2.2 Implementation

Cumulative prospect theory certainty equivalent

When presented with an uncertain payoff, the *certainty equivalent* is the guaranteed payoff at which a person is indifferent between accepting the uncertain payoff and the guaranteed payoff. Certainty
equivalent varies according to individuals’ risk preferences, and for a risk averse individual the certainty equivalent will be less than the expected value of the gamble. An attractive property of certainty equivalent is that so long as risk preferences are known, it reduces a probability distribution to a single value. This has obvious advantages for an investor who wishes to compare distributions of returns.

A new investment performance measurement algorithm is developed, which is an implementation of Tversky and Kahneman’s cumulative prospect theory (Tversky and Kahneman, 1992) (explained on p. 38). The measure is known as Cumulative prospect theory certainty equivalent (or CPTCE). This measure tells us that, on average, people would wish to be paid £22.30 to take the gamble offered on p. 76. The equations used to derive this figure follow.

Wakker (2010)’s step-by-step description of the procedure for calculating the PT (prospect theory) value of a prospect follows. Note that steps 1 and 2 together determine the complete sign-ranking, and losses (steps 6–8) are treated symmetrically to gains (steps 3–5).

1. Completely rank outcomes from best to worst.
2. Determine which outcomes are positive and which are negative.
3. For each positive outcome, calculate the gain-rank $g$.
4. For all resulting gain-ranks, calculate their $w^+$ value.
5. For each positive outcome $a$, calculate the marginal $w^+$ contribution of its outcome probability $p$ to its rank; i.e., calculate $w^+(p + g) - w^+(g)$.
6. For each negative outcome, calculate the loss-rank $\ell$.
7. For all resulting loss-ranks, calculate their $w^-$ value.
8. For each negative outcome $b$, calculate the marginal $w^-$ contribution of its probability $q$ to its loss-rank; i.e., calculate $w^-(q + \ell) - w^-(\ell)$.

$p + g$ is the gain-rank of the gain in the prospect ranked worse than but next to $a$ considered in step 3.
$q + \ell$ is the loss-rank of the loss in the prospect ranked better than but next to $b$ considered in step 6.

Figure 4.5: Maximal Sharpe ratio (Goetzmann et al., 2002)
9. Determine the utility of each outcome, \( U(x) \).

10. Multiply the utility of each outcome by its decision weight.

11. PT value is the sum of the results of step 10.

12. Certainty equivalent is then a function of PT value, \( \alpha \), \( \beta \) and \( \lambda \), as described in the code in Appendix I (pp. 156–165).

The two weighting functions, \( w^+ \) for gain-ranked probabilities and \( w^- \) for loss-ranked probabilities:

\[
\begin{align*}
  w^+(p) &= \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^\frac{1}{\gamma}} \\
  w^-(p) &= \frac{p^\delta}{(p^\delta + (1-p)^\delta)^\frac{1}{\delta}}
\end{align*}
\]

(4.2.2) (4.2.3)

\( \gamma \) and \( \delta \) are parameters that Tversky and Kahneman determined empirically as \( \gamma = 0.61 \) and \( \delta = 0.69 \).

The value (utility) function (taken from Kobberling (2002)) has a loss aversion parameter \( \lambda \), and is as follows:

\[
U(x) = \begin{cases} 
  f(x) & \text{if } x > 0 \\
  0 & \text{if } x = 0 \\
  \lambda g(x) & \text{if } x < 0
\end{cases}
\]

(4.2.4)

where \( f(x) \) and \( g(x) \) are defined as follows:

\[
\begin{align*}
  f(x) &= \begin{cases} 
    x^\alpha & \text{if } \alpha > 0 \\
    \log(x) & \text{if } \alpha = 0 \\
    1 - (1 + x)^\alpha & \text{if } \alpha < 0
  \end{cases} \\
  g(x) &= \begin{cases} 
    -(-x)^\beta & \text{if } \beta > 0 \\
    -\log(-x) & \text{if } \beta = 0 \\
    (1 - x)^\beta - 1 & \text{if } \beta < 0
  \end{cases}
\end{align*}
\]

(4.2.5) (4.2.6)

Again, the parameters \( \alpha \), \( \beta \) and \( \lambda \) were determined by Tversky and Kahneman empirically, \( \alpha = 0.88 \), \( \beta = 0.88 \) and \( \lambda = 2.25 \).

Cumulative prospect theory certainty equivalent makes up part of a more general performance measurement calculator which I wrote in PHP for the Web and in Visual Basic for Excel, both of which are freely available online [7]. It calculates mean return, standard deviation, skewness, kurtosis, beta, Jensen’s alpha, Sharpe ratio, Sortino ratio, Treynor’s measure, information ratio, Stutzer ratio, Omega, \( M^2 \), \( T^2 \) and maximum drawdown. To avoid ambiguity, the source code for CPTCE is included in Appendix I (pp. 156–165). Note that Kahneman and Tversky’s prospect theory is concerned with absolute gains and losses, whilst here the concept is mapped onto returns (because, in terms of assessing the performance of an investment, that is what an investor is interested in).
4.2.3 Testing

A Monte Carlo simulation was used to simulate 20,000 funds, each with 15 daily returns. Each fund allocated a randomly-chosen proportion (between 0% and 100%) of their assets to a risky asset, and put the rest in a risk-free asset. The risk-free rate and the threshold used for the Sortino ratio, Omega and the upside potential ratio are both set to 3% (0.012% per day), the mean return for the risky asset 0.0905% per day and the risky asset standard deviation 1.62% per day (averages taken from Taylor (2005) and repeated in Table 4.1 (p. 73)). The probability used by Conditional VaR and Modified VaR was set to 0.05.

Results are given in Table 4.4. The table shows the correlations between the Sharpe ratio and CPTCE versus various statistics and performance metrics. The statistics show that CPTCE is more risk averse than the Sharpe ratio, penalizing both the proportion of funds allocated to a risky asset and the standard deviation of returns to a greater degree. Significantly, the Sharpe ratio is indifferent towards skewness, but CPTCE rewards positive skewness, and the latter is more consistent with investors’ risk preferences (Scott and Horvath [1980]). Further, CPTCE punishes maximum drawdown to a greater extent than the Sharpe ratio, which is also more consistent with many investors’ utility. This experiment was conducted using a Performance Metric Analysis Excel spreadsheet I wrote in Visual Basic, which is freely available online[^1]. A reasonable criticism of CPTCE is that prospect theory is *descriptive*, and

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sharpe ratio</th>
<th>CPTCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion in risky asset</td>
<td>0.01</td>
<td>-0.48</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>0.84</td>
<td>0.76</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.00</td>
<td>-0.53</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>1.00</td>
<td>0.70</td>
</tr>
<tr>
<td>Sortino ratio</td>
<td>0.80</td>
<td>0.61</td>
</tr>
<tr>
<td>Omega</td>
<td>0.65</td>
<td>0.50</td>
</tr>
<tr>
<td>CPTCE</td>
<td>0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>Maximum drawdown</td>
<td>-0.49</td>
<td>-0.87</td>
</tr>
<tr>
<td>Upside potential ratio</td>
<td>0.75</td>
<td>0.59</td>
</tr>
<tr>
<td>Calmar ratio</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Conditional VaR</td>
<td>0.18</td>
<td>0.43</td>
</tr>
<tr>
<td>Modified VaR</td>
<td>0.14</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 4.4: *The correlations between the Sharpe ratio and various statistics, and CPTCE and various statistics*

one could argue that an investment manager is responsible for implementing an algorithm that employs sensible *prescriptive* risk preferences, as I argue in Sewell (2009a).

[^1]: [http://www.performance-measurement.org](http://www.performance-measurement.org)
4.3 Conclusion and Summary

Those heuristics and biases which contribute to behavioural finance were identified, and used to build a theoretical model of market action which simulates the aggregates of many interacting agents. The artificial stock market exposed the effect of varying the proportion of technical analysts to fundamental analysts. The artificial market replicates mean returns, the standard deviation of returns, the absolute returns correlation and the squared returns correlation of a real stock market, but failed to accurately replicate the skewness, kurtosis and autocorrelation of returns. This implies that the model has failed to capture some of the dynamics underlying the process of price formation. Finally, a contribution was made to investment performance measurement in the form of a new metric, CPTCE, which is based on prospect theory. From a descriptive perspective, the risk metric should be superior to any existing methods of performance measurement, as it accurately incorporates people’s risk profiles.
Chapter 5

Forecasting

This is the chapter that traders, hedge fund managers and fortune tellers are likely to turn to first. If the author had indulged similarly without first covering the necessary groundwork in the previous chapters, he would be at a distinct disadvantage because domain knowledge is necessary to provide the assumptions that supervised machine learning relies upon. The chapter is published as Sewell and Shawe-Taylor (2012). In addition, Yan et al. (2008) provides a head-to-head evaluation of GP and SVM forecasting, similar to the work in this chapter, and draws the same conclusion.

5.1 Design

VC Dimension

The VC dimension (for ‘Vapnik Chervonenkis dimension’) measures the capacity of a hypothesis space, and the concept is central to support vector machines, a learning algorithm used in this chapter. Capacity is a measure of complexity and measures the expressive power, richness or flexibility of a set of functions by assessing how wiggly its members can be. The definitions below are taken from Vapnik (1999).

Definition 3 The VC Dimension of a Set of Indicator Functions (Vapnik and Chervonenkis, 1968; Vapnik and Chervonenkis, 1971)

The VC dimension of a set of indicator functions \( Q(z, \alpha), \alpha \in \Lambda \), is the maximum number \( h \) of vectors \( z_1, \ldots, z_h \) that can be separated into two classes in all \( 2^h \) possible ways using functions of the set \( Q \) (i.e., the maximum number of vectors that can be shattered by the set of functions). If for any \( n \) there exists a set of \( n \) vectors that can be shattered by the set \( Q(z, \alpha), \alpha \in \Lambda \), then the VC dimension is equal to infinity.

Definition 4 The VC Dimension of a Set of Real Functions (Vapnik, 1979)

Let \( A \leq Q(z, \alpha) \leq B, \alpha \in \Lambda \), be a set of real functions bounded by constants \( A \) and \( B \) (\( A \) can be \(-\infty\) and \( B \) can be \( \infty \)).

The indicator of level \( \beta \) for the function \( Q(z, \alpha) \) shows for which \( z \) the function \( Q(z, \alpha) \) exceeds \( \beta \) and for which it does not. The function \( Q(z, \alpha) \) can be described by the set of all its indicators.

\(^1\) Any indicator function separates a given set of vectors into two subsets: the subset of vectors for which this indicator function takes the value 0 and the subset of vectors for which this indicator function takes the value 1.
Let us consider along with the set of real functions $Q(z, \alpha)$, $\alpha \in \Lambda$, the set of indicators
\[ I(z, \alpha, \beta) = \theta\{Q(z, \alpha) - \beta\}, \alpha \in \Lambda, \beta \in (A, B), \] (5.1.1)
where $\theta(z)$ is the step function
\[ \theta(z) = \begin{cases} 
0 & \text{if } z < 0, \\
1 & \text{if } z \geq 0. 
\end{cases} \] (5.1.2)

The VC dimension of a set of real functions $Q(z, \alpha)$, $\alpha \in \Lambda$, is defined to be the VC dimension of the set of corresponding indicators (5.1.1) with parameters $\alpha \in \Lambda$ and $\beta \in (A, B)$.

No Free Lunch Theorems

The two main no free lunch theorems (NFL) are introduced and then evolutionary algorithms and statistical learning theory are reconciled with the NFL theorems. The theorems are novel, non-trivial, frequently misunderstood and profoundly relevant to optimization, machine learning and science in general (and often conveniently ignored by the evolutionary algorithms and statistical learning theory communities). I run the world’s only No Free Lunch website.

No Free Lunch Theorem for Optimization/Search

The no free lunch theorem for search and optimization applies to finite spaces and algorithms that do not resample points. The theorem tells us that all algorithms that search for an extremum of a cost function perform exactly the same when averaged over all possible cost functions. So, for any search/optimization algorithm, any elevated performance over one class of problems is exactly paid for in performance over another class. See Wolpert and Macready (1997).

The no free lunch theorem for search implies that putting blind faith in evolutionary algorithms as a blind search/optimization algorithm is misplaced. For example, on average, a genetic algorithm is no better, or worse, than any other search algorithm. In practice, in our universe, one will only be interested in a subset of all possible functions. This means that it is necessary to show that the set of functions that are of interest has some property that allows a particular algorithm to perform better than random search on this subset.

No Free Lunch Theorem for Supervised Machine Learning

Hume (1739–1740) pointed out that ‘even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience’. More recently, and with increasing rigour, Mitchell (1980), Schaffer (1994) and Wolpert (1996) showed that bias-free learning is futile. The no free lunch theorem for supervised machine learning Wolpert (1996) shows that in a noise-free scenario where the loss function is the misclassification rate, in terms of off-training-set error, there are no a priori distinctions between learning algorithms.

More formally, where
\[ d = \text{training set}; \]

http://www.no-free-lunch.org
5.1. Design

\( m = \) number of elements in training set;
\( f = \) ‘target’ input-output relationships;
\( h = \) hypothesis (the algorithm’s guess for \( f \) made in response to \( d \)); and
\( c = \) off-training-set ‘loss’ associated with \( f \) and \( h \) (‘generalization error’ or ‘test set error’)

all algorithms are equivalent, on average, by any of the following measures of risk: \( E(c|d), E(c|m), E(c|f,d) \) or \( E(c|f,m) \).

How well you do is determined by how ‘aligned’ your learning algorithm \( P(h|d) \) is with the actual posterior, \( P(f|d) \). This result, in essence, formalizes Hume, extends him and calls all of science into question.

The NFL proves that if you make no assumptions about the target functions, or if you have a uniform prior, then \( P(c|d) \) is independent of one’s learning algorithm. Vapnik appears to ‘prove’ that given a large training set and a small VC dimension, one can generalize well. The VC dimension is a property of the learning algorithm, so no assumptions are being made about the target functions. So, has Vapnik found a free lunch? VC theory tells us that the training set error, \( s \), converges to \( c \). If \( \epsilon \) is an arbitrary real number, the VC framework actually concerns

\[ P(|c - s| > \epsilon|f,m). \]

VC theory does not concern

\[ P(c|s,m, \text{VC dimension}). \]

So there is no free lunch for Vapnik, and no guarantee that SVMs generalize well. This foray into the no free lunch theorems is to place the work in this chapter in context: we cannot make any general claims about the superiority or otherwise of the algorithms used or developed, at best we can claim that they are well suited to the data sets employed here.

**Kernel Methods**

**Terminology**

The term kernel is derived from a word that can be traced back to c. 1000 and originally meant a seed (contained within a fruit) or the softer (usually edible) part contained within the hard shell of a nut or stone-fruit. The former meaning is now obsolete. It was first used in mathematics when it was defined for integral equations in which the kernel is known and the other function(s) unknown, but now has several meanings in mathematics. As far as I am aware, the machine learning term kernel trick was first used in 1998. A literature review was given on pp. 40–43.

**Definition**

The kernel of a function \( f \) is the equivalence relation on the function’s domain that roughly expresses the idea of ‘equivalent as far as the function \( f \) can tell’.

**Definition 5** Let \( X \) and \( Y \) be sets and let \( f \) be a function from \( X \) to \( Y \). Elements \( x_1 \) and \( x_2 \) of \( X \) are equivalent if \( f(x_1) \) and \( f(x_2) \) are equal, i.e. are the same element of \( Y \). Formally, if \( f : X \rightarrow Y \), then

\[ \ker(f) = \{(x_1, x_2) \in X \times X : f(x_1) = f(x_2)\}. \]
5.1. Design

The kernel trick (described below) uses the kernel as a similarity measure and the term *kernel function* is often used for $f$ above.

Motivation and Description

Firstly, linearity is rather special, and outside quantum mechanics no real system is truly linear [Meiss, 2003]. Secondly, detecting linear relations has been the focus of much research in statistics and machine learning for decades and the resulting algorithms are well understood, well developed and efficient. Naturally, one wants the best of both worlds. So, if a problem is non-linear, instead of trying to fit a non-linear model, one can map the problem from the input space to a new (higher-dimensional) space (called the feature space) by doing a non-linear transformation using suitably chosen basis functions and then use a linear model in the feature space. This is known as the ‘kernel trick’. The linear model in the feature space corresponds to a non-linear model in the input space. This approach can be used in both classification and regression problems. The choice of kernel function is crucial for the success of all kernel algorithms because the kernel constitutes prior knowledge that is available about a task. Accordingly, there is no free lunch (see p. 83) in kernel choice.

Kernel Trick

The kernel trick was first published by Aizerman et al. [1964]. Mercer’s theorem states that any continuous, symmetric, positive semi-definite kernel function $K(x, y)$ can be expressed as a dot product in a high-dimensional space.

If the arguments to the kernel are in a measurable space $X$, and if the kernel is positive semi-definite—i.e.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) c_i c_j \geq 0$$

for any finite subset $\{x_1, \ldots, x_n\}$ of $X$ and subset $\{c_1, \ldots, c_n\}$ of objects (typically real numbers, but could even be molecules)—then there exists a function $\varphi(x)$ whose range is in an inner product space of possibly high dimension, such that

$$K(x, y) = \varphi(x) \cdot \varphi(y).$$

Advantages

- The kernel defines a similarity measure between two data points and thus allows one to incorporate prior knowledge of the problem domain.

- Most importantly, the kernel contains all of the information about the relative positions of the inputs in the feature space and the actual learning algorithm is based only on the kernel function and can thus be carried out without explicit use of the feature space. The training data only enter the algorithm through their entries in the kernel matrix (a Gram matrix, see Appendix J.1 (p. 166)), and never through their individual attributes. Because one never explicitly has to evaluate the feature map in the high dimensional feature space, the kernel function represents a computational shortcut.

- The number of operations required is not necessarily proportional to the number of features.
Support Vector Machines

A support vector machine (SVM) is a supervised learning technique from the field of machine learning applicable to both classification and regression. Rooted in the statistical learning theory developed by Vladimir Vapnik and co-workers, SVMs are based on the principle of structural risk minimization (Vapnik and Chervonenkis 1974).

The background mathematics required includes probability, linear algebra and functional analysis. More specifically: vector spaces, inner product spaces, Hilbert spaces (defined in Appendix J.2 (p. 166)), operators, eigenvalues and eigenvectors. A good book for learning the background maths is Introductory Real Analysis (Kolmogorov and Fomin 1975).

Support vector machines (reviewed briefly on p. 43) are the best-known example of kernel methods, the history of which was given on pp. 40–43. The literature on the application of SVMs to the financial domain was covered on pp. 43–51.

The basic idea of an SVM is as follows:

1. Non-linearly map the input space into a very high dimensional feature space (the ‘kernel trick’).

2. • In the case of classification, construct an optimal separating hyperplane in this space (a maximal margin classifier); or

   • in the case of regression, perform linear regression in this space, but without penalising small errors.

Preprocessing

Preprocessing the data is a vital part of forecasting. Filtering the data is a common procedure, but should be avoided altogether if it is suspected that the time series may be chaotic (there is little evidence for low dimensional chaos in financial data (Hsieh 1991)). In the following work, simple averaging was used to deal with missing data. It is good practice to normalize the data so that the inputs are in the range [0, 1] or [-1, 1], here I used [-1, 1]. Care was taken to avoid multicollinearity in the inputs, as this would increase the variance (in a bias-variance sense). Another common task is outlier removal, however, if an ‘outlier’ is a market crash, it is obviously highly significant, so no outliers were removed. Useful references include Masters (1995), Pyle (1999) and (to a lesser extent) Theodoridis and Koutroumbas (2008).

Model Selection

For books on model selection, see Burnham and Anderson (2002) and Claeskens and Hjort (2008). For a Bayesian approach to model selection using foreign exchange data (not reported in this thesis), see Sewell (2008a) and Sewell (2009b). Support vector machines are implemented here, which employ structural risk minimization, and a validation set is used for meta-parameter selection.

Typically, the data is split thus: the first 50% is the ‘training set’, the next 25% the ‘validation set’ and the final 25% the ‘test set’. However, in the experiments below I split the data set in the same manner as that of a published work (Neely et al. 2009), for comparative purposes. The training set is used for
5.1. Design

training the SVM, the validation set for parameter selection and the test set is the out of sample data. The parameters that generated the highest net profit on the validation set are used for the test set.

Can one use $K$-fold cross-validation (rather than a sliding window) on a time series? In other words, what assumptions are made if one uses the data in an order other than that in which it was generated? It is only a problem if the function that you are approximating is also a function of time (or order). Surprisingly, this is a contribution. To be safe, a system should be tested using a data set that is both previously unseen and forwards in time, a rule that I adhered to in the experiments that follow.

**Feature Selection**

First and foremost, when making assumptions regarding selecting inputs, I (among other things) subscribe to Tobler’s first law of geography [Tobler, 1970] that tells us that ‘everything is related to everything else, but near things are more related than distant things’. That is, for example, the following common sense notion is applied: when predicting tomorrow’s price change, yesterday’s price change is more likely to have predictive value than the daily price change, say, 173 days ago. With such noisy data, standard feature selection techniques such as principal component analysis (PCA), factor analysis and independent component analysis (ICA), which are all examples of unsupervised learning, risk overfitting the training set by extracting data structures based on noise. For reasons of market efficiency, it is safest to take the view that there are no privileged features in financial time series, over and above keeping the inputs potentially relevant, orthogonal and utilizing Tobler’s first law of geography. To a degree, the random subspace method (RSM) [Ho, 1998] alleviates the problem of feature selection in areas with little domain knowledge, but was not used here.

**Software**

I wrote two Windows versions of support vector machines, both of which are freely available online (including source code): SVM\textsubscript{dark} is based on SVM\textsuperscript{light} and written in C for Win32, whilst winSVM is based on mySVM and written in C++ for Win32. My Windows SVM software has been used by the financial industry. Both products include a model/parameter selection tool which randomly selects the SVM kernel and/or parameters within the range selected by the user. Results for each parameter combination are saved in a spreadsheet and the user can narrow down the range of parameters and home in on the optimum solution for the validation set. The software comes with a tutorial, and has received a great deal of positive feedback from academia, banks and individuals. The programs make a very real practical contribution to SVM model and parameter selection, as they each present the user with an easy-to-use interface that allows them to select a subset of the search space of parameters to be parsed randomly, and enables them to inspect and sort the results with ease in Excel. The random model/parameter selection is particularly beneficial in applications with limited domain knowledge, such as financial time series. Figure 5.1 and Figure 5.2 (p. 89) show screenshots of my Windows SVM software. The software used for the experiments on forecasting, some of which are reported in this chapter, includes mySVM [Rüping, 2000], SVM\textsuperscript{light} [Joachims, 2004], SVM\textsubscript{dark}, winSVM, LIBSVM [Chang and Lin, 2001] and MATLAB.

---

[3] \url{http://winsvm.martinsewell.com/} and \url{http://svmdark.martinsewell.com/}
Fisher Kernel

Introduction

To save space, my literature review on Fisher kernels is omitted here, but is available for download on the Web ([Sewell](2011)). In common with all kernel methods, the support vector machine technique involves two stages: first non-linearly map the input space into a very high dimensional feature space, then apply a learning algorithm designed to discover linear patterns in that space. The novelty in this section concerns the first stage. The basic idea behind the Fisher kernel method is to train a (generative) hidden Markov model (HMM) on data to derive a Fisher kernel for a (discriminative) support vector machine (SVM). The Fisher kernel gives a ‘natural’ similarity measure that takes into account the underlying probability distribution. If each data item is a (possibly varying length) sequence, the sequences may be used to train a HMM. It is then possible to calculate how much a new data item would ‘stretch’ the parameters of the existing model. This is achieved by, for two data items, calculating and comparing the gradient of the log-likelihood of the data item with respect to the model with a given set of parameters. If these ‘Fisher scores’ are similar it means that the two data items would adapt the model in the same way, that is from the point of view of the given parametric model at the current parameter setting they are similar in the sense that they would require similar adaptations to the parameters.
Figure 5.2: winSVM
5.1. Design

Markov Chains

Markov chains were introduced by the Russian mathematician Andrey Markov in 1906 (Markov 1906), although the term did not appear for over 20 years when it was used by Bernstein (1927). A Markov chain is a discrete-state Markov process. Formally, a discrete time Markov chain is a sequence of random variables \( X_n, n \geq 0 \) such that for every \( n \),

\[
P(X_{n+1} = x | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n).
\]

In words, the future of the system depends on the present, but not the past.

