Signed Path Dependence in Financial Markets: Applications and Implications

Fabio Silva Dias

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Statistical Science

University College London

February 17, 2021

I, Fabio Silva Dias, 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 work.

Abstract

Despite decades of studies, there is still no consensus on what type of serial dependence, if any, might be present in risky asset returns. The serial dependence structure in asset returns is complex and challenging to study, it varies over time, it varies over observed time resolution, it varies by asset type, it varies with liquidity and exchange and it even varies in statistical structure. The focus of the work in this thesis is to capture a previously unexplored notion of serial dependence that is applicable to any asset class and can be both parameteric or non-parameteric depending on the modelling approach preferred. The aim of this research is to develop new approaches by providing a model-free definition of serial dependence based on how the sign of cumulative innovations for a given lookback horizon correlates with the future cumulative innovations for a given forecast horizon. This concept is then theoretically validated on well-known time series model classes and used to build a predictive econometric model for future market returns, which is applied to empirical forecasting by means of a profit-seeking trading strategy. The empirical experiment revealed strong evidence of serial dependence in equity markets, being statistically and economically significant even in the presence of trading costs. Subsequently, this thesis provides an empirical study of the prices of Energy Commodities, Gold and Copper in the futures markets and demonstrates that, for these assets, the level of asymmetry of asset returns varies through time and can be forecast using past returns. A new time series model is proposed based on this phenomenon, also empirically validated. The thesis concludes by embedding into option pricing theory the findings of previous chapters pertaining to signed path dependence structure. This is achieved by devising a model-free empirical risk-neutral distribution based on Polynomial Chaos Expansion and Stochastic Bridge Interpolators that includes information from the entire set of observable European call option prices under all available strikes and maturities for a given underlying asset, whilst the real-world measure includes the effects of serial dependence based on the sign of previous returns. The risk premium behaviour is subsequently inferred from the two distributions using the Radon-Nikodym derivative of the empirical risk-neutral distribution with respect to the modelled real-world distribution.

Impact Statement

The knowledge presented in this thesis could be put to a beneficial use in several manners, both inside and outside academia.

The proposed models on signed path dependence can generate material benefits inside academia in the field of Econometrics, enriching the existing suite of models by proposing a new paradigm to deal with stochastic processes that exhibit memory that decays much slower than what is implied by an exponential decay but that at the same time does not imply an explosion in variance. I believe there is still great potential to enhance standard econometric theory by further exploring and applying the concept of "signed path dependence" that is introduced in this thesis. Moreover, the knowledge presented in this thesis can also generate material benefits inside academia in the field of Empirical Asset Pricing. There is still not much consensus on how serial dependence manifests itself in risky markets such as Equities, Currencies and short to mid-term Commodity Futures. The empirical findings of this thesis are relevant to such an interesting field. Further, the final chapter of this thesis demonstrates how such findings can be put to a beneficial use also in the field of Derivatives Pricing, by challenging usual assumptions of market behaviour and risk-premium.

Nevertheless, the benefits that this thesis can generate are not limited to inside academia. Outside academia, there is great demand from banks and investment managers for models that can capture as many empirical features of financial markets as possible and be used to generate better informed market predictions and enhanced risk management activity. In addition to financial institutions, commodity producers can benefit from the application of the econometric models here introduced by including into their pricing and hedging strategies the additional information extracted from the markets using the models here proposed.

In the limit, the impact of this thesis could manifest itself internationally, as the knowledge here disseminated propagates to individuals and organisations. The benefits to any individual use of the models here proposed are expected to manifest within a relatively short period of time; however, if further research is indeed triggered by the contents of such thesis, a larger benefit can materialise in the context of a broader field of research, over many years, decades or longer.

Acknowledgements

This PhD has been a very long journey and some people had provided me with invaluable help. I'm very thankful to them.

My first manifestation of gratitude goes to my main supervisor, Prof Gareth W Peters. He provided me with vital advice during bad times and with excellent support during good times. I certainly would not have completed this PhD without him.

I am also very thankful to Dr Paul J Northrop, my secondary supervisor, who also came to help me at a critical time halfway through my PhD in a way that not many people would do. And, of course, his feedback and insights have also been invaluable.

My boss at Lloyds Banking Group, Dr Colin Burke, also deserves a special honourable mention. He has facilitated enourmously my task of being a part-time PhD student and a part-time banking employee at the same time. Both tasks added together arguably made me busier than a full-time student or a full time employee, but Colin gave me all the support within his powers to make me navigate through this.