Hidden Markov Models

A hidden Markov model (HMM) is a temporal probabilistic model in which the state of the process is described by a single discrete random variable. Loosely speaking, it is a Markov chain observed in noise. The theory of hidden Markov models was developed in the late 1960s and early 1970s by Baum, Eagon, Petrie, Soules and Weiss (Baum and Petrie 1966), Baum and Eagon (1967), Baum et al. (1970), Baum (1972), whilst the name ‘hidden Markov model’ was coined by L. P. Neuwirth. For more information on HMMs, see the tutorial papers Rabiner and Juang (1986), Poritz (1988), Rabiner (1989) and Eddy (2004), and the books MacDonald and Zucchini (1997), Durbin et al. (1999), Elliot et al. (2004) and Cappé et al. (2005). HMMs have earned their popularity largely from successful application to speech recognition (Rabiner 1989), but have also been applied to handwriting recognition, gesture recognition, musical score following and bioinformatics.

Formally, a hidden Markov model is a bivariate discrete time process \( \{X_k, Y_k\}_{k \geq 0} \), where \( X_k \) is a Markov chain and, conditional on \( X_k \), \( Y_k \) is a sequence of independent random variables such that the conditional distribution of \( Y_k \) only depends on \( X_k \).

The successful application of HMMs to markets is referenced as far back as Kemeny et al. (1976) and Juang (1985). The books Bhar and Hamori (2004) and Mamon and Elliott (2007) cover HMMs in finance.

Fixed Length Strings Generated by a Hidden Markov Model

Parts of the final chapter of Shawe-Taylor and Cristianini (2004)—which covers turning generative models into kernels—are followed below.

Let us assume that one has two strings \( s \) and \( t \) of fixed length \( n \) that are composed of symbols from an alphabet \( \Sigma \). Furthermore it is assumed that they have been generated by a hidden model \( M \), whose elements are represented by strings \( h \) of \( n \) states each from a set \( A \), and that each symbol is generated independently, so that

\[
P(s,t|h) = \prod_{i=1}^{n} P(s_i|h_i)P(t_i|h_i).
\]

Consider the hidden Markov model

\[
P_M(h) = P_M(h_1)P_M(h_2|h_1)\ldots P_M(h_n|h_{n-1}).
\]

Define the states of the model to be

\[
\{a_I\} \cup A \times \{1, \ldots, n\} \cup a_F,
\]
with the transition probabilities given by

\[ P_M((a, i) | a_I) = \begin{cases} P_M(a) & \text{if } i = 1; \\ 0 & \text{otherwise}, \end{cases} \]

\[ P_M((a, i) | (b, j)) = \begin{cases} P_M(a | b) & \text{if } i = j + 1; \\ 0 & \text{otherwise}, \end{cases} \]

\[ P_M(a_F | (b, j)) = \begin{cases} 1 & \text{if } i = n; \\ 0 & \text{otherwise}. \end{cases} \]

This means that in order to marginalise, one needs to sum over a more complex probability distribution for the hidden states to obtain the corresponding marginalisation kernel

\[ \kappa(s, t) = \sum_{h \in A^n} P(s|h)P(t|h)P_M(h) \]

\[ = \sum_{h \in A^n} \prod_{i=1}^n P(s_i|h_i)P(t_i|h_i)P_M(h_i|h_{i-1}), \tag{5.1.3} \]

where the convention that \( P_M(h_1|h_0) = P_M(h_1) \) has been used.

Each hidden sequence \( h \) is considered as a template for the sequences \( s, t \) in the sense that if it is in state \( h_i \) at position \( i \), the probability that the observable sequence has a symbol \( s_i \) in that position is a function of \( h_i \). In the generative model, sequences are generated independently from the hidden template with probabilities \( P(s_i|h_i) \) that can be specified by a matrix of size \( |\Sigma| \times |A| \). So given this matrix and a fixed \( h \), one can compute \( P(s|h) \) and \( P(t|h) \). The problem is that there are \( |A|^n \) different possible models for generating the sequences \( s, t \), that is the feature space is spanned by a basis of \( |A|^n \) dimensions. Furthermore, a special generating process for \( h \) of Markov type, the probability of a state depends only on the preceding state, is considered. The consequent marginalisation step will therefore be prohibitively expensive, if performed in a direct way. Dynamic programming techniques shall be exploited to speed it up.

Consider the set of states \( A_k^a \) of length \( k \) that end with \( a \) given by

\[ A_k^a = \left\{ h \in A^k : h_k = a \right\}. \]

A series of subkernels \( \kappa_{k,a} \) for \( k = 1, \ldots, n \) and \( a \in A \) are introduced as follows

\[ \kappa_{k,a}(s, t) = \sum_{h \in A_k^a} P(s|h)P(t|h)P_M(h) \]

\[ = \sum_{h \in A_k^a} \prod_{i=1}^k P(s_i|h_i)P(t_i|h_i)P_M(h_i|h_{i-1}), \]

where the definitions of \( P(s|h) \) and \( P(h) \) have been implicitly extended to cover the case when \( h \) has fewer than \( n \) symbols by ignoring the rest of the string \( s \).

Clearly, the HMM kernel can be expressed simply by

\[ \kappa(s, t) = \sum_{a \in A} \kappa_{n,a}(s, t). \]
For $k = 1$ one has
\[
\kappa_{1, a}(s, t) = P(s_1 | a)P(t_1 | a)PM(a).
\]

Recursive equations for computing $\kappa_{k+1, a}(s, t)$ in terms of $\kappa_{k, b}(s, t)$ for $b \in A$ are now obtained, as the following derivation shows
\[
\kappa_{k+1, a}(s, t) = \sum_{h \in A_k^{k+1}} \prod_{i=1}^{k+1} P(s_i | h_i)P(t_i | h_i)PM(h_i | h_{i-1})
\]
\[
= \sum_{b \in A} P(s_{k+1} | a)P(t_{k+1} | a)PM(a | b) \sum_{h \in A_k} \prod_{i=1}^{k} P(s_i | h_i)P(t_i | h_i)PM(h_i | h_{i-1})
\]
\[
= \sum_{b \in A} P(s_{k+1} | a)P(t_{k+1} | a)PM(a | b)\kappa_{k, b}(s, t).
\]

When computing these kernels the usual dynamic programming tables, one for each $\kappa_{k, b}(s, t)$, need to be used, though of course those obtained for $k - 1$ can be overwritten when computing $k + 1$. The result is summarized in the pseudocode below.

\begin{verbatim}
Input  | Symbol strings $s$ and $t$, state transition probability matrix $PM(a | b)$,
       |   initial state probabilities $PM(a)$
       |   and conditional probabilities $P(\sigma | a)$ of symbols given states.

Process | Assume $p$ states, $1, \ldots, p$.
2       | for $a = 1 : p$
3       |   $DPr(a) = P(s_1 | a)P(t_1 | a)PM(a)$;
4       | end
5       | for $i = 1 : n$
6       | Kern = 0;
7       | for $a = 1 : p$
8       |   $DP(a) = 0$;
9       | for $b = 1 : p$
10      |   $DP(a) = DP(a) + P(s_i | a)P(t_i | a)PM(a | b)DPr(b)$;
11      | end
12      | Kern = Kern + DP(a);
13      | end
14      | DPr = DP;
15      | end

Output | $\kappa(s, t) = Kern$
\end{verbatim}

Table 5.1: Pseudocode for the fixed length HMM kernel

The complexity of the kernel can be bounded from the structure of the algorithm by
\[
O\left(n|A|^2\right).
\]
Fisher Kernel

The log-likelihood of a data item \( x \) with respect to the model \( m(\theta^0) \) for a given setting of the parameters \( \theta^0 \) is defined to be

\[
\log L_{\theta^0}(x).
\]

Consider the vector gradient of the log-likelihood

\[
g(\theta, x) = \left( \frac{\partial \log L_{\theta}(x)}{\partial \theta_i} \right)_{i=1}^N.
\]

The Fisher score of a data item \( x \) with respect to the model \( m(\theta^0) \) for a given setting of the parameters \( \theta^0 \) is

\[
g(\theta^0, x).
\]

The Fisher information matrix with respect to the model \( m(\theta^0) \) for a given setting of the parameters \( \theta^0 \) is given by

\[
I_M = E \left[ g(\theta^0, x)g(\theta^0, x)' \right],
\]

where the expectation is over the generation of the data point \( x \) according to the data generating distribution.

The Fisher score gives us an embedding into the feature space \( \mathbb{R}^N \) and hence immediately suggests a possible kernel. The matrix \( I_M \) can be used to define a non-standard inner product in that feature space.

**Definition 6** The invariant Fisher kernel with respect to the model \( m(\theta^0) \) for a given setting of the parameters \( \theta^0 \) is defined as

\[
\kappa(x, z) = g(\theta^0, x)'I_M^{-1}g(\theta^0, z).
\]

The practical Fisher kernel is defined as

\[
\kappa(x, z) = g(\theta^0, x)'g(\theta^0, z).
\]

As explained in the introduction, the Fisher kernel gives a ‘natural’ similarity measure that takes into account an underlying probability distribution. It seems natural to compare two data points through the directions in which they ‘stretch’ the parameters of the model, that is by viewing the score function at the two points as a function of the parameters and comparing the two gradients. If the gradient vectors are similar it means that the two data items would adapt the model in the same way, that is from the point of view of the given parametric model at the current parameter setting they are similar in the sense that they would require similar adaptations to the parameters.

**Fisher Kernels for Hidden Markov Models**

The model can now be viewed as the sum over all of the state paths or individual models with the parameters the various transition and emission probabilities, so that for a particular parameter setting the probability of a sequence \( s \) is given by

\[
P_M(s) = \sum_{m \in A^n} P(s|m)P_M(m) = \sum_{m \in A^n} P_M(s, m).
\]
where
\[ P_M(m) = P_M(m_1)P_M(m_2|m_1) \ldots P_M(m_n|m_{n-1}), \]
and
\[ P(s|m) = \prod_{i=1}^{n} P(s_i|m_i) \]
so that
\[ P_M(s, m) = \prod_{i=1}^{n} P(s_i|m_i)P(m_i|m_{i-1}). \]

The parameters of the model are the emission probabilities \( P(s_i|m_i) \) and the transition probabilities \( P_M(m_i|m_{i-1}) \). For convenience parameters are introduced
\[ \theta_{s_i|m_i} = P(s_i|m_i) \quad \text{and} \quad \tau_{m_i|m_{i-1}} = P_M(m_i|m_{i-1}), \]
where the convention that \( P_M(m_1) = P_M(m_1|m_0) \) with \( m_0 = a_0 \) for a special fixed state \( a_0 \notin A \) is used. The difficulty is that these parameters are not independent. The unconstrained parameters are introduced
\[ \theta_{\sigma,a} \quad \text{and} \quad \tau_{a,b} \]
with
\[ \theta_{\sigma|a} = \frac{\theta_{\sigma,a}}{\sum_{\sigma' \in \Sigma} \theta_{\sigma',a}} \quad \text{and} \quad \tau_{a|b} = \frac{\tau_{a,b}}{\sum_{a' \in A} \tau_{a',b}}. \]

These values are assembled into a parameter vector \( \theta \). Furthermore it is assumed that the parameter setting at which the derivatives are computed satisfies
\[ \sum_{\sigma \in \Sigma} \theta_{\sigma,a} = \sum_{a \in A} \tau_{a,b} = 1, \quad (5.1.4) \]
for all \( a, b \in A \) in order to simplify the calculations.

The derivatives of the log-likelihood with respect to the parameters \( \theta \) and \( \tau \) must be computed. The computations for both sets of parameters follow an identical pattern, so to simplify the presentation first a template that assumes both cases is derived. Let
\[ \tilde{\psi}(b, a) = \frac{\psi(b, a)}{\sum_{b' \in B} \psi(b', a)}, \text{ for } a \in A \text{ and } b \in B. \]

Let
\[ Q(a, b) = \prod_{i=1}^{n} \tilde{\psi}(b_i, a_i)c_i, \]
for some constants \( c_i \). Consider the derivative of \( Q(a, b) \) with respect to the parameter \( \psi(b, a) \) at point \((a^0, b^0)\) where
\[ \sum_{b \in B} \psi(b, a_i^0) = 1 \text{ for all } i. \quad (5.1.5) \]
One has

\[
\frac{\partial Q(a, b)}{\partial \psi(b, a)} = \sum_{k=1}^{n} c_k \prod_{i \neq k} \psi(b_i, a_i)c_i \frac{\partial}{\partial \psi(b, a)} \sum_{i' \in B} \psi(b_i, a_k) \frac{\partial \psi(b_i, a_k)}{\partial \psi(b, a)} \sum_{i' \in B} \psi(b_i, a_k)
\]

\[
= \sum_{k=1}^{n} \left( \frac{[\theta_k^0 = b][a_k^0 = a]}{\sum_{b' \in B} \psi(b', a_k)^2} - \psi(b_k, a_k) [a_k^0 = a] \right) c_k \prod_{i \neq k} \psi(b_i, a_i)c_i
\]

\[
= \sum_{k=1}^{n} \left( \frac{[\theta_k^0 = b][a_k^0 = a]}{\psi(b, a)} - [a_k^0 = a] \right) \prod_{i \neq k} \psi(b_i, a_i)c_i
\]

\[
= \sum_{k=1}^{n} \left( \frac{[\theta_k^0 = b][a_k^0 = a]}{\psi(b, a)} - [a_k^0 = a] \right) Q(a^0, b^0),
\]

where use of (5.1.5) has been made to obtain the third line from the second. Now return to considering
the derivatives of the log-likelihood, first with respect to the parameter \(\theta_{\sigma,a}\)

\[
\frac{\partial \log P_M(s|\theta)}{\partial \theta_{\sigma,a}} = \frac{1}{P_M(s|\theta)} \sum_{m \in A^n} \prod_{i=1}^{n} P(s_i|m_i)P_M(m_i|m_{i-1})
\]

\[
= \frac{1}{P_M(s|\theta)} \sum_{m \in A^n} \frac{\partial}{\partial \theta_{\sigma,a}} \prod_{i=1}^{n} \theta_{s_i,m_i} \tau_{m_i|m_{i-1}}.
\]

Letting \(a\) be the sequence of states \(m\) and \(b\) the string \(s\), with \(\psi(a, b) = \theta_{b,a}\) and \(c_i = \tau_{m_i|m_{i-1}}\) one has

\[
Q(a, b) = \prod_{i=1}^{n} \theta_{s_i,m_i} \tau_{m_i|m_{i-1}} = P_M(s, m|\theta).
\]

It follows from the derivative of \(Q\) that

\[
\frac{\partial \log P_M(s|\theta)}{\partial \theta_{\sigma,a}} = \sum_{m \in A^n} \sum_{k=1}^{n} \left( \frac{[s_k = \sigma][m_k = a]}{\theta_{\sigma,a}} - [m_k = a] \right) P_M(s, m|\theta)
\]

\[
= \sum_{k=1}^{n} \sum_{m \in A^n} \left( \frac{[s_k = \sigma][m_k = a]}{\theta_{\sigma,a}} - [m_k = a] \right) P_M(m|s, \theta)
\]

\[
= \sum_{k=1}^{n} \mathbb{E} \left[ \frac{[s_k = \sigma][m_k = a]}{\theta_{\sigma,a}} - [m_k = a] \right| s, \theta
\]

\[
= \frac{1}{\theta_{\sigma,a}} \sum_{k=1}^{n} \mathbb{E}[[s_k = \sigma][m_k = a]|s, \theta] - \sum_{k=1}^{n} \mathbb{E}[[m_k = a]|s, \theta],
\]

where the expectations are over the hidden states that generate \(s\). Now consider the derivatives with
respect to the parameter \(\tau_{a,b}\)

\[
\frac{\partial \log P_M(s|\theta)}{\partial \tau_{a,b}} = \frac{1}{P_M(s|\theta)} \frac{\partial}{\partial \tau_{a,b}} \sum_{m \in A^n} \prod_{i=1}^{n} P(s_i|m_i)P_M(m_i|m_{i-1})
\]

\[
= \frac{1}{P_M(s|\theta)} \sum_{m \in A^n} \frac{\partial}{\partial \tau_{a,b}} \prod_{i=1}^{n} \theta_{s_i,m_i} \tau_{m_i|m_{i-1}}.
\]

Letting \(a\) and \(b\) be the sequence of states \(m\) and \(b\) be the same sequence of states shifted one
position, \(\psi(a, b) = \tau_{a,b}\) and \(c_i = \theta_{s_i|m_i}\), one has

\[
Q(a, b) = \prod_{i=1}^{n} \theta_{s_i,m_i} \tau_{m_i|m_{i-1}} = P_M(s, m|\theta).
\]
It follows from the derivative of \( Q \) that
\[
\frac{\partial \log P_M(s|\theta)}{\partial \tau_{a,b}} = \sum_{k=1}^{n} \sum_{m \in A} \left( \frac{[m_{k-1} = b][m_k = a]}{\tau_{a,b}} - [m_k = a] \right) P_M(m|s, \theta)
\]
\[
= \frac{1}{\tau_{a,b}} \sum_{k=1}^{n} \mathbb{E}[m_{k-1} = b][m_k = a]|s, \theta] \mathbb{E}[m_k = a]|s, \theta].
\]

It remains to compute the expectations in each of the sums. These are the expectations that the particular emissions and transitions occurred in the generation of the string \( s \).

The computation of these quantities will rely on an algorithm known as the forwards-backwards algorithm. As the name suggests this is a two-stage algorithm that computes the quantities
\[
f_a(i) = P(s_1 \ldots s_i, m_i = a),
\]
in other words the probability that the \( i \)th hidden state is \( a \) with the prefix of the string \( s \) together with the probability \( P(s) \) of the sequence. Following this the backwards algorithm computes
\[
b_a(i) = P(s_{i+1} \ldots s_n|m_i = a).
\]

Once these values have been computed it is possible to evaluate the expectation
\[
\mathbb{E}[s_k = \sigma|m_k = a]|s] = P(s_k = \sigma, m_k = a|s)
\]
\[
= [s_k = \sigma] \frac{P(s_{k+1} \ldots s_n|m_k = a)P(s_1 \ldots s_k, m_k = a)}{P(s)}
\]
\[
= [s_k = \sigma] \frac{f_a(k)b_a(k)}{P(s)}.
\]

Similarly
\[
\mathbb{E}[m_k = a]|s] = P(m_k = a|s)
\]
\[
= \frac{P(s_{k+1} \ldots s_n|m_k = a)P(s_1 \ldots s_k, m_k = a)}{P(s)}
\]
\[
= \frac{f_a(k)b_a(k)}{P(s)}.
\]

Finally, for the second pair of expectations the only tricky evaluation is \( \mathbb{E}[m_{k-1} = b|m_k = a]|s] \), which equals
\[
\frac{P(s_{k+1} \ldots s_n|m_k = a)P(s_1 \ldots s_{k-1}, m_{k-1} = b)P(a|b)P(s_k|m_k = a)}{P(s)} = \frac{f_a(k-1)b_a(k)\tau_{a,b|a}}{P(s)}.
\]

Hence, the Fisher scores can be evaluated based on the results of the forwards-backwards algorithm. The forwards-backwards algorithm again uses a dynamic programming approach based on the recursion
\[
f_b(i + 1) = \theta_{s_{i+1}|b} \sum_{a \in A} f_a(i) \tau_{b|a},
\]
with \( f_{a_0}(0) = 1 \) and \( f_a(0) = 0 \), for \( a = a_0 \). Once the forward recursion is complete one has
\[
P(s) = \sum_{a \in A} f_a(n).
The initialisation for the backward algorithm is

$$b_a(n) = 1$$

with the recursion

$$b_a(i) = \sum_{a \in A} \tau_{a|} \theta_{a,\sigma_{i+1}|a} b_a(i + 1).$$

Putting all of these observations together the code in Appendix K (pp. 167–169) is obtained, the calculation of the Fisher scores for the transmission probabilities is my own contribution.

**Test**

This subsection concerns the prediction of synthetic data, generated by a very simple 5-symbol, 5-state HMM, in order to test the Fisher kernel. The hidden Markov model used in this thesis is based on a C++ implementation of a basic left-to-right HMM which uses the Baum-Welch (maximum likelihood) training algorithm written by Richard Myers. The hidden Markov model used to generate the synthetic data is shown below. Following the header are a series of ordered blocks, each of which is two lines long. Each of the 5 blocks corresponds to a state in the model. Within each block, the first line gives the probability of the model recurring (the first number) followed by the probability of generating each of the possible output symbols when it recurs (the following five numbers). The second line gives the probability of the model transitioning to the next state (the first number) followed by the probability of generating each of the possible output symbols when it transitions (the following five numbers).

**states:** 5
**symbols:** 5
0.5 0.96 0.01 0.01 0.01 0.01
0.5 0.96 0.01 0.01 0.01 0.01
0.5 0.01 0.96 0.01 0.01 0.01
0.5 0.01 0.96 0.01 0.01 0.01
0.5 0.01 0.01 0.96 0.01 0.01
0.5 0.01 0.01 0.01 0.96 0.01
0.5 0.01 0.01 0.01 0.01 0.96
0.0 0.0 0.0 0.0 0.0 0.0

The step-by-step methodology follows.

1. Create a HMM with 5 states and 5 symbols, as above. Save as hmm.txt.

2. Use generate_seq on hmm.txt to generate 10,000 sequences, each 11 symbols long, each symbol \( \in \{0, 1, 2, 3, 4\} \). Output will be hmm.txt.seq.

3. Save the output, hmm.txt.seq, in Fisher.xlsx, Sheet 1. Split the data into 5000 sequences for training, 2500 sequences for validation and 2500 sequences for testing. Separate the 11th column, this will be the target and is not used until later.

4. Copy the training data (without the 11th column) into stringst.txt.

5. Run train_hmm on stringst.txt, with the following parameter settings: seed = 1234, states = 5, symbols = 5 and \( \text{min\_delta\_psum} = 0.01 \). The output will be hmmt.txt.

6. From Fisher.xlsx, Sheet 1, copy all of the data except the target column into strings.txt.

7. In strings.txt, replace symbols thus: 4 \( \rightarrow \) 5, 3 \( \rightarrow \) 4, 2 \( \rightarrow \) 3, 1 \( \rightarrow \) 2, 0 \( \rightarrow \) 1 (this is simply an artefact of the software). Save.

8. Run Fisher.exe (code given in Appendix K (pp. 167–169)), inputs are hmmt.txt and strings.txt, output will be fisher.txt.

9. Use formati.exe\(^5\) to convert fisher.txt to LIBSVM format: ‘formati.exe fisher.txt fisherf.txt’.

10. Copy and paste fisherf.txt into Fisher.xlsx, Sheet 2 (cells need to be formatted for text).

11. Copy target data from Fisher.xlsx, Sheet 1 into a temporary file and replace symbols thus: 4 \( \rightarrow \) 5, 3 \( \rightarrow \) 4, 2 \( \rightarrow \) 3, 1 \( \rightarrow \) 2, 0 \( \rightarrow \) 1.

12. Insert the target data into Fisher.xlsx, Sheet 2, column A then split the data into training set, validation set and test set.

13. Copy and paste into training.txt, validation.txt and test.txt.

14. Scale the data.

15. Apply LIBSVM for regression with default Gaussian (rbf) kernel \((e^{-\gamma ||\vec{u} - \vec{v}||^2})\) using the validation set to select \( C \in \{0.1, 1, 10, 100, 1000, 10000, 100000\} \) and \( \epsilon \in \{0.00001, 0.0001, 0.001, 0.01, 0.1\} \), ‘svmtrain.exe -s 3 -t 2 [...]’. In practice, five parameter combinations performed joint best on the validation set, namely \( \{C = 1, \epsilon = 0.00001\}, \{C = 1, \epsilon = 0.0001\}, \{C = 1, \epsilon = 0.001\}, \{C = 1, \epsilon = 0.1\} \) and \( \{C = 1, \epsilon = 0.01\} \), so the median values were chosen, \( C = 1 \) and \( \epsilon = 0.001 \). Run LIBSVM with these parameter settings on the test set.