My son Ian J Dias has been very good to me, understanding the times when I could not play with him or give him some home-schooling because I was too busy with my own schooling. For that, I'm thankful.

And last but not least, I thank my wife Keila R C Dias, who has been very supportive from the start, before I decided to apply for a PhD vacancy at UCL.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter gives a brief introduction on the history behind the subject investigated and what motivates this research. Subsequently, the objectives of the research are delineated, along with the methods used to accomplish such objectives. The chapter is concluded by listing the structure of the rest of the thesis. This introductory chapter is intentionally kept relatively informal and does not give many details on the underlying theories - such details are later given in the Literature Review.

1.1 Motivation

The question on whether it is possible to obtain abnormally high excess returns in investments by identifying predictability in the financial markets is probably as old as the financial markets themselves. And the most baffling feature of this question is that more than 150 years after the first proposed answer, there is still no consensus on what is the right answer.

The foundations of free markets are believed to have been introduced by Adam Smith in his 1776 book colloquially known as “The Wealth of Nations” [1]. In free markets, market participants are self-interested and prices tend to oscillate around some true or fundamental value. Jules Regnault is credited with an early publication of an answer to the question of how financial markets evolve through time, publishing in 1863 his book Calcul des Chances et Philosophie de la Bourse [2]. His book applied several concepts of probability and statistics known at the time and after considerable reasoning, one of his main conclusions is given using capital letters in the middle of page 50: "L'ÉCART DES COURS EST EN RAISON DIRECTE DE LA RACINE CARRÉE DES TEMPS" or, in English, "the changes in course [of the prices] are directly proportional to the square root of time". In 1900, mathematician Louis Bachelier expanded upon Jules Regnault's work to deduce the first known financial forecasting model which assumed price changes followed a random walk in continuous time [3].

The seminal work by Benjamin Graham published in his 1946 book The Intelligent Investor [4] proposed an investment analysis framework involving detailed scrutiny of company accounts, to calculate fundamental values, and thus ascertain when a given investment is cheap or expensive. The objective would be to buy cheap stocks and sell expensive ones. Any excess performance thus obtained would be at the expense of irrational traders, who bought and sold on emotional grounds and without the benefit of detailed analysis.

In 1970 the question about the possibility of developing trading strategies that would yield abnormally high excess returns appeared to be solved when the Efficient Market Hypothesis (EMH) was proposed by Eugene Fama [5], not only advocating unpredictability of the markets, but also giving a strong explanation for it in terms of economic theory, backed by considerable empirical evidence. Such theory was widely accepted by the bulk of the academia, with numerous different types of tests being implemented and confirming its validity. Over the two following decades any publications on the contrary would be seen with some degree of skepticism.

The significant rise in the popularity of index funds that track major market indexes – both mutual funds and ETFs – is due at least in part to widespread popular acceptance of the efficient markets hypothesis. Investors who subscribe to the EMH are more inclined to invest in passive index funds that are designed to mirror the market's overall performance, and less inclined to be willing to pay high fees for expert fund management when they don't expect even the best of fund managers to significantly outperform average market returns.

The economic argument behind the EMH is very compelling. If in a highly competitive market there is a way to get return predictability beyond a drift coming out of being compensated for taking a risk, this will be as inefficient as leaving a €500 bill on the table in a very crowded pub: in no time somebody will go there and will get the bill for themselves, so in a matter of instants there will be no more inefficiency to be exploited.

By the late 1990s the volume of opposing publications – also backed by empirical evidence – would have grown to a level difficult to ignore. Whilst even the proponents of the EMH did not argue in favour of markets being strongly efficient, opponents have found evidence that would put into question even the weak form of market efficienty.

A fringe field of behavioural economics was gaining more and more traction, to the point that in 2002 Daniel Kahneman, a psychologist, was awarded the 2002 Nobel Memorial Prize in Economic Sciences for his contributions based on empirical studies of human behaviour in situations of financial uncertainty. It was an acknowledgement by the academic circles that human behaviour in financial markets is somewhat predictable – but not an acknowledgement that the asset prices themselves are predictable. However, the market crash of 2008-09 and successful exploitation by a number of hedge funds gave further support to the theory that market returns had some element of predictability that could be exploited to enhance the overall returns of an investment portfolio.

Eventually, two different schools of thought evolved and became mainstream in financial economics. As a consequence of such evolution, a rather interesting fact happened in 2013: Eugene Fama was awarded the Nobel Memorial Prize in Economic Sciences for the contribution to the field of economics that his theory and subsequent academic work provided.