Results are given in Table 5.2. There are five symbols, so if the algorithm was no better than random, one would expect a correct classification rate of approximately 20%. The results are impressive, and evidence the fact that my implementation of the Fisher kernel works.

\(^5\)Available from http://format.martinsewell.com/
5.2 Implementation

Introduction

As reported in the literature review (pp. 43–51), there is evidence that, on average, SVMs outperform ANNs when applied to the prediction of financial or commodity markets. Therefore, my approach focuses on kernel methods, and includes an SVM. The no free lunch theorem for supervised machine learning discussed earlier showed us that there is no free lunch in kernel choice, and that the success of our algorithm depends on the assumptions that we make. The kernel constitutes prior knowledge that is available about a task, so the choice of kernel function is crucial for the success of all kernel algorithms. A kernel is a similarity measure, and it seems wise to use the data itself to learn the optimal similarity measure. This section compares a vanilla support vector machine, three existing methods of learning the kernel—the Fisher kernel, the DC algorithm and a Bayes point machine—and a new technique, a DC algorithm-Fisher kernel hybrid, when applied to the classification of daily foreign exchange log returns into positive and negative.

Data

In Park and Irwin (2004)'s excellent review of technical analysis, genetic programming did quite well on foreign exchange data, and Christopher Neely is the most published author within the academic literature on technical analysis [Neely, 1997], [Neely et al., 1997], [Neely, 1998], [Neely and Weller, 2001], so for the sake of comparison, the experiments conducted in this section use the same data sets as employed in Neely et al. (2009), daily foreign exchange (FX) rates and daily interest rate data. The FX rates were originally from the Board of Governors of the Federal Reserve System, and are published online via the H.10 release. The interest rate data was from the Bank for International Settlements (BIS), and is not in the public domain. All of the data was kindly provided by Chris Neely. Missing data was filled in by taking averages of the data points immediately before and after the missing value. The experiments forecast six currency pairs, USD/DEM, USD/JPY, GBP/USD, USD/CHF, DEM/JPY and GBP/CHF, independently. As in Neely et al. (2009), the data set was divided up thus: training set 1975–1977 (783 data points), validation set 1978–1980 (783 data points) and the (out-of-sample) test set spanned 1981–30 June 2005 (6391 data points).

Let \( P_t \) be the exchange rate (such as USD/DEM) on day \( t \), \( I_t \) the annual interest rate of the nominator currency (e.g. USD) and \( I_t^* \) the annual interest rate of the denominator currency (e.g. DEM), \( d = 1 \) Monday to Friday and \( d = 3 \) on Fridays, \( n \) is the number of round trip trades and \( c \) is the one-way transaction cost. Consistent with Neely et al. (2009), \( c \) was taken as 0.0005 from 1978 to 1980, then decreasing in a linear fashion to 0.000094 on 30 June 2005. For the vanilla SVM, Bayes point machine,
DC algorithm and DC-Fisher hybrid, the inputs are

\[ \log \frac{P_t}{P_{t-1}}, \log \frac{P_{t-1}}{P_{t-5}}, \log \frac{P_{t-5}}{P_{t-20}}, \]

plus, for four of the currency pairs, USD/DEM, GBP/USD, USD/CHF and GBP/CHF,

\[ \frac{d}{365} \log \frac{1 + \frac{I_{t-1}}{100}}{1 + I_t}, \sum_{i=t-2}^{t-5} \frac{d}{365} \log \frac{1 + \frac{I_i}{100}}{1 + \frac{I_{t-1}}{100}}, \text{and} \sum_{i=t-6}^{t-20} \frac{d}{365} \log \frac{1 + \frac{I_i}{100}}{1 + \frac{I_{t-1}}{100}}. \]

For the Fisher kernel experiment, the original inputs are

\[ \log(\frac{P_{t-9}}{P_{t-10}}) \ldots \log(\frac{P_t}{P_{t-1}}). \]

So, for the vanilla SVM, Bayes point machine, DC algorithm and DC-Fisher hybrid, for USD/JPY and DEM/JPY there were three inputs and for USD/DEM, GBP/USD, USD/CHF and GBP/CHF there were six inputs. For the Fisher kernel there were ten inputs. In all cases, the target is +1 or −1, depending on whether the following day’s log return, \( \log \frac{P_{t+1}}{P_t} \), is positive or negative.

The cumulative net return, \( r \), over \( k \) days is given by

\[ r = \sum_{t=0}^{k-1} \left( \log \frac{P_{t+1}}{P_t} + \frac{d}{365} \log \frac{1 + \frac{I_t}{100}}{1 + \frac{I_{t-1}}{100}} \right) + n \log \frac{1-c}{1+c}. \]

**Vanilla Support Vector Machine**

The experiment employs LIBSVM \(^{[9]}^{[99]}\) Version 2.91, for classification. In common with all of the experiments in this section, a Gaussian radial basis function \((e^{-\gamma \|\vec{u} - \vec{v}\|^2})\) was chosen as the similarity measure. Whilst systematically cycling through different combinations of values of meta-parameters, the SVM is repeatedly trained on the training set and tested on the validation set. Meta-parameters were chosen thus: \( C \in \{10^{-6}, 10^{-5}, \ldots, 10^{6}\} \) and \( \sigma \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100\} \). For each currency pair, the parameter combination that led to the highest net return on the validation set was used for the (out of sample) test set.

**Fisher Kernel**

1. Data consists of daily log returns of FX.

2. Split the data into many smaller subsequences of 11 data points each (with each subsequence overlapping the previous subsequence by 10 data points).

3. For each subsequence, the target is +1 or −1, depending on whether the following day’s log return, \( \log \frac{P_{t+1}}{P_t} \), is positive or negative.

4. Convert each subsequence of log returns into a 5-symbol alphabet \( \{0, 1, 2, 3, 4\} \). Each log return, \( r \), is replaced by a symbol according to the following table, where centiles are derived from the training set. In other words, the range of returns is split into equiprobable regions, and each allocated a symbol.

5. Split the data into training set, validation set and test set as previously described above (p.29).
Table 5.3: Fisher kernel symbol allocation

<table>
<thead>
<tr>
<th>Range</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r &lt; 20\text{th centile} )</td>
<td>0</td>
</tr>
<tr>
<td>( 20\text{th centile} \leq r &lt; 40\text{th centile} )</td>
<td>1</td>
</tr>
<tr>
<td>( 40\text{th centile} \leq r &lt; 60\text{th centile} )</td>
<td>2</td>
</tr>
<tr>
<td>( 60\text{th centile} \leq r &lt; 80\text{th centile} )</td>
<td>3</td>
</tr>
<tr>
<td>( r \geq 80\text{th centile} )</td>
<td>4</td>
</tr>
</tbody>
</table>

6. Exclude target data until otherwise mentioned.

7. For each training set, generate a left-to-right 5-state hidden Markov model, giving us the following parameters: state transition probability matrix and conditional probabilities of symbols given states.

8. Using the program whose C++ code is provided in Appendix K (pp. 167–169), plus the parameters of the HMM and each string from the training set, determine the Fisher scores.

9. Create a new data set using the Fisher scores as the input vectors and the original targets as the targets. Each input vector will have 50 elements, and each target will be either -1 or +1.

10. Using LIBSVM, proceed with an SVM as described for the vanilla SVM above, but using the data set created in 9.

**DC Algorithm**

This section explores another attempt to ‘learn the kernel’, this time using the DC (difference of convex functions) algorithm. For an overview of DC programming, see Horst and Thoai (1999). The convex hull of a set of points \( X \) in a real vector space \( V \) is the minimal convex set containing \( X \). The idea is to learn convex combinations of continuously-parameterized basic kernels by searching within the convex hull of a prescribed set of basic kernels for one which minimizes a convex regularization functional. The method and software used here is that outlined in Argyriou et al. (2006). An implementation written in MATLAB was downloaded from the website of Andreas Argyriou.

In this instance isotropic Gaussian kernels with constrained diagonal covariance matrices are used, there being an infinite number with the variance lying within a prescribed interval. The covariance matrix is then chosen as a function of the available data. The algorithm was trained using the square loss function, \( q(y, v) = (y - v)^2 \). The validation set was used to select the following parameters. The regularization parameter \( \mu \in \{10^{-3}, 10^{-4}, \ldots, 10^{-11}\} \), for USD/DEM, GBP/USD, USD/CHF and GBP/CHF block sizes \( \in \{[6], [3, 3], [2, 2, 2], [1, 1, 2, 2]\} \), for USD/JPY and DEM/JPY block sizes \( \in \{[3], [1, 2]\} \), and for all cases the interval within which the Gaussian kernel variances lie ranges \( \in \{[75, 25000], [100, 10000], [500, 5000]\} \).

http://ttic.uchicago.edu/~argyriou/code/dc/dc.tar
Bayes Point Machine

Given a sample of labelled instances, the so-called version space is defined as the set of classifiers consistent with the sample. Whilst an SVM singles out the consistent classifier with the largest margin, the Bayes point machine (Herbrich et al., 2001) approximates the Bayes-optimal decision by the centre of mass of version space. Tom Minka’s Bayes Point Machine (BPM) MATLAB toolbox7, which implements the expectation propagation (EP) algorithms for training was used. Expectation propagation is a family of algorithms developed by Tom Minka (Minka, 2001b,a) for approximate inference in Bayesian models. The method approximates the integral of a function by approximating each factor by sequential moment-matching. EP unifies and generalizes two previous techniques: (1) assumed-density filtering, an extension of the Kalman filter, and (2) loopy belief propagation, an extension of belief propagation in Bayesian networks. The BPM attempts to select the optimum kernel width by inspecting the training set. The expected error rate of the BPM was fixed at 0.45, and the kernel width $\sigma \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100\}$. Using LIBSVM (Chang and Lin, 2001) a standard support vector machine was trained on the training set with the optimal $\sigma$ found using the BPM and $C \in \{10^{-6}, 10^{-5}, \ldots, 10^6\}$ selected using the validation set.

DC Algorithm–Fisher Kernel Hybrid

This section describes a novel algorithm. First, the Fisher kernel was derived, as described earlier, using the FX data. The data from step 9 of the Fisher kernel method was used. The input data consists of the parameters of the hidden Markov model in the Fisher kernel, namely the emission and transition probabilities, respectively

$$\frac{\partial \log P_M(s|\theta)}{\partial \tau_{a,b}} \text{ and } \frac{\partial \log P_M(s|\theta)}{\partial \theta_{\tau,a}}.$$ 

The input data was scaled. Next, the data was split into training, validation and test sets as previously described. Then, as above, the DC algorithm was used to find an optimal Gaussian kernel using the training data, and the square loss function used. The validation set was used to select the following parameters used in the DC algorithm: $\mu \in \{10^{-3}, 10^{-4}, \ldots, 10^{-11}\}$, block sizes $\in \{[50], [25, 25], [16, 17, 17], [12, 12, 13, 13]\}$ and ranges $\in \{[75, 25000], [100, 10000], [500, 5000]\}$.

5.3 Testing

Results

Tables 5.4–5.10 below show an analysis of the out of sample results. For the sake of comparison, column NWD/NWU shows the results from Neely et al. (1997) published in Neely et al. (2009). Annual returns (AR) are calculated both gross and net of transaction costs. The Sharpe ratios are annualized, and their standard errors (SE) calculated, in accordance with Lo (2002).

Conclusion

The mean gross returns from all six experiments were positive, with NWD/NWU being the highest, followed by BPM and the Fisher kernel, whilst the vanilla SVM is the lowest. The mean net returns

### 5.3. Testing

<table>
<thead>
<tr>
<th>SVM</th>
<th>BPM</th>
<th>Fisher kernel</th>
<th>DC algorithm</th>
<th>Hybrid</th>
<th>NWD/NWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross AR(%)</td>
<td>0.27</td>
<td>0.27</td>
<td>0.84</td>
<td>2.9</td>
<td>−1.09</td>
</tr>
<tr>
<td>Net AR(%)</td>
<td>−3.82</td>
<td>−3.82</td>
<td>−0.56</td>
<td>1.85</td>
<td>−5.33</td>
</tr>
<tr>
<td>t-stat</td>
<td>−1.78</td>
<td>−1.78</td>
<td>−0.26</td>
<td>0.86</td>
<td>−2.49</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>−0.36</td>
<td>−0.36</td>
<td>−0.05</td>
<td>0.17</td>
<td>−0.51</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.21</td>
<td>0.21</td>
<td>0.2</td>
<td>0.2</td>
<td>0.21</td>
</tr>
<tr>
<td>Trades/year</td>
<td>81.47</td>
<td>81.47</td>
<td>28.82</td>
<td>20.57</td>
<td>85.59</td>
</tr>
</tbody>
</table>

Table 5.4: *Out of sample results, USD/DEM*

<table>
<thead>
<tr>
<th>SVM</th>
<th>BPM</th>
<th>Fisher kernel</th>
<th>DC algorithm</th>
<th>Hybrid</th>
<th>NWD/NWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross AR(%)</td>
<td>0.03</td>
<td>−2.63</td>
<td>−1.54</td>
<td>−1.2</td>
<td>3.52</td>
</tr>
<tr>
<td>Net AR(%)</td>
<td>−5.09</td>
<td>−3.46</td>
<td>−4.15</td>
<td>−2.98</td>
<td>1.66</td>
</tr>
<tr>
<td>t-stat</td>
<td>−2.37</td>
<td>−1.61</td>
<td>−1.94</td>
<td>−1.39</td>
<td>0.77</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>−0.48</td>
<td>−0.32</td>
<td>−0.39</td>
<td>−0.28</td>
<td>0.15</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.2</td>
</tr>
<tr>
<td>Trades/year</td>
<td>105.22</td>
<td>17.14</td>
<td>53.35</td>
<td>36.65</td>
<td>38.33</td>
</tr>
</tbody>
</table>

Table 5.5: *Out of sample results, USD/JPY*

<table>
<thead>
<tr>
<th>SVM</th>
<th>BPM</th>
<th>Fisher kernel</th>
<th>DC algorithm</th>
<th>Hybrid</th>
<th>NWD/NWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross AR(%)</td>
<td>−1.68</td>
<td>−1.68</td>
<td>1.04</td>
<td>−1.25</td>
<td>−3.28</td>
</tr>
<tr>
<td>Net AR(%)</td>
<td>−2.9</td>
<td>−2.9</td>
<td>−1.99</td>
<td>−4.01</td>
<td>−6.56</td>
</tr>
<tr>
<td>t-stat</td>
<td>−1.45</td>
<td>−1.45</td>
<td>−1</td>
<td>−2.01</td>
<td>−3.28</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>−0.28</td>
<td>−0.28</td>
<td>−0.19</td>
<td>−0.4</td>
<td>−0.64</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.21</td>
<td>0.21</td>
<td>0.2</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>Trades/year</td>
<td>24.98</td>
<td>24.98</td>
<td>61.96</td>
<td>55.22</td>
<td>66.16</td>
</tr>
</tbody>
</table>

Table 5.6: *Out of sample results, GBP/USD*
5.3. Testing

Table 5.7: Out of sample results, USD/CHF

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>BPM</th>
<th>Fisher kernel</th>
<th>DC algorithm</th>
<th>Hybrid</th>
<th>NWD/NWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross AR(%)</td>
<td>-1.51</td>
<td>3.48</td>
<td>4.04</td>
<td>1.09</td>
<td>2.71</td>
<td>0.25</td>
</tr>
<tr>
<td>Net AR(%)</td>
<td>-4.51</td>
<td>2.44</td>
<td>4.01</td>
<td>-1.81</td>
<td>0.29</td>
<td>0.01</td>
</tr>
<tr>
<td>t-stat</td>
<td>-1.9</td>
<td>1.03</td>
<td>1.69</td>
<td>-0.77</td>
<td>0.12</td>
<td>0</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>-0.38</td>
<td>0.2</td>
<td>0.33</td>
<td>-0.15</td>
<td>0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.21</td>
<td>0.2</td>
<td>0.21</td>
<td>0.2</td>
<td>0.2</td>
<td>0.28</td>
</tr>
<tr>
<td>Trades/year</td>
<td>60.08</td>
<td>18.53</td>
<td>0.82</td>
<td>60.94</td>
<td>48.33</td>
<td>4.88</td>
</tr>
</tbody>
</table>

Table 5.8: Out of sample results, DEM/JPY

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>BPM</th>
<th>Fisher kernel</th>
<th>DC algorithm</th>
<th>Hybrid</th>
<th>NWD/NWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross AR(%)</td>
<td>-2.78</td>
<td>3.43</td>
<td>-0.17</td>
<td>-0.57</td>
<td>-4.89</td>
<td>3.17</td>
</tr>
<tr>
<td>Net AR(%)</td>
<td>-7.75</td>
<td>3.43</td>
<td>-4.63</td>
<td>-3.84</td>
<td>-9.61</td>
<td>2.04</td>
</tr>
<tr>
<td>t-stat</td>
<td>-3.8</td>
<td>1.68</td>
<td>-2.27</td>
<td>-1.88</td>
<td>-4.71</td>
<td>1.34</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>-0.75</td>
<td>0.32</td>
<td>-0.45</td>
<td>-0.37</td>
<td>-0.97</td>
<td>0.35</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.23</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.24</td>
<td>0.3</td>
</tr>
<tr>
<td>Trades/year</td>
<td>99.10</td>
<td>0.08</td>
<td>88.49</td>
<td>70.04</td>
<td>94.82</td>
<td>23.39</td>
</tr>
</tbody>
</table>

Table 5.9: Out of sample results, GBP/CHF

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>BPM</th>
<th>Fisher kernel</th>
<th>DC algorithm</th>
<th>Hybrid</th>
<th>NWD/NWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross AR(%)</td>
<td>7.22</td>
<td>7.22</td>
<td>5.74</td>
<td>5.28</td>
<td>6.43</td>
<td>-0.06</td>
</tr>
<tr>
<td>Net AR(%)</td>
<td>7.22</td>
<td>7.22</td>
<td>3.72</td>
<td>4.68</td>
<td>5.49</td>
<td>-0.18</td>
</tr>
<tr>
<td>t-stat</td>
<td>4.15</td>
<td>4.15</td>
<td>2.14</td>
<td>2.69</td>
<td>3.15</td>
<td>-0.03</td>
</tr>
<tr>
<td>Sharpe ratio</td>
<td>0.81</td>
<td>0.81</td>
<td>0.43</td>
<td>0.51</td>
<td>0.61</td>
<td>-0.02</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.23</td>
<td>0.23</td>
<td>0.21</td>
<td>0.21</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td>Trades/year</td>
<td>0.08</td>
<td>0.08</td>
<td>42.65</td>
<td>8.98</td>
<td>19.43</td>
<td>2.26</td>
</tr>
</tbody>
</table>
were positive for NWD/NWU and BPM; NWD/NWU performed best, and the vanilla SVM and hybrid algorithm the worst. BPM, the Fisher kernel, the DC algorithm and the hybrid algorithm were all improvements over the vanilla SVM in terms of both gross returns and net returns, but none achieved net returns as high as NWD/NWU. One likely reason is that the genetic programming methodology was better suited to optimally restricting the number of trades per year. However, the performance of the genetic programming trading system described in Neely et al. (1997) was one of the worst reported in Neely et al. (2009). The following three methods performed best. Sweeney (1986) used filter rules, as described in Fama and Blume (1966). Taylor (1994) considered ARIMA(1,0,2) trading rules, prespecifying the ARIMA order and choosing the parameters and the size of a ‘band of inactivity’ to maximize in-sample profitability. Dueker and Neely (2007) used a Markov-switching model on deviations from uncovered interest parity, with time-varying mean, variance, and kurtosis to develop trading rules. In-sample data was used to estimate model parameters and to construct optimal ‘bands of inactivity’ that reduce trading frequency. I would expect the returns generated by all of the models to diminish with time, especially as they are published, and would not be confident that significant profits could be made in today’s market.

5.4 Conclusion and Summary

The chapter began with an exposition of the no free lunch theorems and explanations of kernel methods and support vector machines. The importance of preprocessing one’s data was discussed and the methods used, namely normalization and avoiding multicollinearity, were highlighted. The methodology regarding model selection and feature selection was described and the software introduced. Many trading systems were built, which traded large cap US stocks intra-day, FTSE 100 constituents daily, US stocks weekly, FTSE 100 constituents monthly, US stocks daily with fundamental inputs, commodities daily and FX daily, although to save space, only the experiments on the final data set are reported. The applications of the Fisher kernel, the DC algorithm and Bayes point machine to financial time series are all new. Most novel of all was the use of the DC algorithm to learn the parameters of the hidden Markov model in the Fisher kernel. Table 5.11 gives a summary of the goals achieved in the forecasting chapter.
5.4. Conclusion and Summary

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beat market?</td>
<td>Yes</td>
</tr>
<tr>
<td>Beat standard SVM?</td>
<td>Yes</td>
</tr>
<tr>
<td>Beat state of the art?</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 5.11: Summary of results on forecasting

of this thesis. More precise conclusions are elusive, because a slight change to the data set or the inputs can produce quite different results. Although I believe that machine learning in general, and learning the kernel in particular, have a lot to offer financial time series prediction, financial data is a poor test bed for comparing machine learning algorithms due to its vanishingly small signal-to-noise ratio.
Chapter 6

Assessment

This chapter undertakes a critical assessment of the work by restating the hypothesis (in the context of a well-defined problem) (Section 6.1), demonstrating the precision (Section 6.2) and thoroughness (Section 6.3) of the work on the characterization, modelling and forecasting of financial time series, restating the contributions (Section 6.4) and providing a comparison with the work of others that is most similar to my own (Section 6.5). The chapter ends with a conclusion and summary (Section 6.6).

6.1 Hypothesis

Recently, a great deal of machine learning has been applied to the area of bioinformatics, and the area appears to drive the research within kernel methods (reviewed on pp. 40–43). Meanwhile, the application of machine learning to the financial domain is of a much lower quality (pp. 43–51). There is ample scope for the application of novel kernel methods to financial markets. It was hypothesized that the state of the art in financial time series analysis could be improved upon by applying intelligent systems. Although, in the case of the work on forecasting (Chapter 5), the aims were graduated thus: 1) to ‘beat the market’ (in a foreign exchange market this simply means earn a positive return), 2) to improve standard algorithms, and 3) to beat the ‘state of the art’.

Gershenfeld and Weigend (1994) claim that ‘[t]ime series analysis has three goals: forecasting, modeling, and characterization’. I utilized their time series trichotomy, albeit in a different order, and applied it to financial time series to structure the core of the thesis thus: characterization (Chapter 3), modelling (Chapter 4) and forecasting (Chapter 5).

characterization Characterization attempts with little or no a priori knowledge to determine fundamental properties, such as the stationarity of a system or the amount of randomness.

modelling The goal of modelling is to find a description that accurately captures features of the long-term behaviour of the system.

forecasting The aim of forecasting (also called predicting) is to accurately predict the short-term evolution of the system.

The major time series references were given in Section 3.1.2 (p. 56) and a glossary is provided in Appendix C (p. 135–136).
The thesis research question is: ‘Can one improve upon the state of the art in financial time series analysis through the application of intelligent systems?’

**Well-Defined Problem**

The concept of a *well-defined problem* ([McCarthy, 1956](#) [Newell and Simon, 1972](#) [Nozick, 1993](#) ) is employed in order to make the ‘thesis’ (in the narrow sense of the word) explicit: A *well-defined problem* is one in which each of the following features is explicitly specified and delimited.

1. A *goal*, an evaluative criterion for judging outcomes and states.

2. An *initial state*, consisting of a (starting) situation and the resources that are available to be used.

3. *Admissible operations* that can be used to transform states and resources. These admissible operations are stated in the form of rules that may be applied to transform the initial state and then to transform again and again the resulting transformed states.

4. *Constraints* on what intermediate states can be passed along the way, what final states may be reached, what operations may be done when, how many times, in what order, and so forth.

5. An *outcome*, a final state.

A *solution* to the problem is a sequence of admissible operations that transforms the initial state into an outcome that meets the goal, without violating any constraints at any time along the way.

A well-defined problem applied to the thesis:

**Goal** Improve upon the state of the art in financial time series analysis.

**Initial state** The current state of the art in financial time series analysis.

**Admissible operations** Intelligent systems.

**Constraints** Finite resources such as time, money, processing power, ability and support (‘bounded rationality’).

**Outcome** An improved ability to characterize, model and forecast financial markets.

In other words, a solution to the problem is a sequence of admissible operations (in this case, intelligent systems) that transforms the initial state (the current state of the art in financial time series analysis) into an outcome that meets the goal (improve upon the state of the art in financial time series analysis), without violating any constraints (finite resources such as time, money, ability and support) at any time along the way. The thesis was decomposed into three related hypotheses, as outlined in Section [1.3](#) (p. 15).
6.2 Precision

One should expect a peculiar (non-linear) publication bias in the areas covered in this thesis: the efficient market hypothesis, technical analysis and trading systems. In general, and in common with all other areas of research, positive results are more likely to be published, which in the case of trading systems, technical analysis and behavioural finance, means evidence against the EMH. However, those with a vested interest in supporting the existing paradigm (the EMH) and those with results that are so good that they would rather keep them to themselves are less likely to publish results that highlight market inefficiencies. Further, as academics seek to make a novel contribution there will likely be a bias towards publications showing novel algorithms outperforming established algorithms. For an interesting and short paper on publication bias/positive outcome bias/the ‘file-drawer problem’, see Rosenthal (1979). Ioannidis (2005) claims that most published research findings are false. Assuming that his paper is itself correct, problems with experimental and statistical methods mean that there is less than a 50 per cent chance that the results of any randomly chosen scientific paper are true. The reasons for this include small sample sizes, poor study design, researcher bias and selective reporting. Although I strive to such avoid biases, I cannot guarantee that I am totally immune.

Characterization

For the characterization of financial markets as much data as possible was used. The DJIA was chosen as it is the best-known and second-oldest US stock index, the data set used spans over 81 years. Experiments were conducted on daily, weekly, monthly and annual log returns, de-trended when necessary.

Modelling

The artificial stock market modelled in Chapter 4 replicated mean returns, the standard deviation of returns, the absolute returns correlation and the squared returns correlation of a real stock market. However, the artificial stock market failed to accurately replicate the skewness, kurtosis and autocorrelation of returns.

Forecasting

My foray into using kernel methods for forecasting foreign exchange rates produced mixed results. The best of my models beat the market and improved upon a standard support vector machine, but failed to match the genetic programming methodology of Neely et al. (1997) published in Neely et al. (2009). More generally, it became apparent that attempting to predict individual financial time series is quite possibly a fruitless exercise (especially at intermediate time horizons). One should always accept the only free lunch in finance—diversification—and trade a portfolio of assets, possibly using cointegration. It was after my own prediction that FX markets are the least efficient (explained in Section 6.3 (p. 113)) that James (2006) came to the same conclusion and claimed that only around 10 per cent of the market (the active currency managers) are truly concerned with real returns.

I presented all of the results from my final experiments, and did not cherry-pick either favourable results or data sets that show my algorithms in a favourable light. Data snooping (also known as data snooping)

1Modern portfolio theory (MPT) dictates that the only free lunch in finance is diversification.
Thoroughness and (confusingly, in economics) data mining) occurs when a set of data is used more than once for purposes of inference or model selection. This can lead to biases. When data mining, one has to take into account the fact that one is data mining, also that one has read papers that may have been written on the basis of inferences from the same data set that one’s own work is based on. For example, the S&P 500 has been the subject of an enormous number of studies. Lo and MacKinlay (1990b) noted that tests of financial asset pricing models may yield misleading inferences when properties of the data are used to construct the test statistics. ‘In particular, such tests are often based on returns to portfolios of common stock, where portfolios are constructed by sorting on some empirically motivated characteristic of the securities such as market value of equity. Analytical calculations, Monte Carlo simulations, and two empirical examples show that the effects of this type of data snooping can be substantial.’ In 2000, Halbert White published ‘A reality check for data snooping’ (White, 2000). He specified a new procedure, the ‘Reality Check’, which is a straightforward procedure for testing the null hypothesis that the best model encountered in a specification search has no predictive superiority over a given benchmark model, permitting account to be taken of the effects of data snooping. White claims that his method ‘permits data snooping to be undertaken with some degree of confidence that one will not mistake results that could have been generated by chance for genuinely good results’. Sullivan et al. (1999) utilized White’s Reality Check bootstrap methodology (White, 2000) to evaluate simple technical trading rules while quantifying the data-snooping bias and fully adjusting for its effect in the context of the full universe from which the trading rules were drawn. Aronson (2006) suggests three approaches for dealing with data mining bias. His first, out-of-sample testing, involves excluding one or more subsets of the historical data from the data mining, as used in this thesis. His second approach, randomization methods, includes methods like bootstrapping and the Monte Carlo method. His third suggestion, a data-mining correction factor developed by Markowitz and Xu (1994), deflates the observed performance of the rule that did the best. In theory the best approach of all would be to use Bayesian model selection, such as outlined in Sewell (2009b).

6.3 Thoroughness

Characterization

For the characterization of financial markets four experiments—tests for autocorrelation, long memory and runs tests—were conducted across four time intervals (daily, weekly, monthly and annual), plus newsletters were analysed.

Modelling

The model was built from the bottom up utilizing evolutionary psychology, work so detached from computer science that it was published separately (Sewell, 2011g).

Forecasting

The techniques employed could be applied to any financial or commodity instruments. Of course, one can never guarantee that a system will generate certain returns above the risk-free rate. If one could,
given capital, leverage and enough time, one would eventually own the entire world. One can never be sure that one’s trading system will perform successfully in the future at all. The nature of the markets could change overnight. One cannot predict events such as the ‘September 11 attacks’ in 2001. However, the assumption behind technical analysis (presumably due to aspects of behavioural finance) is that the markets react after the event in a predictable way. If the model fails on the test set, then one must conclude that either the time series is unpredictable or the preprocessing and/or prediction methodology were not suited to the task.

Is it worth trying to predict financial markets?

If a market is weak-form efficient, then technical analysis will fail. If a market is semi-strong-form efficient, then both technical analysis and fundamental analysis will fail. For a discussion of efficient markets, see Section 3.1.3 (pp. 56–60). Just under half of the papers reviewed in my review of the efficient market hypothesis (Sewell, 2011d) support market efficiency, whilst around 30 per cent of the relevant articles reviewed in my review of fund performance (Sewell, 2011c) supported market efficiency. After a century of analysis, there is no clear consensus, and I reject notions that the EMH is clearly true, or that the EMH is clearly false and work with the assumption that market efficiency is relative, not absolute. Recall that a market is said to be efficient with respect to an information set if the price ‘fully reflects’ that information set (Fama, 1970). On the one hand, the definitional ‘fully’ is an exacting requirement, suggesting that no real market could ever be efficient, implying that the EMH is almost certainly false. On the other hand, economics is a social science, and a hypothesis that is asymptotically true puts the EMH in contention for one of the strongest hypothesis in the whole of the social sciences. Strictly speaking the EMH is false, but in spirit is profoundly true. Besides, science concerns seeking the best hypothesis, and until a flawed hypothesis is replaced by a better hypothesis, criticism is of limited value. Due to imperfect arbitrage opportunities and correlated irrational behaviour, I take the view that it is worth trying to predict financial markets due to the potential for high rewards and the enhanced mate-value it provides men (Moxon, 2008; Sewell, 2008b), but recognise that the task is extremely difficult and that the majority of people fail.

What skills are required?

The task of predicting markets should be approached with scientific and statistical rigour. In addition to robust scientific methods, successful system building requires both creativeness (one wishes to identify a signal which others have yet to find) and honesty (avoid data snooping). It is also crucial to attempt to suppress one’s innate overconfidence and optimism.

What is one trying to do?

The no free lunch theorem for supervised machine learning (see p. 83) proves that, under some fairly general conditions, all algorithms are equivalent, on average. In other words, the success of an algorithm says as much about the data as about the algorithm. What’s more, the data of interest here—financial time series—is extremely noisy. The best one can hope for is an algorithm that generalizes well on the data sets of interest. This is achieved by creating an algorithm that successfully exploits one’s intuitive
6.3. Thoroughness

implicit prior knowledge concerning \( P(\text{target}) \) so that it implicitly assumes a \( P(\text{hypothesis} \mid \text{training set}) \) which is aligned with \( P(\text{target} \mid \text{training set}) \), where ‘hypothesis’ is one’s guess for the ‘target’ input-output relationships. In short, one must use their prior knowledge to determine the machine learning bias. SVMs in general assume smoothness priors. Here, domain knowledge has been used to facilitate shrewd subset selection, feature selection and the preprocessing of the data.

What assumptions are being made?

With financial time series, there is little domain knowledge (although, thanks to my review of the literature on the characteristics of financial markets (Sewell [2011]), this thesis uses as much as possible), so one must make do with fairly minimal assumptions. Induction relies upon ‘The principle of uniformity of nature’, which Hume (1748) summed up with the phrase ‘For all inferences from experience suppose, as their foundation, that the future will resemble the past, and that similar powers will be conjoined with similar sensible qualities.’ The author of this thesis—like all technical analysts—makes the assumption that the future signal will be like the past. It is also assumed that the universe is smoother than random. When it comes to feature selection, ‘Tobler’s first law of geography’ is employed (see Section 5.1 (p. 87)).

Which algorithm should one use?

A trading system must identify the relationship between the mean of a dependent variable (log returns) and one or more ‘independent’ variables (technical (and sometimes fundamental) inputs), i.e. perform (possibly non-linear) regression. One may also use classification, and classify the market into ‘up’ or ‘down’ movements, as in this thesis. Certainly, this thesis recommends using a data-driven approach by employing machine learning. Using both theory and domain knowledge, one must, a priori, select a prediction technique. Neftci (1991) showed that technical analysis relies on non-linearities being present. There is ample empirical evidence that a non-linear process contributes to the dynamics of market returns (Hsieh, 1989; Scheinkman and LeBaron 1989; Brock et al. 1991). In their review paper, Park and Irwin (2004) found that, on average, non-linear methods outperformed genetic programming in all three types of market considered: stock markets, futures markets and currency markets. How does a support vector machine compare with its close rival, an artificial neural network (ANN)? Firstly, it should be made clear that SVMs contain a large class of neural networks and radial basis function (RBF) networks as special cases. The development of ANNs followed a heuristic path, with applications and extensive experimentation preceding theory. In contrast, the development of SVMs involved sound theory first, then implementation and experiments. A significant advantage of SVMs is that whilst ANNs can suffer from multiple local minima, the solution to an SVM is global and unique. Two more advantages of SVMs are that they have a simple geometric interpretation and give a sparse solution. Also, unlike ANNs, the computational complexity of SVMs does not depend on the dimensionality of the input space. ANNs use empirical risk minimization, whilst SVMs use structural risk minimization. The reason that SVMs often outperform ANNs in practice is that they deal with the biggest problem with ANNs, SVMs are less prone to overfitting. In addition to the theoretical reasons for preferring support vector machines over neural networks, there are empirical reasons. Of the 38 articles in the literature review on pp. [43][51]
that compare SVMs with ANNs in the financial domain, SVMs outperformed ANNs in 32 cases, ANNs outperformed SVMs in 3 cases, and there was no significance difference in 3 cases. More specifically, of the 22 articles that concern the prediction of financial or commodity markets, 18 favoured SVMs, 2 favoured ANNs and 2 found no significant difference. In light of these findings, this author settled for support vector machines as the prediction tool of choice. SVMs are related to smoothness priors, so satisfy that assumption. On a separate note, the nature of trading dictates that the most profitable algorithm may well be one that identifies a trend.

Which markets should one predict?

For reasons of market efficiency, a priori, one would assume that there is no privileged market. As explained in Table 3.1 (p. 55), due to risk aversion, investors require a small positive expected return in risky markets. In long-only markets—like a stock market—this implies a positive upward drift. In symmetric markets which traders are as likely to be long as they are short, like futures and foreign exchange markets, the implication is that one would expect the price to be predictable to some degree. Furthermore, government intervention in foreign exchange markets may provide a positive sum game for other participants in the short-term. **[Neely] 1998, LeBaron 1999, Neely and Weller 2001.** So, for theoretical reasons, one may expect that foreign exchange markets should be the most predictable, futures markets intermediate and stock markets the least predictable. The empirical evidence found in **[Park and Irwin] 2004** and **[James] 2006** confirms this theory. However, a buy-and-hold strategy in the stock market should make money because stock markets are a positive sum game, whilst the same cannot be said for futures or FX markets. Costs in futures and FX markets are tiny; FX is the lowest with $1m of notional costing $3 to trade, whilst futures costs are considerably less than one tick. Costs are dominated by the spread.

At what time frame should one predict?

Again, for reasons of market efficiency, a priori, one would assume that there is no privileged time frame. The world’s most successful hedge funds trade at both ultra-high frequencies (Renaissance Technologies) and over the very long-term (Warren Buffett). The former employ technical analysis, and this is consistent with the literature that finds evidence of dependence at the tick level, but not at longer time horizons (p. 24-25). Buffett employs fundamental analysis, but my analysis of the dependence of annual returns in Chapter 3 implies that technical analysis should work over the long-term too. Like business in general, finding a niche is ideal. If one has the luck and skill of Warren Buffett, one should trade long-term using fundamental analysis; if one is of a quantitative bent and able to invest heavily in IT, one should trade short-term using technical analysis.

**[Conrad and Kaul] 1998** implemented and analysed a wide spectrum of trading strategies during the 1926–1989 period, and during subperiods within, using the entire sample of available NYSE/AMEX securities. They found that a momentum strategy is usually profitable at the medium (3- to 12-month) horizon, while a contrarian strategy nets statistically significant profits at long horizons, but only during the 1926–1947 subperiod. This implies that markets exhibited both persistence and antipersistence, at different time periods, providing an explanation for the success of trading systems in the past, and hope
for the success of systems in the future.

I include transaction costs as an integral part of the methodology, and changes in costs will effect the frequency with which the systems optimally trade.

Strategies

The strategies employed here fall under active investment management and are most likely to be employed by a hedge fund (described on p. 116) or the proprietary trading desk of an investment bank. The financial industry use the term statistical arbitrage (also known as stat arb) to describe the computer-generated ‘black box’ strategies used here. Any misgivings about such systems have more to do with the illusion of control (Langer [1975], Langer and Roth [1975]) than any rational fear. A long/short strategy involves the combined purchase and sale of two securities. A market neutral strategy is a long/short strategy that aims at balancing long and short positions to ensure a zero or negligible market exposure and consequently returns that are independent of market movements. Market neutral strategies are pure alpha strategies. Note that market neutral means beta neutral, not dollar neutral. An example of an equity market neutral strategy is pairs trading, the combination of long and short positions that trade in the same market, are from the same industry and from the same economic sector. Such strategies often rely on some form of mean reversion. Mean reversion only requires one thing: that the mean exists. For example, the spread between two stock prices may be stationary. Figure 6.1 shows the net market exposure of various strategies.

![Figure 6.1: Net market exposure for various strategies in equities](Image)

Finally, my review of the literature on fund performance [Sewell, 2011c] concluded that stock picking is a worthwhile activity, whilst market timing is not; broadly speaking, this favours fundamental
Marketing

Beating the benchmark is only half the game in the real world: the other half is marketing. There is no point in having a great strategy if one is unable to raise enough funds to implement it profitably. It is worth noting that the marriage of strategies and marketing has generated some terminology and methodology which deserves closer examination. Firstly, the expression *statistical arbitrage* conjures up images of robustness (from ‘statistics’) and risk free (from ‘arbitrage’), whilst the reality is that statistical arbitrage is simply gambling in the markets. Although there is an investor–speculator continuum with someone who holds cash and is long the entire market being the only pure investor, and at least stat arb is gambling with a positive edge. A strategy may be market neutral, and marketed as such so that the potential investor can then invest a portion of their wealth accordingly; whilst the constraints (and resultant transaction costs) would likely compromise expected returns to a greater extent than a less restrictive strategy. The *Sharpe ratio* is a popular but flawed performance metric which is open to manipulation (see p. 77). *Drawdowns* are another favourite method of evaluating performance (no one likes losing money), but the metric relies upon two assumptions (both of which must be satisfied for the use of drawdowns to make sense). Firstly, drawdowns must describe the risk-preferences of the investor and secondly the returns from the trading system must not be independent (if the returns are independent, the shape of the curve is irrelevant). In practice, the second assumption implies that the magnitude of the signal in the market displays persistence, plus the trading system’s predictions are not conditioned on the magnitude of said signal. Using maximum drawdown does make sense, however, when maximum drawdown is calculated repeatedly for a bootstrapped sample. The busy manager’s favourite is the *equity curve*, the idea being that he has neither the time nor the ability to examine the strategy (and everyone likes a picture); again, it only makes sense under assumptions very similar to those under which the use of drawdowns makes sense. *Stop-losses* may appear to mirror investors’ risk preferences, but in practice are often an example of pandering to marketing. The use of stop losses only enhances returns under the assumption of persistence in the market. An assumption that is wrong as often as it is right [Conrad and Kaul (1998) show that this is the case, albeit over long time intervals]. All of the above practices necessarily decrease expected returns, but may be consistent with investors’ risk preferences. Also amusing is the story-telling that takes place, allegedly, to explain why a trading system stops working. The assumption here is that the system worked in the first place, when in reality they may have simply been lucky, and their luck ran out. I am myself guilty of some of the points I’ve outlined. I take the view that any financial time series is close to a martingale, and any trading algorithm must be explained in terms of how the market deviates from a martingale. Of course, this process may take place implicitly.

6.4 Contributions

The major contributions made are listed below.

- Using tests for autocorrelation and the runs test I reconcile the fact that daily DJIA log returns pass linear statistical tests of efficiency, yet nonlinear forecasting methods can still make above-average
risk-adjusted returns. See Chapter 3.

- The evolved heuristics and biases exhibited by fundamental analysts and technical analysts, inducing underreaction and overreaction, were used to build an agent-based artificial stock market that provided insight into the effect of the relative proportion of technical analysts and fundamental analysts. See Chapter 4.

- A novel investment performance measurement metric, CPTCE, was developed from Tversky and Kahneman’s cumulative prospect theory. The statistic models investors’ empirically-observed risk preferences (people care about losses and gains rather than absolute wealth, evaluate probabilities incorrectly, are loss averse, risk averse for gains, risk seeking for losses and have non-linear preferences), whilst no other performance metric does this effectively. The financial industry have taken interest, with offers to commercialize the product. See Chapter 4.

- Two Windows implementations of SVMs with semi-automated parameter selection were built. SVMdark is based on SVMlight and written in C for Win32, whilst winSVM is based on mySVM and written in C++ for Win32. The software is frequently downloaded and has been used by the financial industry. The source code is also freely available to download. See pp. 87–87.

- A (generative) hidden Markov model was trained on market data to derive a Fisher kernel for a (discriminative) support vector machine, the DC algorithm and a Bayes point machine are also used to create kernels. Furthermore, the DC algorithm was used to learn the parameters of the hidden Markov model in the Fisher kernel. The four algorithms were compared with a vanilla SVM and the results of Neely et al. (1997) published in Neely et al. (2009). See Chapter 5.

The reason that this area of research is profound is that machine learning can be viewed as an attempt to automate ‘doing science’.

To whom is this thesis useful? Outside academia, the contributions are most likely to be of interest to the alternative investment industry. A hedge fund is an ‘alternative investment’ fund that aims to maximize absolute returns, charges high fees and pursues high risk investment strategies. In 1949 Alfred Winslow Jones established the first hedge fund. His inspiration came from the research he was doing for an article he was writing for Fortune about technical methods of market analysis (Jones, 1949). He raised $100,000 (including $40,000 of his own capital) and started an equity fund. Jones’s innovation was to merge two known speculative tools: short sales and leverage. Hedge funds remained relatively obscure until the structure and success of Jones’s fund was covered by Carol Loomis in another article in Fortune (Loomis, 1966). The accelerating growth of hedge funds between 1980 and the 2008 credit crunch was phenomenal. Regardless of whether markets are (increasingly) efficient, the growth in hedge funds indicated that the desire to invest in actively managed funds was growing, not diminishing. Today, hedge fund strategies broadly fall within four areas: long/short, relative value/arbitrage, event-driven and directional. I run a hedge fund portal dedicated to academic research. For more on hedge funds, see the primer Lhabitant (2002) or the more comprehensive Lhabitant (2006).
6.5 Comparison with Similar Work of Others

I’ve given my contemporaries working in the same area as much help as possible, by making much of the content of this thesis available online. Readers with access to a soft copy of this thesis should appreciate the full hyperlinking, both internal cross referencing and externally linking most of the 600+ items in the bibliography to articles on the Web.

Characterization

There is plenty of empirical work on the statistical nature of financial markets, but few good all-inclusive review papers. Cont (2001) is probably the best review paper of stylized facts in financial markets in general, whilst Guillaume et al. (1997) gives an excellent review of the foreign exchange market.

Modelling

The best known models of financial time series are autoregressive conditional heteroskedasticity (ARCH, Engle (1982)) and generalized autoregressive conditional heteroskedasticity (GARCH, Bollerslev (1986)) processes. The work here does not attempt to compete with such models in terms of an accurate statistical description of financial markets.

Barberis et al. (1998), Daniel et al. (1998), Hong and Stein (1999), Veronesi (1999) and Lee and Swaminathan (2000) all present very good models of markets which exhibit both underreaction and overreaction (see pp. 36-37 for details). Again, the relevant work here does not attempt to compete with these models in terms of their explanation of market under- and overreaction.

Martinez-Jaramillo (2007) and Martinez-Jaramillo and Tsang (2009) developed an artificial financial market and modelled technical, fundamental and noise traders. They investigated the different conditions under which the statistical properties of an artificial stock market resemble those of a real financial market. Their approach replicated the stylized facts of a financial market far more accurately than my own; this was possible by including and adjusting a much larger number of parameters. They also investigated the effects on the market when the agents learn, whilst in my model, by design, no learning takes place. On average, their model without learning replicated the stylized facts most accurately, but not by much (Martinez-Jaramillo (2007), Table 6.5).

Forecasting

Trading systems for equity and commodity markets which trade at various time scales have been built using support vector machines with varying degrees of success, some of which are reported here. It is hypothesized that using SVMs on financial time series is more effective than linear regression (due to non-linearities in the market) and neural networks (due to overfitting). Due to the necessarily secretive nature of the financial industry, little is known of the methodology in use by the most successful systematic traders. However, one is generally aware of their performance. Certainly, I do not pretend to compete with the best in the world on that basis. Taub (2006) reported that Renaissance Technologies (who also employ a scientific approach to trading) had made approximately a 34 per cent annualized net return since its 1988 inception. The ultimate test: would I trade the systems? Currently, no; but with access to more data, processing power, experience and time I reserve the right to be optimistic about the
future. As mentioned earlier, the no free lunch theorem for supervised machine learning (p. 83) dictates that, under some fairly general conditions, all algorithms are equivalent, on average. Therefore the comparison of algorithms in a general setting is futile. The best one can hope for is that their algorithms (implicitly) exploit prior knowledge in the form of a learning bias more effectively than the competition.

The performance metric, cumulative prospect theory certainty equivalent (CPTCE), is new to the domain of finance. As a descriptive measure of people’s attitude towards risk, it should be superior to any other existing measure, even allowing for seemingly irrational behaviour such as simultaneously purchasing insurance and lottery tickets. As a prescriptive description of how people should invest, something like the trade-off between an optimal growth strategy and the security of holding cash advocated by MacLean et al. [1992] or the iterated log function described in McDonnell [2008] would likely be better.