However, the same 2013 prize was awarded jointly with Robert Shiller, a leading behavioural economist and arguably a heavy critic of Eugene Fama's theory. This time, it was an acknowledgement by the academic circles that nobody had the definite answer to the question of market predictability – and, as a human science problem, probably nobody ever will have the definite answer – but considerable progress had been achieved to form two economically and empirically justifiable theories.

Incidentally, a PhD student supervised by Eugene Fama in the early nineties researched on evidence to disprove Dr Fama's theory, and used the results of his own doctoral research as basis to later establish one of the most successful investment management companies in the world, with hundreds of employees and US$185 billion in assets under management[1].

Compelled by evidence, another credible theory has arisen, merging concepts of evolutionary psychology, behavioural finance and efficient markets. Maybe, there is a €500 bill on the table after all. It is just that the table is not in a crowded pub; the table actually is hidden in a dark corner of a poorly lit haute-cuisine restaurant for exclusive members and getting there to pick it up will need a fair bit of resources: be it in the form or capital, or research, or cultural change, or all of them. The inefficiency will stay there for as long as there is not enough knowledge about it and speculative capital seeking to exploit it. This theory became known as the Adaptive Markets Hypothesis.

Interestingly enough, Eugene Fama himself eventually conceded that there is some predictability in market returns. In a joint publication with another re-

[1]Data as at 30th September 2019. Source: `http://www.aqr.com`

spectable author in the 105th volume of the Journal of Financial Economics in 2012 [6], they write in the conclusions that “there are common patterns in average returns in developed markets. Echoing earlier studies, [they] find value premiums in average returns in all four regions [they] examine (North America, Europe, Japan, and Asia Pacific), and there are strong momentum returns in all regions except Japan.” However, this won’t disprove the efficient market hypothesis, as they explain these potential anomalies to likely arise out of local value risk premiums associated with size and local momentum premium extremes that few mutual funds would be able to exploit.

This context provides exciting motivation for a novel research aiming to contribute to areas of knowledge that have not been explored yet.

1.2 Research Objectives

This research doesn’t aim to definitively unlock the Holy Grail of market predictability. It would be presumptuous of me to say I can do that. Nonetheless, there are three core questions this thesis aims to answer:

1. Is there any way to detect on a systematic basis patterns of serial dependence in financial market returns and use them to make any type of prediction better than a random guess?
2. If yes, can a new econometric model be proposed that embeds such patterns into the information used to make economic forecasts and enhance existing risk management techniques?
3. What are the implications of any previous findings to derivatives pricing theory and models?

An answer to the first question is inspired by the concept of "time-series momentum" introduced in [7]. In this work, the authors show that a simple but elegant linear time-series regression model that uses only the sign of previous returns could be used to generate useful information about the sign of future returns. This satisfies the requisite of the first question that this thesis aims to answer. Therefore, it is natural to expand upon such model and provide contributions to the existing literature.

Whilst the first question is worded in a way that is relatively easy to solve given existing literature, answering the subsequent questions can be relatively more complex. Given the noisy nature of financial markets, I do not expect deterministic economic forecasts to have any meaningful value in this area - and the focus therefore should shift to probabilistic forecasting methods. I note that probabilistic economic forecasting and risk measurement are intrinsically linked and as such, as soon as a good probabilistic method is developed, there is an immediate application of such a method into risk measurement and management.

Finally, answering the second question brings immediately a link to the third question, as the pricing of derivatives is totally driven by probabilistic forecasts, albeit in a different probability measure than the one under which the second question is answered. This leaves the research with the final core objective to link the two probability measures of interest and ascertain if there are any relevant findings when performing such exercise.

Above all, the main objective of this thesis is to produce material valuable to both the investment community and academics alike. All other objectives are just intermediate targets to achieve the final goal.

1.3 Research Methodology

As with any research, the choice of which methods to use and how to implement them is driven by the central questions the researcher wants to answer. Given that these topics certainly engage the interest of thousands of researchers across the world, there is plenty of literature available on this subject backed both by widely accepted theory and plenty of empirical experiments. Hence, the first stage of the research corresponds to an extensive and comprehensive literature review, which will ensure valuable research time is not wasted either directing the research into areas that have already been proven unfruitful or simply "reinventing the wheel". It will also ensure the best techniques available are leveraged upon.

The quantitative nature of the questions and expected answers naturally demands the research to be geared towards quantitative methods. Qualitative methods would bring little relevant information to this particular study as they would be centred into a handful of specific events and heavily depend on unproven assumptions in order to be translated into a generic case relevant to all markets studied. Quantitative methods, on the other hand, can provide quick and unbiased interpretation and direct applicability of the results, being able to capitalise on a substantial volume of intraday data available. These quantitative methods will be applied in the context of an Evaluation Research.

Professor William Trochim gives a good definition of how this type of research is supposed to be conducted [8]: "Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object."

Throughout this thesis, a number of theories and hypothesis will be formulated which will be tested via one or more statistical models calibrated to empirical evidence gathered from Equity markets, Foreign Exchange markets and Commodity Futures markets. Thereafter, several systematic experiments will be devised aiming at validating such models and providing useful conclusions.

1.4 Thesis Structure

The thesis is structured in six chapters, including the present introductory chapter. Chapter 2 provides a literature review presenting key concepts in general statistics, econometrics, risk management models and derivatives pricing which will be used throughout the entire thesis, giving examples and counter-examples when applicable. Stylised facts and other empirical findings in the relevant fields are also discussed. Chapter 3 introduces the concept of "signed path dependence," demonstrates how this concept ties up with existing econometric models and develops a machine learning method that can be used to predict the sign of future market returns using the sign of previous market returns. Chapter 4 provides a framework to make probabilistic economic forecasts of financial markets by understanding how the shape of the return distribution is influenced by previous realisations of market returns. Chapter 5 develops a new method to obtain an arbitrary risk-neutral measure based on observed European call option prices and to link it to a real-world measure that exhibits signed path dependence. Chapter 6 concludes the work.

1.5 Research Output

The research output from this thesis has been published in/submitted to the following journals and conferences:

- F.S. Dias, F. Kiraly, G.W. Peters, "Non-parametric Price Momentum Models for Global Equity Index and Currency Markets", Royal Economic Society Conference, Bristol, United Kingdom, 2017

- F.S. Dias, F. Kiraly, G.W. Peters, "Testing for Serial Correlation of Unknown Form Using Signed Path Dependence", 2nd International Conference on Econometrics and Statistics, Hong Kong, 2018

- F.S. Dias and G.W. Peters, "A Non-Parametric Test and Predictive Model for Signed Path Dependence", published in Computational Economics, 2019

- F.S. Dias, "Using Conditional Asymmetry to Predict Commodity Futures Prices", submitted to the International Journal of Financial Markets and Derivatives, 2020

- F.S. Dias and G.W. Peters, "Option Pricing with Polynomial Chaos Expansion Stochastic Bridge Interpolators and Signed Path Dependence", Preprint, 2020, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3588871

Chapter 2

Literature Review

This chapter provides a review of important theoretical concepts and empirical evidence already documented in existing literature. Such results will be relied upon throughout the remainder of the thesis.

2.1 The Efficient Markets Hypothesis

The arrival of new information into asset price formation has been subject of extensive discussion. If such information is always fully reflected and incorporated into market prices, asset returns should be largely unpredictable and in the long run driven by compensation for taking market risk. Such a perspective on the driver for returns on investment is most directly justified in settings in which there exists a competitive double auction market with informed rational market participants that will not make irrational errors on a systematic basis, and instead will exploit any known informational inefficiency until its exhaustion. This is typically formulated in financial mathematics pricing theory through concepts such as the Efficient Market Hypothesis (EMH), von Neumann Morgenstern rationality and other related financial assumptions on markets and agent behaviours as captured in, for instance, the technical discussions on such assumptions in Chapters 9 and 14 of [9].

Initial work on the EMH focussed heavily on statistical properties of stock prices ([10],[11],[12]), arguing that in strongly efficient markets asset prices fluctuate in unpredictable and completely random ways. The seminal work in [5] has subsequently provided the definition of three forms of financial market efficiency:

- Weakly efficient markets, where the market price of an investment incorporates all information contained in the price history of that investment, meaning that historical price series and trends have no value to predict future prices. A mathematical interpretation of this level of efficiency is that changes in market prices exhibit no form of serial dependence and knowledge of a stock's price history cannot produce excess performance as this information is already incorporated in the market price. This form, if true, also means that technical analysis (or *chartism*) techniques (i.e. analysing charts of prices and spotting patterns) will not produce excess performance.

- Semi-strong efficient markets, where the market price of an investment incorporates all publicly available information. Knowledge of any public information cannot produce excess performance, as this information is already incorporated in the market price. This form, if true, means that fundamental analysis techniques (i.e. analysing accounting statements and other pieces of financial information) will not produce excess performance. A mathematical interpretation of this level of efficiency is that changes in market prices exhibit no form of serial dependence and that there is no other lagged exhogenous variable known to the general public that can be used to predict future changes in market prices.