**Thesis**

Taken as a whole, how does the thesis compare with similar work by others? The fast pace of computer science makes anything other than recent comparisons unfair. I have compiled a list of publications that in some ways may be considered to be similar to my own in Appendix L (p. 170–171).

Below, the approach of the discipline used here, machine learning, is contrasted with other approaches.

**Commercial world** According to Nassim Nicholas Taleb[3], while Wall Street research departments may be way ahead of academia in pure derivatives pricing (and other abstractions), they, surprisingly, lag in the more relevant area of quantitative empiricism. Hedge funds are notoriously secretive, so little is known of their strategies.

**Engineering** Engineers take a robust and pragmatic approach, but care is needed to avoid the over-enthusiastic application of an engineer’s tools. For example, whilst the Kalman filter is frequently applicable in the physical world, the assumptions on which it is built (linear and Gaussian) are not relevant to financial time series. Having said that, one always has to make some assumptions.

**Economics** The Victorian historian Thomas Carlyle gave economics the nickname ‘dismal science’. Economics is often criticised for being founded upon dubious assumptions of rationality, unrealistic risk preferences and fanciful normally distributed returns (see, for example, the capital asset pricing model (CAPM) (Treynor [1962], Sharpe [1964], Lintner [1965], Mossin [1966])). You may have heard the joke about the three hungry castaways on a deserted island who are trying to open a can of food. The physicist proposes breaking it open with a sharp rock, the chemist suggests heating the can until it bursts, and the economist says ‘Assume we have a can opener…’.

Econometricians seem obsessed with linear regression analysis. The optimal nature of least squares linear regression is often justified by the Gauss-Markov theorem, which rests on the assumption of linearity, which itself rests on little. This is an important point because outside quantum mechanics, no model of a real system is truly linear (Meiss [2003]). Having said that,
6.5. Comparison with Similar Work of Others

A linear model may still be useful for modelling a non-linear process. For example, the simplest non-trivial model obtainable from the Taylor expansion of any infinitely-differentiable function is a linear model (the first-order expansion of the Taylor series).

Economics is unique among the human and social sciences in that it is egalitarian. It starts from the premise that all races, social groups, societies and individuals are created equal, i.e. have equal potential. Economists speak of ‘developing nations’ and ‘developed nations’, there is never any question of whether or not the developing nations will one day be developed, it is taken that they will catch up. The problem with the assumption of egalitarianism is that scientific psychology and work on intelligence shows that it is profoundly wrong. See, for example, Herrnstein and Murray (1994), Lynn (2002), Lynn (2006), Lynn and Vanhanen (2006) and Lynn (2008).

Physics Physics is a natural science and a physical science, whilst financial markets concern human aspects of the world and are therefore better described as a social science. When tackling finance, physicists tend to either shoehorn a social science into their own paradigm (e.g. modelling the market using spin glass theory (Bornholdt, 2001)) or at the very least make unintuitive assumptions (such as treating the market as a minority game (Challet et al., 2000)). A profound, yet often overlooked, difference is that in physics there are constants and absolute sizes, whilst in economics and finance there are not. For an excellent introduction to ‘econophysics’, see Mantegna and Stanley (2000). As an aside, I often wonder if encouraging physicists (the brightest of us all (Motl, 2006)) into the financial domain (which simply moves money around) is such a good idea, as they could be doing something useful.

Statistics At the risk of oversimplifying, statistics is largely concerned with testing a given hypothesis, whilst machine learning is more concerned with formulating the process of generalization as a search through possible hypotheses in an attempt to find the best hypothesis. See Witten and Frank (2005) pp. 29–30. Machine learning strikes me as a much better paradigm than statistics for the sort of problems of inference that we are interested in solving, and for conducting science in general. In fact, classical statistical inference seems wrong-headed—who cares about the null hypothesis and p-values? We’re generally more interested in finding the best hypothesis than testing a particular hypothesis, and being able to put a probability on a hypothesis.

Machine learning The task of forecasting financial markets is one of predicting a time series generated from a social science, which in practice may be considered as purely an exercise in information processing. This was achieved by making minimal assumptions and using a data-driven, model-free, flexible and nonparametric approach. In other words, I used machine learning, in the guise of supervised learning, which encompasses both theoretical soundness and experimental effectiveness. Multiagent systems were the most natural way of modelling a market with many agents. Being a relatively fast changing discipline, computer science can be rushed, which naturally compromises quality. Also, computer scientists are far too overconfident and optimistic, seemingly unaware just how efficient markets are. My other main criticism is that a lot of the machine learn-
6.6 Conclusion and Summary

The work undertaken during the course of this thesis is important across various disciplines. The use of the DC algorithm to learn the parameters of the HMM in the Fisher kernel is a novel algorithmic contribution to computer science, whilst my support vector machine software for Windows has proved popular and introduced SVMs to a wider audience. The investment performance measurement metric I developed, CPTCE, is in use by, and is truly beneficial to, the financial industry. My genuinely novel work on the evolutionary foundations of heuristics and biases (Sewell, 2011g) (not reported here) should be of great interest to psychologists.

Should the thesis be judged from an engineering perspective, here lies a summary of the programs written:

- Rescaled Range Analysis in C++ and Visual Basic for Excel
- Runs Test in Visual Basic for Excel
- Performance Measurement Calculator in PHP and Visual Basic for Excel
- Performance Metric Analysis in Visual Basic for Excel
• SVM_{dark} in C for Win32

• winSVM in C++ for Win32

• Fisher Kernel in C++

• Monte Carlo Portfolio Optimization in Visual Basic for Excel

• Kelly Criterion in PHP and Visual Basic for Excel

• Order Book Reconstruction in C#
Chapter 7

Conclusion and Future

The final chapter starts with a conclusion that includes a summary of the thesis, and finishes with ideas for extending the current work and several suggestions for further work.

7.1 Conclusion

The central argument of the thesis is that one can improve upon the state of the art in financial time series analysis through the application of intelligent systems. The following section concludes by summarizing the thesis and highlighting the contributions made.

Summary

This thesis set out to do three things. It attempted to (1) characterize, (2) model and (3) forecast financial time series using the best methods available to a computer scientist, in the hope that it is possible to improve upon existing methodologies. Along the way, various contributions to both computer science and other disciplines have been made.

Chapter 1: Introduction

This was a short chapter that ‘set the scene’. It started by explaining what is meant by ‘artificial intelligence’ and covered the controversies, failings and subfields of AI. Next, the problems found in the area were addressed. The thesis statement is that *one can improve upon the state of the art in financial time series analysis through the application of intelligent systems*. ‘Financial time series analysis’ was split into three goals: characterization (Chapter 3), modelling (Chapter 4) and forecasting (Chapter 5). It was explained that the thesis question is important primarily because of the sheer amount of money involved in financial markets. The research undertaken in the thesis was then outlined. Possible future directions were suggested, namely, algorithmic trading, cointegration, copulas, ensemble learning, an equity trading system, evolutionary algorithms, funds of funds, global macro strategies, market-making, merger arbitrage, money management, option pricing, order book, particle filter, investment performance measurement, portfolio optimization, profit-objective error function and yield curve analysis. The contributions to the characterization, modelling and forecasting of financial markets were listed. The chapter finished with an annotated guide to the rest of the thesis.
Chapter 2: Background

The second chapter in the thesis details the work of others: it’s a survey and critical assessment of previous related work and its relation to the research in this thesis. The literature in the following areas is reviewed: the efficient market hypothesis, dependence and long memory in market returns, investment newsletters, technical analysis, behavioural finance, multiagent systems, investment performance measurement, kernel methods and support vector machines, with an emphasis on the financial applications of SVMs.

Chapter 3: Characterization

The chapter defined a stylized fact, and then provided a list of them. The blind alleys that we’ve been led down (stable distributions, long memory in returns and chaos theory) were also highlighted, and used as evidence for the importance of a data-driven approach. Stochastic processes were introduced, a note on time series was included, the efficient market hypothesis discussed and long memory defined. Three experiments were conducted on DJIA returns. First, a test of autocorrelation was run. Second, implementations of Hurst’s rescaled range (R/S) analysis found little evidence of long memory. Third, a runs test was performed showed that daily returns, in particular, exhibit dependence. An analysis of investment letters was undertaken, and it was found that technical analysis performed poorly, evidencing weak-form market efficiency, whilst fundamental analysis gave mixed results.

Chapter 4: Modelling

To set the scene, fundamental analysis, technical analysis, behavioural finance, multiagent systems and investment performance measurement were introduced and discussed. Two experiments are conducted, both utilize behavioural finance. In the first experiment, the evolved heuristics and biases exhibited by fundamental analysts and technical analysts, inducing underreaction and overreaction, are used to build an agent-based artificial stock market. Results showed that whether a fundamental analyst, or a technical analyst, it pays to be in a small majority. The artificial stock market replicated mean returns, the standard deviation of returns, the absolute returns correlation and the squared returns correlation of a real stock market, but failed to accurately replicate the skewness, kurtosis and autocorrelation of returns. In a second experiment, risk preferences were modelled. A novel performance metric, cumulative prospect theory certainty equivalent (CPTCE), was described and developed from prospect theory.

Chapter 5: Forecasting

First, the all-important no free lunch theorems were introduced. Next kernel methods, support vector machines, preprocessing, model selection, feature selection, SVM software and the Fisher kernel are introduced and discussed. The Fisher kernel, the DC algorithm and the Bayes point machine were used to learn the kernel on financial data. Most novel of all, the DC algorithm was used to learn the parameters of the hidden Markov model in the Fisher kernel. Not unsurprisingly, the various systems showed mixed results. Two implementations of SVMs for Windows with semi-automated parameter selection were built.
Chapter 6: Assessment

This chapter undertook a critical assessment of the work by restating the hypothesis, demonstrating precision, thoroughness and contribution, and comparing the work with similar work by others. The approach used in this thesis, machine learning, was contrasted with approaches taken by the commercial world and the disciplines of engineering, economics, physics and statistics.

Chapter 7: Conclusion and Future

The current chapter concludes the thesis, and also addresses potential ideas for further work. Work which is a direct extension to the work on characterization, modelling and forecasting financial time series is considered. In addition, potential new avenues for the application of intelligent techniques are explored which include algorithmic trading, cointegration, copulas, ensemble learning, equity trading system, evolutionary algorithms, funds of funds, global macro strategies, market-making, merger arbitrage, money management, option pricing, order book, particle filter, investment performance measurement, portfolio optimization, profit-objective error function and yield curve analysis.

The major contributions made are listed below.

- Using tests for autocorrelation and the runs test I reconcile the fact that daily DJIA log returns pass linear statistical tests of efficiency, yet nonlinear forecasting methods can still make above-average risk-adjusted returns. See Chapter 3.

- The evolved heuristics and biases exhibited by fundamental analysts and technical analysts, inducing underreaction and overreaction, were used to build an agent-based artificial stock market that exposed the effect of the relative number of technical and fundamental analysts. See Chapter 4.

- A novel investment performance measurement metric, CPTCE, was developed from Tversky and Kahneman’s cumulative prospect theory. The statistic models investors’ empirically-observed risk preferences (people care about losses and gains rather than absolute wealth, evaluate probabilities incorrectly, are loss averse, risk averse for gains, risk seeking for losses and have non-linear preferences), whilst no other performance metric does this effectively. The financial industry have taken interest, with offers to commercialize the product. See Chapter 4.

- Two Windows implementations of SVMs with semi-automated parameter selection were built. SVM\textsubscript{dark} is based on SVM\textsubscript{light} and written in C for Win32, whilst winSVM is based on mySVM and written in C++ for Win32. The software is frequently downloaded and has been used by the financial industry. The source code is also freely available to download. See p. 87.

- A (generative) hidden Markov model was trained on market data to derive a Fisher kernel for a (discriminative) support vector machine (SVM), the DC algorithm and the Bayes point machine are also used to create kernels. Furthermore, the DC algorithm was used to learn the parameters of the hidden Markov model in the Fisher kernel. The four algorithms were compared with a standard SVM and the genetic programming model developed in Neely et al. (1997) and reported in Neely et al. (2009). See Chapter 5.
7.2 Further Work

In this chapter potential ideas for further work in this field are addressed. If the field of research were entirely efficient, there wouldn’t be any point in expending time and effort in seeking novel research, as if it was viable, someone else would have already done it. Fortunately, perfect efficiency is impossible (Grossman and Stiglitz [1980]). However, it is true that if I had thought of anything worth doing in the allotted time, I would have already done it. Thankfully, a PhD takes a finite amount of time\(^1\), so there is always room for further work.

Firstly, logical extensions to the work covered in this thesis on the characterization, modelling and forecasting of financial markets are considered. Secondly, this chapter explores further potential applications, further removed from the current work, but still within areas that could benefit from the ‘computer science in finance’ paradigm, again with, whenever possible, an emphasis on intelligent systems.

Characterization

Despite the seriousness of the implications for risk and the willingness of many physicists to tackle the problem, the precise distribution and scaling properties of financial time series returns is still an open question. Both the size of available data sets and computing power can only increase, so even in the unlikely event that there is no improvement in current methods of analysis, one should always be in an increasingly better position to accurately characterize financial time series.

Modelling

Although not reported here, I did some research on the evolutionary foundations of heuristics (Sewell [2011g]). The fields of evolutionary psychology and behavioural finance are both relatively young and growing. The intersection of the two fields is certainly ripe for future research. More specifically, an attempt could be made to represent all six general purpose heuristics identified by Gilovich and Griffin [2002] by introducing more parameters to the artificial stock market. Further, some of the six special purpose heuristics also identified could be introduced. The recognition heuristic (Goldstein and Gigerenzer [1999, 2002], in particular, appears to be the most accepted. As and when psychologists identify more heuristics and biases, one will be increasingly able to capture the irrationalities of human behaviour in the model. Model validation is currently unsatisfactory (although this is a problem with agent-based modelling in general, as highlighted on p. 67), which leaves an obvious area for future work.

Forecasting

All methods of investment analysis are limited by the amount of data available; but supervised learning suffers the most, so as the quantity and quality of available data increases, and the faster computers get, the more of an edge systematic trading should have.

When building the trading systems, not every facet of the extensive literature review of stylized facts has been utilized. If (say) a January effect were to exist and gradually become known, the only way to profit from it would be to trade on the effect before fellow investors. As all investors strive to get in

\(^1\)Ex post.
before the crowd, the January effect would then become a December effect. Such a moving signal would require a more dynamic approach to algorithm development.

Investors may wish to utilize the models as part of a larger portfolio. For this reason, a shrewd investor would wish to know the distribution of returns, or at least what strategy the system employs. The distributions of returns could be analysed.

Volatility is non-stationary both in the short-term and the long-term. But what about the market signal? Is volatility a proxy for noise? This is of interest because in support vector regression the insensitivity parameter, $\epsilon$, should vary linearly with the noise \cite{Smola1998}.

Regarding the work which involves the Fisher kernel, a great deal more experimentation is required to optimise the various parameters, such as string length, number of symbols and number of states. Also, this author believes that the real value of the Fisher kernel lies with its application to high frequency trading where the structure of the order book comes into play. Future work on the DC algorithm-Fisher kernel hybrid could involve increasing the efficiency of the algorithm (in practice, it was rather slow).

Renaissance Technologies, the world’s most successful hedge fund has averaged 35 per cent annual returns, after fees, since 1989. It is clearly worth attempting to mimic the strategy employed by such a successful fund. Like this author, they use quantitative trading models. They trade with such high-frequency that their Nova fund accounts for over 10 per cent of all the trades occurring on NASDAQ some days. Depending on transaction costs, ultra-high frequency trading utilizing the order book looks attractive.

Hopefully, more of the methods outlined in this thesis will be used in the real world with real money.

**New Directions**

Potential new avenues of research are outlined below.

**Algorithmic Trading**

*Market impact* is the effect that a market participant has when they buy or sell an asset; it is the extent to which the buying or selling moves the price against the buyer or seller, i.e. upward when buying and downward when selling. A *block trade* is the sale or purchase of a large quantity of securities, normally in excess of 10,000 shares. Due to the adverse market impact of block trades, *algorithmic trading* developed, the aim of which is to split up the order in an optimal manner. The most common benchmark is *VWAP* (volume-weighted average price), which is the ratio of the value traded to total volume traded over a particular time horizon (usually one day). Machine learning could be used to improve algorithmic trading, specifically to alleviate the problem of market impact when placing large orders.

**Cointegration**

*Cointegration* \cite{Engle1987} is an econometric technique for testing the relationship between non-stationary time series variables. If two or more series each have a unit root, that is $I(1)$, but a linear combination of them is stationary, $I(0)$, then the series are said to be cointegrated. For example, a stock market index and the price of its associated futures contract, whilst both following a random walk, will be in a long-run equilibrium and deviations from this equilibrium will be stationary. Robert Engle
and Clive Granger shared the 2003 Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, the latter’s portion due to his contribution to the development of cointegration. In an excellent book [Alexander (2001)] points out that if the allocations in a portfolio are designed so that the portfolio tracks an index, then the portfolio should be cointegrated with the index. She proposes that in a long-only fund one should track $\log(\text{index}) + \alpha$, whilst I would extend this and define $\alpha$ as $\alpha = \text{constant} \times \text{standard deviation of log returns over the training period}$. Secondly, triangular cointegration is (as far as I am aware) a novel idea which may be employed whenever there exists a linear relationship between a linear combination of two cointegrated assets and a third asset. The benefits include reduced transaction costs when compared to ordinary cointegration because only the third asset is traded.

**Copulas**

A *copula* is a function that joins univariate distribution functions to form multivariate distribution functions. One could employ copulas to inspect the relationships between the inputs and the target in supervised learning to ensure that the inputs are combined in an optimal manner.

**Ensemble Learning**

Now that the author has completed an extensive literature review on ensemble learning ([Sewell 2011b](#)) and also has the ability to build several varieties of trading system, it is tempting to attempt to combine their predictions in an optimal way.

**Equity Trading System**

Going back to basics, a rule-based equity trading system could be employed (after all, the signal may turn out to be simple). The main advantage is that it would avoid the ‘black box’ nature of the other predictive systems in this thesis, and so would have greater explanatory powers.

Potential inputs to an equity trading expert system:

- Previous day’s performance of the stock ([French and Roll 1986](#))
- Previous week’s performance of the stock ([Campbell et al. 1996](#))
- Previous month’s performance of the stock ([Jegadeesh 1990](#))
- Previous 9-month performance of the stock relative to the market ([Conrad and Kaul 1998](#))
- Previous day’s performance of the stock index ([Campbell et al. 1996](#))
- Previous week’s performance of the stock index ([Campbell et al. 1996](#))
- Previous month’s performance of the stock index ([Campbell et al. 1996](#))
- Volume ([Llorente et al. 2002](#))
- Day of the month ([Ariel 1987](#)) (depending on the market ([Kunkel et al. 2003](#))
- Month of the year ([Rozell and Kinney 1976](#))
- Does the day precede a holiday? ([Ariel 1990](#))
Evolutionary Algorithms

Evolutionary Algorithms have been fruitfully applied to financial time series prediction (the literature was reviewed on p. 52), and the genetic programming model described in Neely et al. (1997) performed better than the kernel methods reported in Chapter 5 (but not as well as other methods reported in Neely et al. (2009)).

Funds of Funds

A fund of funds is an investment partnership that invests in a series of other funds, the object being to diversify. Intelligent techniques could be employed to select funds and allocate capital according to various performance metrics. I have written a Monte Carlo simulation for portfolio optimization, which calculates mean return, standard deviation, skewness, kurtosis, beta, Jensen’s alpha, Sharpe ratio, Sortino ratio, Treynor’s measure, information ratio, Stutzer ratio, Omega, $M^2$, $T^2$, maximum drawdown, Cornish-Fisher-VaR, Mean of all Values and Value at Risk.

Global Macro Strategies

Global macro is a directional hedge fund strategy that invests globally based upon macro economic or ‘top-down’ analysis. Typified by George Soros’s Quantum Fund they have long been the most successful and most visible category of hedge funds. The ultimate ego trip, the strategy seeks to profit by making leveraged bets on anticipated price movements of global stock markets, interest rates, foreign exchange rates and physical commodities. Such strategies are not normally associated with quantitative techniques. It is this author’s belief that they should be.

Market-Making

A market-maker is an intermediary who creates a market for a financial obligation. In a given market, he must quote two prices: the lower is the bid (the price at which he is willing to buy) and the higher is the offer (or ask) (the price at which he is willing to sell). The difference between an offer price and the bid price is known as the spread. A market-maker receives the full order flow, so is in a unique position to profit from the stream of data received. An automated market-making algorithm could be designed using machine learning; it would need to accommodate the following three objectives: attract order flow, control inventories and avoid losses to informed traders (‘adverse selection’). Limit orders are disproportionately more likely to come from informed traders (Harris and Hasbrouck, 1996; Kaniel and Liu, 2006). Bluffers are profit-motivated traders who try to fool other traders into trading unwisely; to avoid losing to bluffers, market-makers must adjust their prices so that buy and sell orders have equal (but opposite) market impact per quantity traded. If most traders use market orders, spreads will be narrow; if most traders use limit orders, spreads will be wide. A market-maker may discover the equilibrium spread by adjusting his spread so that limit orders and market orders are equally likely. Spreads increase with (1) the degree of information asymmetry among traders, (2) volatility, and (3) utilitarian trading interest (a utilitarian trader trades because they expect to obtain some benefit from trading besides profits). For more details on ultra high frequency trading, see Sewell and Yan (2008).

2 Available from http://www.portfoliooptimization.co.uk
7.2. Further Work

Merger arbitrage

Traditionally, *merger arbitrage*, also called *risk arbitrage*, concerns estimating the probability of a deal being approved and how long it will take for the deal to close. There are many factors to consider. For example, a deal may be friendly or hostile, and an offer can consist of shares, cash or debt, or any combination of the three. Furthermore, it is widely recognised that mergers come in waves that tend to coincide with bull markets and economic growth. For background reading, the primer on hedge funds, Lhabitant (2002), includes a chapter on event-driven strategies, and the more comprehensive Lhabitant (2006) a chapter on merger arbitrage. For books specifically on merger arbitrage, see Wyser-Pratte (2009) (originally published in 1982), Moore (1999) and (by far the best) Kirchner (2009). The academic article Block (2006) provides a good overview on merger arbitrage hedge funds. A potentially lucrative area of research would be to forecast future takeover targets, and thus profit from the price jump that generally coincides with a takeover announcement. The most common methods are discriminant analysis and logistic regression, but several authors have applied techniques from machine learning. Słowiński et al. (1997) predicted company acquisitions in Greece, and found that the rough set approach was superior to discriminant analysis. Cheh et al. (1999) successfully predicted US non-financial takeover targets using data from 1985–1993 using a feed-forward backpropagation neural network. Superior results were obtained using the neural network together with discriminant analysis. Tartari et al. (2003) considered four methods for predicting corporate acquisitions, which performed individually as follows (ranked from best to worst): 1st probabilistic neural network, 2nd rough sets, 3rd UTADIS, 4th linear discriminant analysis. The models were then combined using stacked generalization, which performed better than any of the individual methods. Doumpos et al. (2004) predicted acquisition targets in the UK using data from 2000–2002 using four models. UTADIS did best, artificial neural networks were good, logistic regression was bad, and discriminant analysis was the worst. Pasiouras et al. (2005) considered the prediction of acquisition targets within the EU banking industry acquired between 1998 and 2002 and compared and evaluated seven classification methodologies (discriminant analysis, logit analysis, UTilités Additiives DIScriminantes (UTADIS), Multi-group Hierarchical Discrimination (MHDIS), classification and regression trees (CART), k-nearest neighbour (k-NN) and support vector machines (SVMs)) and found that discriminant analysis and SVMs performed best. Tsagkanos et al. (2007) predicted takeover targets in Greece using data from 1995–2002, and found that the machine learning algorithm J4.8 outperformed a classical regression tree, although their predictive accuracy was not promising. Pasiouras et al. (2008) successfully applied SVMs to the prediction of acquisition targets in the EU banking sector. Perhaps the major intellectual interest is that the task concerns classification with an unbalanced data set (a small number of positive cases).

Money Management

For a speculative investor, there are two aspects to optimizing a trading strategy. The first and most important goal of a trader is to achieve a positive expected risk-adjusted return. Once this has been achieved, the trader needs to know what percentage of his capital to risk on each trade. The underlying principals of money management apply to both gambling and trading, and were originally developed for...
7.2. Further Work

I’ve written an implementation of the Kelly criterion and an exponentially-weighted version which gives greater weight to more recent trades. Position sizing is ripe for future development.

Option Pricing

In finance, an option is a type of derivative which is a contract whereby the holder has the right but not the obligation to purchase (a ‘call option’) or sell (a ‘put option’) a specified amount of a security up to (an ‘American option’) or on (a ‘European option’) the expiry date. For an excellent book on options, futures and other derivatives, see Hull (2010). Intelligent techniques as described in this thesis could be applied to option pricing. Fuzziness (Yoshida, 2001), genetic algorithms (Chen and Lee, 1997; Grace, 2000), genetic programming (Chidambaram et al., 2000), neural networks (Garcia and Gençay, 2000), support vector machines (Pires and Marwala, 2004) and an agent-based approach (Suzuki et al., 2009) have all been applied. The field is still fertile for further development.

Order Book

There is structure and some information contained in the order book. Cao et al. (2004) found that the order book beyond the first step provides 30 per cent of the information. Farmer et al. (2004) showed that for the London Stock Exchange when a market order removes all the volume at the best price, it creates a change in the best price equal to the size of the gap, so large price fluctuations occur when there are gaps in the occupied price levels in the limit order book. Weber and Rosenow (2006) found that a low density of limit orders in the order book, i.e. a small liquidity, is a necessary prerequisite for the occurrence of extreme price fluctuations. One could aim to exploit gaps in the order book. For more on ultra high frequency trading, see Sewell and Yan (2008). I have written some software using C# that reconstructs the order book.

Particle Filter

A particle filter (also known as a sequential Monte Carlo (SMC) method) (Fearnhead, 1998; Liu and Chen, 1998; Doucet et al., 2000), is an on-line Bayesian model estimation technique based on simulation. Particle filtering is to on-line learning what Markov chain Monte Carlo (MCMC) is to batch learning; and particle filtering is to non-linear non-Gaussian state-space models what the Kalman filter is to linear Gaussian state-space models. Particle filters approximate posterior distributions by using swarms of points (‘particles’) with associated weights. The method is recursive and involves Monte Carlo integration and importance sampling. As an alternative to regression, if it is suspected that a latent variable is in play, the use of a particle filter may facilitate the forecasting of the price of an asset. I experimented with an implementation written by Adam Johansen.
Investment Performance Measurement

An implementation of CPTCE was considered that would allow investors to tailor their risk preferences \cite{Sewell2009}. The investor could be presented with various questions, and the choices made determine the five parameters, $\alpha$, $\beta$, $\lambda$, $\gamma$ and $\delta$. I have been approached by the financial industry who wished to build and commercialize this implementation as a product.

Portfolio Optimization

Techniques such as stochastic programming could be used for robust portfolio optimization. Also, as with the suggestion for funds of funds above, intelligent techniques could be employed to select assets and allocate capital according to various performance metrics.

Profit-Objective Error Function

There is evidence that, rather than optimizing for mean squared error, one should optimize for profits \cite{Leitch1991, LeBaron1994, Caldwell1995, Bengio1997, Moody1997, Harland2000, Yan2008}. My forecasting experiments in Chapter 5 maximize net profit at the validation stage, but there is ample scope to explore this area further, such as maximizing profit at the training stage (via regression).

Yield Curve Analysis

In fixed income markets in finance, the yield curve is the relation between the interest rate (or cost of borrowing) and the time to maturity of the debt for a given borrower in a given currency. More formally, the yield curve is referred to as the term structure of interest rates. For more information on estimating and interpreting yield curves, see \cite{Anderson1996}. Intelligent techniques, as used in this thesis, could be employed by fixed income analysts to interpolate and predict yield curves with a view to seeking profitable trading opportunities.
Appendix A

ISO 4217 Currency Codes

ISO 4217 currency codes (including some obsolete Euro-zone currencies).

<table>
<thead>
<tr>
<th>Code</th>
<th>Currency</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATS</td>
<td>Schilling</td>
<td>Austria</td>
</tr>
<tr>
<td>AUD</td>
<td>Dollar</td>
<td>Australia</td>
</tr>
<tr>
<td>BEF</td>
<td>Franc</td>
<td>Belgium</td>
</tr>
<tr>
<td>BRL</td>
<td>Real</td>
<td>Brazil</td>
</tr>
<tr>
<td>CAD</td>
<td>Dollar</td>
<td>Canada</td>
</tr>
<tr>
<td>CHF</td>
<td>Franc</td>
<td>Switzerland</td>
</tr>
<tr>
<td>CNY</td>
<td>Yuan (Renminbi (RMB))</td>
<td>Mainland China</td>
</tr>
<tr>
<td>DEM</td>
<td>Deutsche Mark</td>
<td>Germany</td>
</tr>
<tr>
<td>ESP</td>
<td>Peseta</td>
<td>Spain</td>
</tr>
<tr>
<td>EUR</td>
<td>Euro</td>
<td>Euro member countries</td>
</tr>
<tr>
<td>FIM</td>
<td>Markka</td>
<td>Finland</td>
</tr>
<tr>
<td>FRF</td>
<td>Franc</td>
<td>France</td>
</tr>
<tr>
<td>GBP</td>
<td>Pound Sterling</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>GRD</td>
<td>Drachma</td>
<td>Greece</td>
</tr>
<tr>
<td>HKD</td>
<td>Hong Kong</td>
<td>Hong Kong Dollar</td>
</tr>
<tr>
<td>IEP</td>
<td>Pound</td>
<td>Ireland</td>
</tr>
<tr>
<td>INR</td>
<td>Rupee</td>
<td>India</td>
</tr>
<tr>
<td>ITL</td>
<td>Lira</td>
<td>Italy</td>
</tr>
<tr>
<td>JPY</td>
<td>Yen</td>
<td>Japan</td>
</tr>
<tr>
<td>LUF</td>
<td>Franc</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>NLG</td>
<td>Guilder (also called Florin)</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>PTE</td>
<td>Escudo</td>
<td>Portugal</td>
</tr>
<tr>
<td>SEK</td>
<td>Kronor</td>
<td>Sweden</td>
</tr>
<tr>
<td>TWD</td>
<td>New Dollar</td>
<td>Taiwan</td>
</tr>
<tr>
<td>USD</td>
<td>Dollar</td>
<td>United States of America</td>
</tr>
<tr>
<td>VAL</td>
<td>Lira</td>
<td>Vatican City</td>
</tr>
</tbody>
</table>
Appendix B

Exchanges and Stock Market Indices

B.1 Exchanges

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEX</td>
<td>American Stock Exchange</td>
</tr>
<tr>
<td>Athens Stock Exchange (ASE)</td>
<td>Greece’s stock exchange</td>
</tr>
<tr>
<td>Australian Securities Exchange (ASX)</td>
<td>The primary stock exchange in Australia</td>
</tr>
<tr>
<td>Bombay Stock Exchange</td>
<td>Stock exchange in India</td>
</tr>
<tr>
<td>Bolsa de Madrid</td>
<td>The largest stock exchange in Spain</td>
</tr>
<tr>
<td>Borsa Italiana</td>
<td>Italy’s main stock exchange, based in Milan</td>
</tr>
<tr>
<td>Chicago Board of Trade (CBOT)</td>
<td>The world’s oldest futures and options exchange</td>
</tr>
<tr>
<td>Chicago Mercantile Exchange (CME)</td>
<td>Financial and commodity derivative exchange</td>
</tr>
<tr>
<td>Euronext</td>
<td>a pan-European stock exchange based in Amsterdam</td>
</tr>
<tr>
<td>Frankfurt Stock Exchange</td>
<td>The largest stock exchange in Germany</td>
</tr>
<tr>
<td>Helsinki Stock Exchange</td>
<td>Finland’s stock exchange</td>
</tr>
<tr>
<td>London Metal Exchange (LME)</td>
<td>The world’s premier non-ferrous metals market</td>
</tr>
<tr>
<td>London Stock Exchange (LSE)</td>
<td>The most international equities exchange in the world</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>American electronic stock exchange</td>
</tr>
<tr>
<td>National Stock Exchange of India (NSE)</td>
<td>A Mumbai-based stock exchange</td>
</tr>
<tr>
<td>New York Mercantile Exchange (NYMEX)</td>
<td>The world’s largest physical commodity futures exchange</td>
</tr>
<tr>
<td>NYSE</td>
<td>New York Stock Exchange</td>
</tr>
<tr>
<td>Paris Bourse</td>
<td>The historical Paris stock exchange</td>
</tr>
<tr>
<td>SWX Swiss Exchange</td>
<td>Switzerland’s stock exchange, based in Zürich</td>
</tr>
<tr>
<td>Shanghai Stock Exchange (SSE)</td>
<td>A Chinese stock exchange</td>
</tr>
<tr>
<td>Taiwan Stock Exchange (TWSE)</td>
<td>The securities trading center in Taiwan</td>
</tr>
<tr>
<td>Tel Aviv Stock Exchange (TASE)</td>
<td>Israel’s stock exchange</td>
</tr>
<tr>
<td>Tokyo Stock Exchange (TSE)</td>
<td>Japan’s largest stock exchange</td>
</tr>
<tr>
<td>Toronto Stock Exchange (TSX, was TSE)</td>
<td>The largest stock exchange in Canada</td>
</tr>
</tbody>
</table>
## B.2 Stock Market Indices

<table>
<thead>
<tr>
<th>Index Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAC 40</td>
<td>French stock market index representing a capitalization-weighted measure of the 40 most significant values among the 100 highest market caps on the Paris Bourse (now Euronext Paris)</td>
</tr>
<tr>
<td>DAX 30</td>
<td>A stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange</td>
</tr>
<tr>
<td>Dow Jones Industrial Average (DJIA or Dow 30)</td>
<td>Major US stock market index, the average consists of 30 of the largest and most widely held public companies in the United States</td>
</tr>
<tr>
<td>FT 30</td>
<td>An index based on the share prices of 30 British companies</td>
</tr>
<tr>
<td>FTSE 100 index</td>
<td>A share index of the 100 most highly capitalised UK companies listed on the London Stock Exchange</td>
</tr>
<tr>
<td>Hang Seng Index (HSI)</td>
<td>A freefloat-adjusted market capitalization-weighted stock market index in Hong Kong</td>
</tr>
<tr>
<td>Korea Composite Stock Price Index (KOSPI)</td>
<td>The index of all common stocks traded on the Stock Market Division of the Korea Exchange</td>
</tr>
<tr>
<td>Madrid Stock Exchange General Index (IGBM)</td>
<td>The principal index for the Bolsa de Madrid (Madrid Stock Exchange)</td>
</tr>
<tr>
<td>NASDAQ-100</td>
<td>A stock market index of 100 of the largest domestic and international non-financial companies listed on the NASDAQ stock exchange</td>
</tr>
<tr>
<td>Nikkei 225</td>
<td>A stock market index for the Tokyo Stock Exchange</td>
</tr>
<tr>
<td>S&amp;P 100</td>
<td>US stock market index, comprised of 100 leading US stocks</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>Major US stock market index, a value weighted index of the prices of 500 large cap common stocks actively traded in the United States</td>
</tr>
<tr>
<td>S&amp;P CNX Nifty</td>
<td>The leading index for large companies on the National Stock Exchange of India</td>
</tr>
<tr>
<td>Taiwan Stock Exchange Capitallization Weighted Stock Index (TAIEX)</td>
<td>A stock market index for companies traded on the Taiwan Stock Exchange</td>
</tr>
</tbody>
</table>
Appendix C

Time Series Glossary

**AR (autoregressive) model** Models for a time series where the next point is dependent on the previous $n$ points: $AR(n)$. (AR(1) is a Markov chain.)

**ARCH (autoregressive conditional heteroskedasticity)** In econometrics, ARCH ([Engle, 1982](#)) is a model used for forecasting volatility which captures the conditional heteroskedasticity (serial correlation of volatility) of financial returns. Today’s conditional variance is a weighted average of past squared unexpected returns. ARCH is an AR process for the variance.

**ARIMA (autoregressive integrated moving average) models** Models for time series which resemble ARMA models except in that it is presumed the time series has a steady underlying trend. The models therefore work with the differences between the successive observed values, instead of the values themselves. To retrieve the original data from the differences requires a form of integration and the models are therefore called autoregressive integrated moving average models.

**ARMA (autoregressive moving average) models** Models for a time series with no trend (the constant mean is taken as 0). They incorporate the terms in both an autoregressive (AR) model and a moving average (MA) model.

**autocorrelation** A measure of the linear relationship between two separate instances of the same random variable.

**Box-Jenkins procedure** A general strategy for the analysis of time series based on the use of ARIMA models or, for seasonal data, SARIMA models. The procedure was set out by Box and Jenkins in their classic 1970 book ([Box et al., 2008](#)). The first stage consists of removing trends or cycles from the data. An appropriate type of model must then be identified and its parameters estimated. The estimated model is then compared with the original data and adjustments are made if necessary.

**deseasonalize** To remove regular seasonal fluctuations from a time series for the purposes of analysis (for example, to estimate an underlying trend).

---

1The current edition is [Box et al., 2008](#).
GARCH (generalized autoregressive conditional heteroskedasticity) GARCH (Bollerslev 1986) generalizes the ARCH model. Today’s conditional variance is a function of past squared unexpected returns and its own past values. The model is an infinite weighted average of all past squared forecast errors, with weights that are constrained to be geometrically declining. GARCH is an ARMA(p,q) process in the variance.

Holt-Winters forecasting An application of exponential smoothing to a time series that displays a trend and seasonality.

MA (moving average) models Models for a time series with constant mean (taken as 0) where the next point is dependent on the previous n errors: MA(n).

serial correlation See autocorrelation.

trend If the mean of a time series changes steadily over time then it is said to exhibit a trend.

unit root In autoregressive models in econometrics, a unit root is present if $y_t = y_{t-1} + c + \epsilon_{t-1}$. 
## Appendix D

### Key Articles on the Efficient Market Hypothesis

<table>
<thead>
<tr>
<th>Feature</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price $S$, time $t$, $dS \propto \sqrt{t}$</td>
<td>Regnault (1863), Osborne (1959)</td>
</tr>
<tr>
<td>Martingale</td>
<td>Bachelier (1900), Samuelson (1965)</td>
</tr>
<tr>
<td>Leptokurtosis</td>
<td>Mitchell (1915), Mitchell (1921), Olivier (1926), Mills (1927), Osborne (1959), Larson (1960), Alexander (1961)</td>
</tr>
<tr>
<td>Nonstationarity</td>
<td>Kendall (1953), Houthakker (1961), Osborne (1962)</td>
</tr>
<tr>
<td>Distribution of price changes is log normal</td>
<td>Osborne (1959)</td>
</tr>
<tr>
<td>Non-linear</td>
<td>Alexander (1961), Houthakker (1961)</td>
</tr>
<tr>
<td>Scaling</td>
<td>Mandelbrot (1962), Mandelbrot (1963)</td>
</tr>
<tr>
<td>Joint hypothesis problem</td>
<td>Fama (1970)</td>
</tr>
<tr>
<td>EMH neither implies nor is implied by a random walk</td>
<td>LeRoy (1973), Lucas (1978)</td>
</tr>
<tr>
<td>Not risk neutral implies submartingale</td>
<td>Cox and Ross (1976), Lucas (1978), Harrison and Kreps (1979)</td>
</tr>
<tr>
<td>Different sorts of efficiency</td>
<td>Stiglitz (1981)</td>
</tr>
</tbody>
</table>

Table D.1: *Key articles on the EMH*
Appendix E

Runs Tests Source Code

'Runs Test

Option Base 1
Option Explicit

Function ExpectedNoTotalRuns(Up As Long, Down As Long, Total As Long)
    If Application.IsNumber(Up) And Application.IsNumber(Down) And Application.IsNumber(Total) And Total <> 0 Then
        ExpectedNoTotalRuns = (2 * Up * Down / Total) + 1
    Else
        ExpectedNoTotalRuns = "undefined"
    End If
End Function

Function SDTotalRuns(Up As Long, Down As Long, Total As Long)
    Dim part1 As Double
    Dim part2 As Double
    If Application.IsNumber(Up) And Application.IsNumber(Down) And Application.IsNumber(Total) And Total <> 0 And Total <> 1 Then
        part1 = 2 * Up * Down / Total ^ 2 'split to avoid overflow
        part2 = (2 * Up * Down - Total) / (Total - 1)
        SDTotalRuns = Sqr(part1 * part2)
    Else
        SDTotalRuns = "undefined"
    End If
End Function

Function ExpectedNoRuns(I As Long, N As Long)
    Dim ERunsUp As Double
    Dim Den As Double
    If Application.IsNumber(I) And Application.IsNumber(N) Then
        If I <= N - 2 Then
            ERunsUp = N * ((1 ^ 2 + 3 * I + 1) - (1 ^ 3 + 3 * I ^ 2 - 1 - 4))
            Den = Application.Fact(I + 3)
            ExpectedNoRuns = ERunsUp / Den
        End If
    End If
If I = N − 1 Then
    If Application.IsNumber(Application.Fact(N)) And Application.Fact(N) <> 0 Then
        ExpectedNoRuns = 1 / Application.Fact(N)
    Else
        ExpectedNoRuns = "undefined"
    End If
End If

If I >= N Then
    ExpectedNoRuns = 0
End If
Else
    ExpectedNoRuns = "undefined"
End If
End Function

Function SDNoRuns(I As Long, N As Long) As Double
    Dim Arg As Double
    Dim SNRTL As Double

    'Constants from code written by James J. Filliben, National Bureau Of Standards, via
    'Alan Heckert, National Institute of Standards and Technology

    Dim C1(15) As Double
    C1(1) = 0.4236111111
    C1(2) = 0.1126675485
    C1(3) = 0.04191688713
    C1(4) = 0.01076912487
    C1(5) = 0.002003959238
    C1(6) = 0.0003023235799
    C1(7) = 0.00003911555473
    C1(8) = 0.000004459038843
    C1(9) = 0.000000455110521
    C1(10) = 4.207466837E−08
    C1(11) = 3.555930927E−09
    C1(12) = 2.768273257E−10
    C1(13) = 1.997821524E−11
    C1(14) = 1.343876568E−12
    C1(15) = 8.465610177E−14

    Dim C2(15) As Double
    C2(1) = −0.4819444444
    C2(2) = −0.1628284832
    C2(3) = −0.0969069649
    C2(4) = −0.03778106786
    C2(5) = −0.009289228716
    C2(6) = −0.001724429252
    C2(7) = −0.0002638557888
    C2(8) = −0.00003466965096
    C2(9) = −0.000004004129153
    C2(10) = −4.130382587E−07
    C2(11) = −3.851876069E−08
C2(12) = −3.279103786E-09
C2(13) = −2.568491117E-10
C2(14) = −1.863433868E-11
C2(15) = −1.259220466E-12

If Application.IsNumber(I) And Application.IsNumber(N) Then
    Arg = C1(I) * N + C2(I)
    SNRTL = 0
    If Arg > 0 Then
        SNRTL = Sqr(Arg)
    End If
    SDNoRuns = Sqr(0.5) * SNRTL
Else
    SDNoRuns = "undefined"
End If
End Function

Function Last(rng As Range)
' Finds last row. Adapted from Ron de Bruin, 5 May 2008.
On Error Resume Next
    Last = rng.Find(What:="" , After:=rng.Cells(1), Lookat:=xlPart, Lookln:=xlFormulas,
          SearchOrder:=xlByRows, SearchDirection:=xlPrevious, MatchCase:=False).Row
On Error GoTo 0
End Function

Sub RunsTest()
    Dim RunsUp() , RunsDown() As Long
    Dim I As Long
    Dim AboveMean, BelowMean As Long
    Dim Up, Down As Long
    Dim lp , fA , lA As Long
    Dim errorA As Boolean
    Dim N As Long
    Dim Data() As Double
    Dim Differences() As Double
    Dim TotalNoOfRuns As Long
    Dim sum , mean As Double
    Dim RunsAboveMean() , RunsBelowMean() As Long

    Worksheets("Runs_test").Cells(1, 3).Font.Bold = True
    Worksheets("Runs_test").Cells(1, 3).Value = "Runs_Test"
    Worksheets("Runs_test").Cells(2, 3).Value = "Martin_Sewell@mvs25@cam.ac.uk>"
    Worksheets("Runs_test").Cells(3, 3).Value = "31_May_2010"
    Worksheets("Runs_test").Cells(8, 3).Value = "Enter returns in Column A"
    Worksheets("Runs_test").Cells(9, 3).Value = "in date order, oldest at the top."
    Worksheets("Runs_test").Cells(10, 3).Value = "'Runs_up' refers to a sequence"
    Worksheets("Runs_test").Cells(11, 3).Value = "of increasing returns such as"
    Worksheets("Runs_test").Cells(12, 3).Value = "−0.2,−0.1,0,0.1,0.2;"
Worksheets("Runs\_test").Cells(13, 3).Value = "'runs\_down' refers to a sequence"
Worksheets("Runs\_test").Cells(14, 3).Value = "of\_decreasing\_returns\_such\_as"
Worksheets("Runs\_test").Cells(15, 3).Value = "0.2, 0.1, 0.0, -0.1, -0.2."
Worksheets("Runs\_test").Cells(16, 3).Value = "You do not need to detrend the data."

'Delete previous results'
Worksheets("Runs\_test").Range("E:O").ClearContents
Worksheets("Runs\_test").Range("C18").ClearContents
Worksheets("Runs\_test").Cells(1, 5).Font.Italic = True
Worksheets("Runs\_test").Cells(1, 5) = "Runs above and below the mean"
Worksheets("Runs\_test").Cells(2, 5) = "Total"
Worksheets("Runs\_test").Cells(2, 6) = "Expected no."
Worksheets("Runs\_test").Cells(2, 7) = "Standard deviation"
Worksheets("Runs\_test").Cells(2, 8) = "z-score"
Worksheets("Runs\_test").Cells(5, 5).Font.Italic = True
Worksheets("Runs\_test").Cells(5, 5) = "Runs\_up"
Worksheets("Runs\_test").Cells(6, 5) = "Length of run"
Worksheets("Runs\_test").Cells(6, 6) = "Number"
Worksheets("Runs\_test").Cells(6, 7) = "Expected no."
Worksheets("Runs\_test").Cells(6, 8) = "Standard deviation"
Worksheets("Runs\_test").Cells(6, 9) = "z-score"
Worksheets("Runs\_test").Cells(5, 11).Font.Italic = True
Worksheets("Runs\_test").Cells(5, 11) = "Runs\_down"
Worksheets("Runs\_test").Cells(6, 11) = "Length of run"
Worksheets("Runs\_test").Cells(6, 12) = "Number"
Worksheets("Runs\_test").Cells(6, 13) = "Expected no."
Worksheets("Runs\_test").Cells(6, 14) = "Standard deviation"
Worksheets("Runs\_test").Cells(6, 15) = "z-score"

'Find last possible number'
lp = Last(Worksheets("Runs\_test").Cells)

'Column A'
'Find first and last number'
FA = 0
For I = 1 To lp
    If Application.IsNumber(Worksheets("Runs\_test").Cells(I, 1)) And FA = 0 Then
        FA = I
    End If
    If Application.IsNumber(Worksheets("Runs\_test").Cells(I, 1)) Then
        I = 1
    End If
Next
If FA = 0 And lA = 0 Then
    N = 0
Else
    N = I - FA + 1
End If
If \( N > 1 \) Then

\[
\text{errorA} = \text{False}
\]

'Check that there are no gaps

For \( I = fA \) To \( lA \)
    If Not Application.IsNumber(Worksheets("Runs_test").Cells(I, 1)) Then
        errorA = True
    End If
Next

If errorA = True Then

Worksheets("Runs_test").Cells(18, 3) = "Column_A_must_contain_numbers_with_no_gaps."
Else

ReDim Data(N)
For \( I = 1 \) To \( N \)
    Data(I) = Worksheets("Runs_test").Cells(I, 1)
Next

ReDim Differences(N - 1)
For \( I = 1 \) To \( N - 1 \)
    Differences(I) = Data(I + 1) - Data(I)
Next

ReDim RunsUp(N - 1)
ReDim RunsDown(N - 1)
ReDim RunsAboveMean(N)
ReDim RunsBelowMean(N)

'Used to calculate expectation and standard deviation of number of runs

sum = 0
For \( I = 1 \) To \( N \)
    sum = sum + Data(I)
Next

mean = sum / \( N \)

'Calculate runs above and below the mean

For \( I = 1 \) To \( N \)
    RunsAboveMean(I) = 0
    RunsBelowMean(I) = 0
Next

Up = 0
Down = 0
For \( I = 1 \) To \( N \)
    If Data(I) = mean And Up \geq 1 Then
        Up = Up + 1
    End If
    If Data(I) = mean And Down \geq 1 Then
        Down = Down + 1
    End If
    If Data(I) = mean And Up = 0 And Down = 0 Then
        Up = Up + 1
End If
If Data(I) > mean And Down >= 1 Then
   RunsBelowMean(Down) = RunsBelowMean(Down) + 1
End If
If Data(I) > mean Then
   Down = 0
   Up = Up + 1
End If
If Data(I) < mean And Up >= 1 Then
   RunsAboveMean(Up) = RunsAboveMean(Up) + 1
End If
If Data(I) < mean Then
   Up = 0
   Down = Down + 1
End If
If I = N And Down >= 1 Then
   RunsBelowMean(Down) = RunsBelowMean(Down) + 1
End If
If I = N And Up >= 1 Then
   RunsAboveMean(Up) = RunsAboveMean(Up) + 1
End If
Next I
AboveMean = 0
BelowMean = 0
TotalNoOfRuns = 0
For I = 1 To N
   AboveMean = AboveMean + I * RunsAboveMean(I)
   BelowMean = BelowMean + I * RunsBelowMean(I)
   TotalNoOfRuns = TotalNoOfRuns + RunsAboveMean(I) + RunsBelowMean(I)
Next
'Calculate runs up and down
For I = 1 To N - 1
   RunsUp(I) = 0
   RunsDown(I) = 0
Next
Up = 0
Down = 0
For I = 1 To N - 1
   If Differences(I) = 0 And Up >= 1 Then 'if no change after an up, count as an
      Up = Up + 1
   End If
   If Differences(I) = 0 And Down >= 1 Then 'if no change after a down, count as a
      Down = Down + 1
   End If
If Differences(1) = 0 And Up = 0 And Down = 0 Then 'if no change at start, count as an up
Up = Up + 1
End If
If Differences(1) > 0 And Down >= 1 Then 'a run of downs followed by an up
RunsDown(Down) = RunsDown(Down) + 1
End If
If Differences(1) > 0 Then 'up
Down = 0
Up = Up + 1
End If
If Differences(1) < 0 And Up >= 1 Then 'a run of ups followed by a down
RunsUp(Up) = RunsUp(Up) + 1
End If
If Differences(1) < 0 Then 'down
Up = 0
Down = Down + 1
End If
If I = N - 1 And Down >= 1 Then 'last difference is down
RunsDown(Down) = RunsDown(Down) + 1
End If
If I = N - 1 And Up >= 1 Then 'last difference is up
RunsUp(Up) = RunsUp(Up) + 1
End If
Next I

'Output results
'Total number of runs
Worksheets("Runs_test").Cells(3, 5) = TotalNoOfRuns
Worksheets("Runs_test").Cells(3, 6) = ExpectedNoTotalRuns(AboveMean, BelowMean, N)
Worksheets("Runs_test").Cells(3, 7) = SDTotalRuns(AboveMean, BelowMean, N)
If SDTotalRuns(AboveMean, BelowMean, N) > 0 Then
    Worksheets("Runs_test").Cells(3, 8) = (TotalNoOfRuns - ExpectedNoTotalRuns(AboveMean, BelowMean, N)) / SDTotalRuns(AboveMean, BelowMean, N)
Else
    Worksheets("Runs_test").Cells(3, 8) = "undefined"
End If

'Runs up and runs down
For I = 1 To N - 1
    If I <= 30 Then
        Worksheets("Runs_test").Cells(I + 6, 5) = 1
        Worksheets("Runs_test").Cells(I + 6, 6) = RunsUp(I)
        Worksheets("Runs_test").Cells(I + 6, 7) = ExpectedNoRuns(I, N)
    If I <= 15 Then
        Worksheets("Runs_test").Cells(I + 6, 8) = SDNoRuns(I, N)
        If SDNoRuns(I, N) > 0 Then
Worksheets("Runs_test").Cells(1 + 6, 9) = (RunsUp(I) − ExpectedNoRuns(I, N)) / SDNoRuns(I, N) 'z-score
Else

Worksheets("Runs_test").Cells(1 + 6, 9) = "undefined"
End If
End If

Worksheets("Runs_test").Cells(1 + 6, 11) = I
Worksheets("Runs_test").Cells(1 + 6, 12) = RunsDown(I)
Worksheets("Runs_test").Cells(1 + 6, 13) = ExpectedNoRuns(I, N)
If I <= 15 Then

Worksheets("Runs_test").Cells(1 + 6, 14) = SDNoRuns(I, N)
End If
If I <= 15 Then

If SDNoRuns(I, N) > 0 Then

Worksheets("Runs_test").Cells(1 + 6, 15) = (RunsDown(I) − ExpectedNoRuns
(I, N)) / SDNoRuns(I, N) 'z-score
Else

Worksheets("Runs_test").Cells(1 + 6, 15) = "undefined"
End If
End If
End If
Next
End If
End If
End If
End Sub
Appendix F

Runs Tests on DJIA Returns

In the following tables and figures, an increasing run refers to a subsequence of increasing returns such as -0.2, -0.1, 0, 0.1, 0.2, whilst a decreasing run refers to a subsequence of decreasing returns such as 0.2, 0.1, 0, -0.1, -0.2. * indicates statistical significance at the 10% level ** 5%, *** 1%, **** 0.5% and ***** 0.1%.

<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3947</td>
<td>4367.5417</td>
<td>66.6337</td>
<td>-6.3112</td>
<td>0.0000***</td>
</tr>
<tr>
<td>2</td>
<td>1995</td>
<td>1921.5833</td>
<td>34.3642</td>
<td>2.1364</td>
<td>0.0163**</td>
</tr>
<tr>
<td>3</td>
<td>556</td>
<td>553.1514</td>
<td>20.9601</td>
<td>0.1359</td>
<td>0.4459</td>
</tr>
<tr>
<td>4</td>
<td>122</td>
<td>120.6056</td>
<td>10.6237</td>
<td>0.1313</td>
<td>0.4478</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>21.3128</td>
<td>4.5827</td>
<td>-1.3775</td>
<td>0.0842*</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3.1765</td>
<td>1.7799</td>
<td>0.4626</td>
<td>0.3218</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.4100</td>
<td>0.6402</td>
<td>-0.6405</td>
<td>0.2609</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0467</td>
<td>0.2162</td>
<td>-0.2162</td>
<td>0.4144</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0048</td>
<td>0.0691</td>
<td>-0.0691</td>
<td>0.4725</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0004</td>
<td>0.0210</td>
<td>-0.0210</td>
<td>0.4916</td>
</tr>
</tbody>
</table>

Table F.1: DJIA daily returns: increasing
<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3701</td>
<td>4367.5417</td>
<td>66.6337</td>
<td>-10.0031</td>
<td>0.0000***</td>
</tr>
<tr>
<td>2</td>
<td>2027</td>
<td>1921.5833</td>
<td>34.3642</td>
<td>3.0676</td>
<td>0.0011**</td>
</tr>
<tr>
<td>3</td>
<td>673</td>
<td>553.1514</td>
<td>20.9601</td>
<td>5.7180</td>
<td>0.0000****</td>
</tr>
<tr>
<td>4</td>
<td>193</td>
<td>120.6056</td>
<td>10.6237</td>
<td>6.8144</td>
<td>0.0000****</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>21.3128</td>
<td>4.5827</td>
<td>4.0778</td>
<td>0.0000****</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3.1765</td>
<td>1.7799</td>
<td>-0.0992</td>
<td>0.9385</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.4100</td>
<td>0.6402</td>
<td>0.9215</td>
<td>0.1784</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0467</td>
<td>0.2162</td>
<td>-0.2162</td>
<td>0.1784</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0048</td>
<td>0.0691</td>
<td>-0.0691</td>
<td>0.1784</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0004</td>
<td>0.0210</td>
<td>-0.0210</td>
<td>0.1784</td>
</tr>
</tbody>
</table>

Table F.2: DJIA daily returns: decreasing

<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>926</td>
<td>885.88</td>
<td>30.01</td>
<td>1.34</td>
<td>0.0906*</td>
</tr>
<tr>
<td>2</td>
<td>369</td>
<td>389.65</td>
<td>15.47</td>
<td>-1.33</td>
<td>0.0910*</td>
</tr>
<tr>
<td>3</td>
<td>97</td>
<td>112.14</td>
<td>9.44</td>
<td>-1.60</td>
<td>0.0543*</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>24.45</td>
<td>4.78</td>
<td>0.12</td>
<td>0.4538</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4.32</td>
<td>2.06</td>
<td>-2.09</td>
<td>0.0181**</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.64</td>
<td>0.80</td>
<td>-0.80</td>
<td>0.2109</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.08</td>
<td>0.29</td>
<td>-0.29</td>
<td>0.3866</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.01</td>
<td>0.10</td>
<td>-0.10</td>
<td>0.4613</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.00</td>
<td>0.03</td>
<td>-0.03</td>
<td>0.4876</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.4962</td>
</tr>
</tbody>
</table>

Table F.3: DJIA weekly returns: increasing
<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>958</td>
<td>906.9167</td>
<td>30.3603</td>
<td>1.6826</td>
<td>0.0462**</td>
</tr>
<tr>
<td>2</td>
<td>373</td>
<td>398.9083</td>
<td>15.6569</td>
<td>-1.6548</td>
<td>0.0490**</td>
</tr>
<tr>
<td>3</td>
<td>99</td>
<td>114.8056</td>
<td>9.5490</td>
<td>-1.6552</td>
<td>0.0489**</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>25.0264</td>
<td>4.8394</td>
<td>-0.0055</td>
<td>0.4978</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4.4217</td>
<td>2.0873</td>
<td>-2.1183</td>
<td>0.0171**</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.6589</td>
<td>0.8106</td>
<td>-0.8128</td>
<td>0.2082</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.0850</td>
<td>0.2916</td>
<td>-0.2917</td>
<td>0.3853</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0097</td>
<td>0.0984</td>
<td>-0.0984</td>
<td>0.4608</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0010</td>
<td>0.0314</td>
<td>-0.0314</td>
<td>0.4875</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0001</td>
<td>0.0096</td>
<td>-0.0096</td>
<td>0.4962</td>
</tr>
</tbody>
</table>

Table F.4: DJIA weekly returns: decreasing

<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>891</td>
<td>906.9167</td>
<td>30.3603</td>
<td>-0.5243</td>
<td>0.3000</td>
</tr>
<tr>
<td>2</td>
<td>387</td>
<td>398.9083</td>
<td>15.6569</td>
<td>-0.7606</td>
<td>0.2235</td>
</tr>
<tr>
<td>3</td>
<td>131</td>
<td>114.8056</td>
<td>9.5490</td>
<td>1.6959</td>
<td>0.0450**</td>
</tr>
<tr>
<td>4</td>
<td>38</td>
<td>25.0264</td>
<td>4.8394</td>
<td>2.6808</td>
<td>0.0037****</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>4.4217</td>
<td>2.0873</td>
<td>1.2352</td>
<td>0.1084</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.6589</td>
<td>0.8106</td>
<td>0.4208</td>
<td>0.3370</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.0850</td>
<td>0.2916</td>
<td>-0.2917</td>
<td>0.3853</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0097</td>
<td>0.0984</td>
<td>-0.0984</td>
<td>0.4608</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0010</td>
<td>0.0314</td>
<td>-0.0314</td>
<td>0.4875</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0001</td>
<td>0.0096</td>
<td>-0.0096</td>
<td>0.4962</td>
</tr>
</tbody>
</table>

Table F.5: DJIA monthly returns: increasing
<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>208.3750</td>
<td>14.5453</td>
<td>-0.5758</td>
<td>0.2824</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>91.5500</td>
<td>7.5002</td>
<td>2.4599</td>
<td>0.0069***</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>26.3236</td>
<td>4.5727</td>
<td>-0.2895</td>
<td>0.3861</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5.7333</td>
<td>2.3164</td>
<td>-0.7483</td>
<td>0.2271</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1.0121</td>
<td>0.9987</td>
<td>-1.0135</td>
<td>0.1554</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.1507</td>
<td>0.3877</td>
<td>-0.3887</td>
<td>0.3487</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.0194</td>
<td>0.1394</td>
<td>-0.1394</td>
<td>0.4446</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0022</td>
<td>0.0470</td>
<td>-0.0470</td>
<td>0.4812</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0002</td>
<td>0.0150</td>
<td>-0.0150</td>
<td>0.4940</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0000</td>
<td>0.0046</td>
<td>-0.0046</td>
<td>0.4982</td>
</tr>
</tbody>
</table>

Table F.6: DJIA monthly returns: decreasing

<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>17.3333</td>
<td>4.1640</td>
<td>0.1601</td>
<td>0.4364</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7.4917</td>
<td>2.1434</td>
<td>-1.1625</td>
<td>0.1225</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.1250</td>
<td>1.3004</td>
<td>-0.0961</td>
<td>0.4617</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.4569</td>
<td>0.6542</td>
<td>-0.6984</td>
<td>0.2425</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.0797</td>
<td>0.2802</td>
<td>-0.2843</td>
<td>0.3881</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.0117</td>
<td>0.1081</td>
<td>-0.1084</td>
<td>0.4568</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.0015</td>
<td>0.0386</td>
<td>-0.0386</td>
<td>0.4846</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0002</td>
<td>0.0130</td>
<td>-0.0130</td>
<td>0.4948</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0000</td>
<td>0.0041</td>
<td>-0.0041</td>
<td>0.4984</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0000</td>
<td>0.0012</td>
<td>-0.0012</td>
<td>0.4995</td>
</tr>
</tbody>
</table>

Table F.7: DJIA annual returns: increasing
<table>
<thead>
<tr>
<th>Length of run</th>
<th>Number of runs</th>
<th>Expectation</th>
<th>Standard deviation</th>
<th>z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>17.3333</td>
<td>4.1640</td>
<td>-1.7611</td>
<td>0.0391**</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>7.4917</td>
<td>2.1434</td>
<td>1.1702</td>
<td>0.1210</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2.1250</td>
<td>1.3004</td>
<td>2.9798</td>
<td>0.0014****</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.4569</td>
<td>0.6542</td>
<td>-0.6984</td>
<td>0.2425</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.0797</td>
<td>0.2802</td>
<td>-0.2843</td>
<td>0.3881</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.0117</td>
<td>0.1081</td>
<td>-0.1084</td>
<td>0.4568</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.0015</td>
<td>0.0386</td>
<td>-0.0386</td>
<td>0.4846</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.0002</td>
<td>0.0130</td>
<td>-0.0130</td>
<td>0.4948</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.0000</td>
<td>0.0041</td>
<td>-0.0041</td>
<td>0.4984</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0000</td>
<td>0.0012</td>
<td>-0.0012</td>
<td>0.4995</td>
</tr>
</tbody>
</table>

*Table F.8: DJIA annual returns: decreasing*
Appendix G

Rescaled Range Analysis Source Code

'This program calculates an estimate of the Hurst coefficient.

Option Base 1
Option Explicit

Sub Hurst ()
Dim Data (), Array1 (), Array2 (), Result () As Double
Dim NoOfDataPoints, i, j, counter, N, NoOfPlottedPoints, NoOfPeriods, PeriodNo,
PlottedPointNo As Integer
Dim logten, totalR, totalS, Summ, SumSquared, Mean, Maxi, Mini, R, S, RS, Sumx, Sumxy,
Sumxx, H As Double
Dim curCell As Object

logten = Log(10)
Worksheets("Data").Range("B1").Value = "Total="
Worksheets("Data").Range("B3").Value = "H=
Worksheets("Data").Range("G1").Value = "Put data (e.g., log returns) in column A."
Worksheets("Data").Range("G2").Value = "The input sequence should be stationary with mean zero."
Worksheets("Data").Range("G3").Value = "So if analysing financial data, the input data must be detrended (zero mean)."
Worksheets("Data").Range("G4").Value = "The values of the Hurst exponent range between 0 and 1."
Worksheets("Data").Range("G5").Value = "0 < H < 0.5"
Worksheets("Data").Range("G6").Value = "H = 0.5"
Worksheets("Data").Range("G7").Value = "0.5 < H < 1"
Worksheets("Data").Range("H5").Value = "anti-persistence"
Worksheets("Data").Range("H6").Value = "random walk"
Worksheets("Data").Range("H7").Value = "persistence"

'Delete any previous results
Worksheets("Data").Range("C3").Value = Null
Worksheets("Data").Range("D:D").Value = Null
Worksheets("Data").Range("E:E").Value = Null

'Get and output total number of data points
NoOfDataPoints = Application.Count(Range("Data"))
Worksheets("Data").Range("C1").Value = NoOfDataPoints

If NoOfDataPoints > 3 Then
    ReDim Data(NoOfDataPoints)

' Get data, ignoring any spaces
i = 1
counter = 1
Do While counter <= NoOfDataPoints
    Set curCell = Worksheets("Data").Cells(i, 1)
    If Application.WorksheetFunction.IsNumber(curCell.Value) Then
        Data(counter) = curCell.Value
        counter = counter + 1
    End If
    i = i + 1
Loop
NoOfPlottedPoints = NoOfDataPoints - 2
ReDim Result(NoOfPlottedPoints, 2)

' Begin main loop
For N = 3 To NoOfDataPoints
    totalR = 0
    totalS = 0
    NoOfPeriods = NoOfDataPoints - N + 1
    For PeriodNo = 1 To NoOfPeriods
        ReDim Array1(N)
        ReDim Array2(N)
        For i = 1 To N
            Array1(i) = Data((PeriodNo - 1) + i)
            Array2(i) = 0
        Next i
        Summ = 0
        SumSquared = 0
        For i = 1 To N
            Summ = Summ + Array1(i)
            SumSquared = SumSquared + ((Array1(i)) * (Array1(i)))
        Next i
        Mean = Summ / N
        S = Sqr((SumSquared - (Summ * Summ) / N) / N)
        For i = 1 To N
            Array1(i) = Array1(i) - Mean
        Next i
        For i = 1 To N
            For j = 1 To i
                Array2(i) = Array2(i) + Array1(j)
            Next j
        Next i
        Maxi = Array2(1)
Mini = Array2(1)

For i = 1 To N
    If Array2(i) > Maxi Then Maxi = Array2(i)
    If Array2(i) < Mini Then Mini = Array2(i)
Next i

R = Maxi - Mini
totalR = totalR + R
totalS = totalS + S

Next PeriodNo

R = totalR / NoOfPeriods
S = totalS / NoOfPeriods
RS = R / S

PlotPointNo = N - 2

Result(PplotPointNo, 1) = (Log(N)) / logten
Result(PplotPointNo, 2) = (Log(RS)) / logten

Next N

Sumx = 0
Sumy = 0
Sumxy = 0
Sumxx = 0

Worksheets("Data").Cells(1, 4).Value = "Log(Time)"
Worksheets("Data").Cells(1, 5).Value = "Log(R/S)"

For i = 1 To NoOfPlotPoint
    Worksheets("Data").Cells(i + 1, 4).Value = Result(i, 1)
    Worksheets("Data").Cells(i + 1, 5).Value = Result(i, 2)
    Sumx = Sumx + Result(i, 1)
    Sumy = Sumy + Result(i, 2)
    Sumxy = Sumxy + (Result(i, 1) * (Result(i, 2))
    Sumxx = Sumxx + (Result(i, 1)) * (Result(i, 1))
Next i

'Calculate Hurst coefficient

H = (Sumxy - ((Sumx * Sumy) / NoOfPlotPoints)) / (Sumxx - ((Sumx * Sumx) / NoOfPlotPoints))

Worksheets("Data").Range("C3").Value = H

Else
    Worksheets("Data").Range("C3").Value = "undefined"
End If

End Sub
Appendix H

Technical Analysis Taxonomy

The taxonomy below is the syllabus of the Society of Technical Analyst’s Diploma.


3. Candle charts and candle patterns.

4. Point and figure charts. Construction, scale, box reversal, objective counting. Advantages and disadvantages compared to other types of chart.

5. Dow Theory

6. Chart patterns, e.g. triangles, flags, pennants, diamonds, broadening patterns (megaphones), wedges.

7. Reversal patterns and how to identify/anticipate them. Rounding tops and bottoms, head and shoulders, spikes, double/treble/multiple tops and bottoms.


9. Consolidation - how and why it occurs. Breakouts and how to recognise them.

10. Corrections: when and how far.

11. Support and resistance. The various chart points and facets that can act as such.

12. Basic elements of Gann theory.

13. Basic elements of Elliott wave theory.

http://www.sta-uk.org/sta_diploma.html#syllabus
14. Fibonacci series, fan lines, arcs and time zones.


16. Relative performance and how to interpret relative strength charts.

17. Momentum indicators and oscillators including:

18. Rate of change - Welles Wilder’s RSI - Stochastics (%K & D)

19. Moving Average Convergence Divergence (MACD) & MACD histogram

20. Directional Movement Indicator - Parabolics - Commodity Channel Index

21. Volume signals and indicators, including On-Balance Volume, Volume Accumulator etc. Open interest.

22. Breadth indicators.

23. Sentiment indicators and contrary opinion.


25. Investor psychology - individual and group.

Appendix I

Cumulative Prospect Theory Certainty
Equivalent Source Code

*Cumulative Prospect Theory*

Option Base 0
Option Explicit

Public Class Class1
    Public out As Double
    Public pro As Double
    Public pos As Double
    Public wei As Double
End Class

Function Upos (a As Double, alpha As Double) As Double
    If alpha > 0 Then
        Upos = a ^ alpha
    ElseIf alpha = 0 Then
        Upos = Log(a)
    Else
        Upos = 1 - (a + 1) ^ alpha
    End If
End Function

Function Uneg (b As Double, beta As Double, lambda As Double) As Double
    If beta > 0 Then
        Uneg = (-1) * lambda * (((-1) * b) ^ beta)
    ElseIf beta = 0 Then
        Uneg = (-1) * lambda * Log((-1) * b)
    Else
        Uneg = (-1) * lambda * (1 - (((-1) * b + 1) ^ beta))
    End If
End Function
Function Utility(y As Double, alpha As Double, beta As Double, lambda As Double) As Double
  If (y > 0) Then
    Utility = Upos(y, alpha)
  ElseIf y < 0 Then
    Utility = Uneg(y, beta, lambda)
  End If
End Function

Function CEpos(c As Double, alpha As Double) As Double
  If alpha > 0 Then
    CEpos = c ^ (1 / alpha)
  ElseIf alpha = 0 Then
    CEpos = Exp(c)
  Else
    CEpos = ((1 - c) ^ (1 / alpha)) - 1
  End If
End Function

Function CEneg(D As Double, beta As Double, lambda As Double) As Double
  If beta > 0 Then
    CEneg = (-1) * (((-1) * D / lambda) ^ (1 / beta))
  ElseIf beta = 0 Then
    CEneg = (-1) * Exp((-1) * D / lambda)
  Else
    CEneg = 1 - (1 + D / lambda) ^ (1 / beta)
  End If
End Function

Function CertaintyEquivalent(x As Double, alpha As Double, beta As Double, lambda As Double) As Double
  If x > 0 Then
    CertaintyEquivalent = CEpos(x, alpha)
  ElseIf x < 0 Then
    CertaintyEquivalent = CEneg(x, beta, lambda)
  End If
End Function

Function Wplus(z As Double, gamma As Double) As Double
  On Error Resume Next
  Wplus = (z ^ gamma) / (((z ^ gamma) + ((1 - z) ^ gamma)) ^ (1 / gamma))
End Function

Function Wminus(T As Double, delta As Double) As Double
  On Error Resume Next
  Wminus = (T ^ delta) / (((T ^ delta) + ((1 - T) ^ delta)) ^ (1 / delta))
End Function
Function sumoutcomes(a As Integer, b As Integer, Outcomes() As Class1)
        Dim counter As Integer
        Dim sum As Double
        sum = 0
        For counter = a To b Step 1
            sum = sum + Outcomes(counter).pro
        Next counter
        sumoutcomes = sum
End Function

'Mean
Function Mean(data() As Double, p() As Double) As Variant
        Dim n As Integer
        Dim i As Integer
        Dim sum As Double
        n = UBound(data)
        If n > 0 Then
            sum = 0
            For i = 0 To n - 1
                sum = sum + data(i) * p(i)
            Next
            Mean = sum
        Else
            Mean = "undefined"
        End If
End Function

'Standard deviation
Function SD(data() As Double, p() As Double) As Variant
        Dim n As Integer
        Dim i As Integer
        Dim M As Double
        Dim moment2 As Double
        n = UBound(data)
        If n > 0 Then
            M = Mean(data(), p())
            moment2 = 0
            For i = 0 To n - 1
                moment2 = moment2 + (((data(i) - M) ^ 2) * p(i))
            Next
            SD = Sqr(moment2)
        Else
            SD = "undefined"
        End If
End Function

'Skewness
Function Skewness(data() As Double, p() As Double) As Variant
Dim n As Long
Dim i As Integer
Dim moment3 As Double
Dim M As Double
Dim S As Double
n = UBound(data)
If n > 0 Then
    moment3 = 0
    M = Mean(data(), p())
    S = SD(data(), p())
    If Application.IsNumber(M) And Application.IsNumber(S) And S <> 0 Then
        For i = 0 To n - 1
            moment3 = moment3 + ((data(i) - M) ^ 3) * p(i)
        Next i
        Skewness = moment3 / (S ^ 3)
    Else
        Skewness = "undefined"
    End If
Else
    Skewness = "undefined"
End If
End Function

'Kurtosis
Function Kurtosis(data() As Double, p() As Double) As Variant
Dim n As Long
Dim i As Integer
Dim M As Double
Dim S As Double
Dim moment4 As Double
n = UBound(data)
If n > 0 Then
    M = Mean(data(), p())
    S = SD(data(), p())
    moment4 = 0
    If Application.IsNumber(M) And Application.IsNumber(S) And S <> 0 Then
        For i = 0 To n - 1
            moment4 = moment4 + ((data(i) - M) ^ 4) * p(i)
        Next i
        Kurtosis = moment4 / (S ^ 4) - 3
    Else
        Kurtosis = "undefined"
    End If
Else
    Kurtosis = "undefined"
End If
End Function
Function Last(choice As Integer, rng As Range)
' http://www.rondebruin.nl/last.htm
' Ron de Bruin, 20 Feb 2007
' 1 = last row
' 2 = last column
' 3 = last cell
Dim lrw As Long
Dim lcol As Integer
Select Case choice
Case 1:
On Error Resume Next
Last = rng.Find(What:=""s", _
               After:=rng.Cells(1), _
               Lookat:=xlPart, _
               LookIn:=xlFormulas, _
               SearchOrder:=xlByRows, _
               SearchDirection:=xlPrevious, _
               MatchCase:=False).Row
On Error GoTo 0
Case 2:
On Error Resume Next
Last = rng.Find(What:=""s", _
               After:=rng.Cells(1), _
               Lookat:=xlPart, _
               LookIn:=xlFormulas, _
               SearchOrder:=xlByColumns, _
               SearchDirection:=xlPrevious, _
               MatchCase:=False).Column
On Error GoTo 0
Case 3:
On Error Resume Next
lrw = rng.Find(What:=""s", _
               After:=rng.Cells(1), _
               Lookat:=xlPart, _
               LookIn:=xlFormulas, _
               SearchOrder:=xlByRows, _
               SearchDirection:=xlPrevious, _
               MatchCase:=False).Row
On Error GoTo 0
On Error Resume Next
lcol = rng.Find(What:=""s", _
               After:=rng.Cells(1), _
               Lookat:=xlPart, _
               LookIn:=xlFormulas, _
               SearchOrder:=xlByColumns, _
               SearchDirection:=xlPrevious, _
               MatchCase:=False).Column
On Error GoTo 0
On Error Resume Next
Last = Cells(lr, lcol).Address(False, False)
If Err.Number > 0 Then
    Last = rng.Cells(1).Address(False, False)
    Err.Clear
End If
On Error GoTo 0
End Select
End Function

Sub Calculate()
Dim maxn, n, counter, arraycounter, i, j, k As Integer
Dim UT() As Double
Dim Outcomes() As Class1
Dim h As Class1
Dim CPTvalue As Double
Dim OutcomeData() As Double
Dim ProbabilityData() As Double
Dim Row() As Integer
Dim SumOfProbabilities As Double
Dim alpha, beta, lambda, gamma, delta As Double
Worksheets("Cumulative_Prospect_Theory").Cells(1, 5).Font.Bold = True
Worksheets("Cumulative_Prospect_Theory").Cells(1, 5).Value = "Cumulative_Prospect_TheoryCalculator"
Worksheets("Cumulative_Prospect_Theory").Cells(2, 5).Value = "Martin Sewell < mvs25@cam.ac.uk >"
Worksheets("Cumulative_Prospect_Theory").Cells(3, 5).Value = "26 October 2010"
Worksheets("Cumulative_Prospect_Theory").Cells(4, 5).Value = "Based on Tversky and Kahneman (1992)"
Worksheets("Cumulative_Prospect_Theory").Cells(1, 1).Value = "Outcome"
Worksheets("Cumulative_Prospect_Theory").Cells(1, 2).Value = "Probability"
Worksheets("Cumulative_Prospect_Theory").Columns(3) = ""
Worksheets("Cumulative_Prospect_Theory").Cells(1, 3).Value = "Decision Weight"
Worksheets("Cumulative_Prospect_Theory").Cells(6, 5).Value = "Power for gains, " & ChrW$(945)
Worksheets("Cumulative_Prospect_Theory").Cells(7, 5).Value = "Power for losses, " & ChrW$(946)
Worksheets("Cumulative_Prospect_Theory").Cells(8, 5).Value = "Loss aversion, " & ChrW$(955)
Worksheets("Cumulative_Prospect_Theory").Cells(9, 5).Value = "Probability weighting parameter for gains, " & ChrW$(947)
Worksheets("Cumulative_Prospect_Theory").Cells(10, 5).Value = "Probability weighting parameter for losses, " & ChrW$(948)
Worksheets("Cumulative_Prospect_Theory").Cells(6, 7).Value = "(0.88 in T&K)"
Worksheets("Cumulative_Prospect_Theory").Cells(7, 7).Value = "(0.88 in T&K)"
Worksheets("Cumulative_Prospect_Theory").Cells(8, 7).Value = "(2.25 in T&K)"
Worksheets("Cumulative_Prospect_Theory").Cells(9, 7).Value = "(0.61 in T&K)"
Worksheets("Cumulative_Prospect_Theory").Cells(10,7).Value = "(0.69 in T&K)"
Worksheets("Cumulative_Prospect_Theory").Cells(15,5).Value = "Number of outcomes"
Worksheets("Cumulative_Prospect_Theory").Cells(15,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(16,5).Value = "Mean"
Worksheets("Cumulative_Prospect_Theory").Cells(16,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(17,5).Value = "Standard deviation"
Worksheets("Cumulative_Prospect_Theory").Cells(17,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(18,5).Value = "Skewness"
Worksheets("Cumulative_Prospect_Theory").Cells(18,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(19,5).Value = "Kurtosis"
Worksheets("Cumulative_Prospect_Theory").Cells(19,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(20,5).Value = "CPT Value"
Worksheets("Cumulative_Prospect_Theory").Cells(20,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(21,5).Value = "Certainty equivalent"
Worksheets("Cumulative_Prospect_Theory").Cells(21,6).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(23,5).Value = ""
Worksheets("Cumulative_Prospect_Theory").Cells(24,5).Value = ""
maxn = Last(1, Worksheets("Cumulative_Prospect_Theory").Columns(1))

'Get number of (valid) data points
n = 0
For counter = 1 To maxn
  If Application.IsNumber(Worksheets("Cumulative_Prospect_Theory").Cells(counter,1).Value) And Application.IsNumber(Worksheets("Cumulative_Prospect_Theory").Cells(counter,2).Value) Then
    n = n + 1
  End If
Next counter
ReDim OutcomeData(n)
ReDim ProbabilityData(n)
ReDim Row(n)

'Having determined the array sizes, parse the data again
arraycounter = 0
For counter = 1 To maxn
  If Application.IsNumber(Worksheets("Cumulative_Prospect_Theory").Cells(counter,1).Value) And Application.IsNumber(Worksheets("Cumulative_Prospect_Theory").Cells(counter,2).Value) Then
    OutcomeData(arraycounter) = Worksheets("Cumulative_Prospect_Theory").Cells(counter,1).Value
    ProbabilityData(arraycounter) = Worksheets("Cumulative_Prospect_Theory").Cells(counter,2).Value
    Row(arraycounter) = counter
    arraycounter = arraycounter + 1
  End If
Next counter

'Read in constants
alpha = Worksheets("CumulativeProspectTheory").Cells(6,6).Value
beta = Worksheets("CumulativeProspectTheory").Cells(7,6).Value
lambda = Worksheets("CumulativeProspectTheory").Cells(8,6).Value
gamma = Worksheets("CumulativeProspectTheory").Cells(9,6).Value
delta = Worksheets("CumulativeProspectTheory").Cells(10,6).Value

n = UBound(OutcomeData)
If n > 0 Then
  ReDim UT(n)
  ReDim Weight(n)
  ReDim Outcomes(n)
End If

SumOfProbabilities = 0
For i = 0 To n - 1
  SumOfProbabilities = SumOfProbabilities + ProbabilityData(i)
Next

If n > 0 And SumOfProbabilities > 0.999 And SumOfProbabilities < 1.001 Then
  For i = 0 To n - 1
    Set Outcomes(i) = New Class1
    Outcomes(i).out = OutcomeData(i)
    Outcomes(i).pro = ProbabilityData(i)
    Outcomes(i).pos = i
    Outcomes(i).wei = 0
  Next
  Set h = New Class1

  'Rank outcomes
  For i = 0 To n - 2
    For j = i + 1 To n - 1
      If Outcomes(i).out < Outcomes(j).out Then
        Set h = Outcomes(i)
        Set Outcomes(i) = Outcomes(j)
        Set Outcomes(j) = h
        Set h = Nothing
      End If
    Next
  Next

  'Apply probability weighting functions for gains and losses
  If Outcomes(0).out >= 0 Then
    Outcomes(0).wei = Wplus(Outcomes(0).pro, gamma)
  Else
    Outcomes(0).wei = 1 - Wminus(1 - Outcomes(0).pro, delta)
  End If
  For i = 1 To n - 2
    If Outcomes(i).out >= 0 Then
      Outcomes(i).wei = Wplus(sumoutcomes(0, i, Outcomes()), gamma) - Wplus(sumoutcomes(0, i - 1, Outcomes()), gamma)
    Else

Outcomes(i).wei = Wminus(sumoutcomes(i, n - 1, Outcomes()), delta) - Wminus(sumoutcomes(i + 1, n - 1, Outcomes()), delta)

End If
Next
If Outcomes(n - 1).out >= 0 Then
  Outcomes(n - 1).wei = 1 - Wplus(1 - Outcomes(n - 1).pro, gamma)
Else
  Outcomes(n - 1).wei = Wminus(Outcomes(n - 1).pro, delta)
End If

' Output weights
For k = 0 To n - 1
  For i = 0 To n - 1
    If Outcomes(k).pos = i Then
      Weight(i) = Outcomes(k).wei
      Worksheets("Cumulative Prospect Theory").Cells(Row(i), 3).Value = Weight(i)
    End If
  Next
Next

'Determine the utility of each outcome (apply the value function)
For i = 0 To n - 1
  UT(i) = Utility(OutcomeData(i), alpha, beta, lambda)
Next

'Calculate CPT value
CPTvalue = 0
For i = 0 To n - 1
  CPTvalue = CPTvalue + Weight(i) * UT(i)
Next
Worksheets("Cumulative Prospect Theory").Cells(15, 6).Value = n
Worksheets("Cumulative Prospect Theory").Cells(16, 6).Value = Mean(OutcomeData(), ProbabilityData())
Worksheets("Cumulative Prospect Theory").Cells(17, 6).Value = SD(OutcomeData(), ProbabilityData())
Worksheets("Cumulative Prospect Theory").Cells(18, 6).Value = Skewness(OutcomeData(), ProbabilityData())
Worksheets("Cumulative Prospect Theory").Cells(19, 6).Value = Kurtosis(OutcomeData(), ProbabilityData())
Worksheets("Cumulative Prospect Theory").Cells(20, 6).Value = CPTvalue
Worksheets("Cumulative Prospect Theory").Cells(21, 6).Value = CertaintyEquivalent(CPTvalue, alpha, beta, lambda)
Else
  If n >= 1 Then
    Worksheets("Cumulative Prospect Theory").Cells(23, 5).Value = "Probabilities_sum to\n  & SumOfProbabilities"
    Worksheets("Cumulative Prospect Theory").Cells(24, 5).Value = "Probabilities_must\n  sum to 1."
  End If
End If
End If
End Sub
Appendix J

Kernel Methods/Support Vector Machines

J.1 Gram Matrix

**Definition 7** Given a set \( S = \{\vec{x}_1, \ldots, \vec{x}_n\} \) of vectors from an inner product space \( X \), the \( n \times n \) matrix \( G \) with entries \( G_{ij} = \langle \vec{x}_i \cdot \vec{x}_j \rangle \) is called the Gram matrix (or kernel matrix) of \( S \).

J.2 Hilbert Space

**Definition 8** A Hilbert space is a Euclidean space which is complete, separable and infinite-dimensional. In other words, a Hilbert space is a set \( H \) of elements \( f, g, \ldots \) of any kind such that

- \( H \) is a Euclidean space, i.e. a real linear space equipped with a scalar product;
- \( H \) is complete with respect to the metric \( \rho(f, g) = \|f - g\| \);
- \( H \) is separable, i.e. \( H \) contains a countable everywhere dense subset;
- \( H \) is infinite-dimensional, i.e., given any positive integer \( n \), \( H \) contains \( n \) linearly independent elements.
Appendix K

Fisher Kernel Source Code

// line numbers refer to Code Fragment 12.4 (page 435) in "Kernel Methods for Pattern Analysis" by John Shawe-Taylor and Nello Cristianini
// use symbols 1, 2, 3, etc.

#include <iostream>
#include <fstream>
#include <sstream>
#include <math.h>
#include <string>

using namespace std;

int main()
{
  int string_length = 10;
  int number_of_states = 5;
  int number_of_symbols = 5;
  int p = number_of_states;
  int n = string_length;
  int a, b;
  double Prob = 0;
  string stringstring;

  ifstream hmmstream("hmmt.txt"); //INPUT: Hidden Markov model, contains one line of parameters
  ifstream stringfile("strings.txt"); //INPUT: symbol strings, one per line
  ofstream fisherfile("fisher.txt"); //OUTPUT: Fisher scores, one data item per line
  int s[n+1]; //symbol string, uses s[1] to s[n] (s[0] is never used)
  double P[p+1][p+1]; //state transition probability matrix
  double P[number_of_symbols+1][p+1]; //conditional probabilities of symbols given states
  double score[p+1][number_of_symbols+1]; //Fisher scores for the emission probabilities
  double score[p+1][p+1]; //Fisher scores for the transmission probabilities
  double forw[p+1][n+1];
  double back[p+1][n+1];

  //initialize to zero
  for (int i=0; i<p; i++)
for (int j=0; j<p; j++)
    PM[i][j] = 0;
for (int i=0; i<number_of_symbols; i++)
    for (int j=0; j<p; j++)
        P[i][j] = 0;
PM[1][0] = 1.0; //because it is a left-to-right hidden Markov model
for (int i=2; i<p; i++)
    PM[i][0] = 0;
for (int i=1; i<p; i++)
    for (int j=1; j<p; j++)
        hmmstream >> PM[i][j];
for (int i=1; i<number_of_symbols; i++)
    for (int j=1; j<p; j++)
        hmmstream >> P[i][j];
while (getline(stringfile, stringstring)) {
    stringstream stringstream (stringstring);

    //initialize to zero
    for (int i=0; i<p; i++)
        for (int j=0; j<n; j++)
            forw[i][j] = 0;
    for (int i=0; i<p; i++)
        for (int j=0; j<n; j++)
            back[i][j] = 0;
    for (int i=0; i<p; i++)
        for (int j=0; j<number_of_symbols; j++)
            scoree[i][j] = 0;
    for (int i=0; i<p; i++)
        for (int j=0; j<p; j++)
            scoret[i][j] = 0;
    s[i] = 0;
    for (int i=1; i<n; i++)
        stringstream >> s[i];
    for (int i=0; i<p; i++)
        for (int j=1; j<number_of_symbols; j++)
            scoree[i][j] = 0; //line 2
    for (int i=0; i<p; i++)
        for (int j=1; j<p; j++)
            scoret[i][j] = 0; //mvs
    for (int i=0; i<p; i++)
        forw[i][0] = 0; //line 3
    for (int i=0; i<p; i++)
        back[i][n] = 1;
    forw[0][0] = 1; //line 4 (corrected)
    Prob = 0;
    for (int i=1; i<n; i++) { //line 5
        for (a=1; a<p; a++) { //line 7
forw[a][i] = 0;  // line 8
for (b=0; b<p; b++)  // line 9 (corrected)
  forw[a][i] = forw[a][i] + PM[a][b]*forw[b][i-1];  // line 10
forw[a][i] = forw[a][i]*P[s[i]][a];  // line 12
}
}
for (a=1; a<p; a++)  // line 15
Prob = Prob + forw[a][n];
for (int i=n-1; i>=1; i--)  // line 18
for (a=1; a<p; a++)  // line 19
  back[a][i] = 0;  // line 20
for (b=1; b<p; b++)
  back[a][i] = back[a][i] + PM[b][a]*P[s[i+1]][b]*back[b][i+1];  // line 22

// Fisher scores for the emission probabilities
for (int i=n-1; i>=1; i--)
for (a=1; a<p; a++)  // line 18
  for (int sigma = 1; sigma<number_of_symbols; sigma++)
    scoree[a][sigma] = scoree[a][sigma] + back[a][i]*forw[a][i]/(P[s][a]*Prob);  // line 24
for (int sigma = 1; sigma<number_of_symbols; sigma++)
  scoree[a][sigma] = scoree[a][sigma] - back[a][i]*forw[a][i]/Prob;  // line 26 (corrected)
}

// Fisher scores for the transmission probabilities
for (int i=n-1; i>=1; i--)
for (b=1; b<p; b++)
  for (a=1; a<p; a++)
  scoret[b][a] = scoret[b][a] + (back[a][i]*forw[b][i-1]*P[s][i][a]/Prob -
  back[b][i]*forw[b][i]/Prob);
for (int i=1; i<p; i++)
  for (int j=1; j<number_of_symbols; j++)
    fisherfile << scoret[i][j] << "\n";
for (int j=1; j<number_of_symbols; j++)
  for (int i=1; i<p; i++)
    fisherfile << scoree[i][j] << "\n";
fisherfile << endl;
}
hmmstream.close();
fisherfile.close();
system("PAUSE");
}
Appendix L

Similar Publications

The following list consists of publications that in some ways may be considered to be similar to my own. The items are given in chronological order, and books dedicated to neural networks in finance are excluded (as they are too numerous).

**Peters (1991)** The book that sparked off popular interest in chaos in the markets (Peters claimed to have found chaos in the markets, whilst most subsequent studies suggest that there is no evidence of low dimensional chaos).

**Trippi and Lee (1992)** Portfolio selection using knowledge-based systems.

**Deboeck (1994)** An interesting, but low-brow, book that caught my interest.


**Peters (1996)** In the second edition, Peters still maintains that there is chaos in the markets.


**Kingdon (1997)** An excellent book that examines the design of an automated system for financial time series forecasting that uses neural networks and genetic algorithms.

**Viner (1998)** A precursor to this thesis, but in my opinion not enough emphasis on domain knowledge, no theoretical insight and no real contribution (i.e. a typical ‘good’ PhD thesis).

**Burgess (1999)** An excellent, if rather long, PhD thesis on statistical arbitrage. More rigorous than this thesis in some areas, less so in others.
Shadbolt and Taylor (2002) An excellent (though naturally disjointed) compilation of relevant techniques, the best there is and the most similar to this thesis.


Almanza (2008) A PhD thesis that uses genetic programming to develop novel approaches for dealing with the classification of unbalanced data sets.

In addition to the above, there will likely be relevant works in progress by the PhD students in the groups led by the following academics:

- Professor Shu-Heng Chen, Department of Economics, National Chengchi University
- Professor Nick Jennings, School of Electronics and Computer Science, University of Southampton
- Professor Han La Poutré, Multi-agent and Adaptive Computation, Centrum Wiskunde & Informatica
- Professor Berç Rustem, Department of Computing, Imperial College London
- Professor Philip Treleaven, Department of Computer Science, University College London
- Professor Edward Tsang, Centre for Computational Finance and Economic Agents (CCFEA), University of Essex
Bibliography


Ahmed, A. H. M. T. (2000), Forecasting of foreign exchange rate time series using support vector regression, 3rd year project. Computer Science Department, University of Manchester.


BIBLIOGRAPHY


Boyson, N. M. (2003), Why do experienced hedge fund managers have lower returns? Working paper, Purdue University.


Chang, C.-C. and Lin, C.-J. (2001), LIBSVM: A Library for Support Vector Machines, National Taiwan University.


Kirchner, T. (2009), Merger Arbitrage: How to Profit from Event-Driven Arbitrage Wiley Finance, Wiley, Hoboken, NJ.


Kumar, A. and Pons, V. (2002), Behavior and performance of investment newsletters analysts Working paper ysm275, Yale School of Management, New Haven, CT.


Mandelbrot, B. (1962), The variation of certain speculative prices, Research Note NC-87, IBM.


Mercer, J. (1909), Functions of positive and negative type, and their connection with the theory of integral equations *Philosophical Transactions of the Royal Society of London. Series A. Containing Papers of a Mathematical or Physical Character* 209, 415–446.


Micchelli, C. A. (1986), Interpolation of scattered data: Distance matrices and conditionally positive definite functions *Constructive Approximation* 2(1), 11–22.


BIBLIOGRAPHY


Powell, N., Foo, S. Y. and Weatherspoon, M. (2008), Supervised and unsupervised methods for stock 
trend forecasting in 40th Southeastern Symposium on System Theory, IIEEE, Piscataway, NJ, pp.203–
205.


Rabiner, L. R. (1989), A tutorial on hidden Markov models and selected applications in speech recogni-


Railsback, S. F. (2001), Getting "results": The pattern-oriented approach to analyzing natural systems 

Rasmussen, C. E. and Williams, C. K. I. (2005), Gaussian Processes for Machine Learning, The MIT
Press, Cambridge, MA.

and Practice 12(1), 75–84.


Roberts, H. (1967), Statistical versus clinical prediction of the stock market, Unpublished manuscript.


BIBLIOGRAPHY


Society of Technical Analysts (1999), Analysis, STA Diploma Course.


Treynor, J. L. (1962), *Toward a theory of market value of risky assets* The first paper on CAPM, yet rarely cited and often incorrectly referred to as ‘Treynor (1961)’.