- Strongly efficient markets, where the market price of an investment incorporates all information, both publicly available and privately available, for example information about a share price that is available only to company insiders. Knowledge available only to insiders cannot produce excess performance as this information is already incorporated in market prices. Stock markets around the world are subject to regulation. Often rules exist to prevent individuals with access to price sensitive information, which is not yet public, from using this information for personal gain. For example, senior management involved in merger and acquisition talks are often banned from trading in the stock of their company. Such rules would be unnecessary if strong form efficiency held. However, one can argue that if senior management were allowed to trade their own company's stock, then strong form EMH would be possible.

Such school of thought has evolved considerably and several different tests have been developed.

Tests for the weak form of financial market efficiency are similar in nature to the test given in [13]: let P_t be the price of a stock at time t. Assume that P_t obeys the following equation:

$$\begin{aligned} \log\left(\frac{P_t}{P_{t-1}}\right) &= \alpha_0 + \sum_{i=1}^{K} \alpha_i \log\left(\frac{P_{t-i}}{P_{t-i-1}}\right) + \delta h_t^2 + \varepsilon_t \\ h_t^2 &= \omega + \beta \varepsilon_{t-1}^2 + \gamma h_{t-1}^2 \varepsilon_{t-1}^2. \end{aligned} \tag{2.1}$$

If a market is weakly efficient, $\alpha_i = 0$ for any $K \geq 1$, $K \in \mathbb{N}$ and ε_t are i.i.d. In [13], Equation 2.1 was applied to the daily changes of the FTSE 30 index (representing the UK stock market) and the authors have concluded that in their experiment α_i was statistically equal to zero for any $K \geq 1$ and they failed to reject the hypothesis that the residuals of the model were i.i.d.

Similar experiments were applied to other markets, finding similar results.

In [14] a Genetic Algorithm was developed to test the weak form of the EMH in the EUR/USD rate by finding the best parameters of a technical trading system that would yield the best returns in a training sample and then the returns of the same algorithm were tested out-of-sample, with the algorithm being updated on a rolling basis. Their procedure found that best performing trading rules in the training sample could not perform well out-of-sample, yielding support for the weak form of the EMH.

There is some evidence, however, that the weak form of the EMH might be violated only during small intervals of time. In [15] the authors split the time series of concern into disjoint partitions and found that, at times, the residuals were not i.i.d. within a partition even though they could not reject the hypothesis of the residuals being i.i.d. across the entire time series.

It's clear, however, that the fact that a specific test does not detect serial dependence in a given price series does not guarantee in itself that serial dependence is not present: it is only evidence that a particular form of dependence is not present, but there might still be another form of serial dependence that was undetected due to test misspecification. This motivates a continuous research on finding potential new models that can cope with such alleged dependence and, ideally, use the detected dependence to make informed predictions of future market behaviour.

Tests of semi-strong and strong form market efficiency were also extensively studied. These are outside of the scope of the present thesis as the goal of this research is to formulate models for serial dependence that can provide information about future market returns without the aid of any exogenous variables, either of public knowledge or private knowledge. Testing the strong form of the EMH is particularly problematic as a puristic test would require the researcher to have access to information that is not in the public domain. Nonetheless, studies can be made focussing on how well-timed were the purchases and sales made by company insiders. A notable study [16] found that purchases made by insiders earned abnormal returns of more than 6% per year; however, sales made by insiders did not earn significant abnormal returns.

Tests of EMH are fraught with difficulty. There is a substantial body of literature proving the existence of mispricings, in contravention of EMH. Unfortunately, there is also a substantial body of literature providing evidence for various forms of EMH. Both schools of thought can cite a great deal of empirical evidence and an impressive wealth of statistical tests.

It is reasonable to ask, from a philosophical point of view, how it could come about that we have categorical proof of mutually contradictory statements. One possible explanation is that many published tests make implicit and explicit, but possibly invalid, assumptions (for example normality of returns, or stationarity of time series). Some of the differences are purely differences of terminology. For example, do we regard anomalies as disproving EMH, if transaction costs prevent their exploitation?

More subtle is the need to make an appropriate allowance for risk. The EMH is not contradicted by a strategy which produces higher profits than the market portfolio by taking higher risks. The market rewards investors for taking risks, so we expect, on average, a high-risk strategy to result in higher returns. What would contradict the EMH is an investment strategy that provided returns over and above those necessary to compensate an investor for the risk they faced. Unfortunately, there is no universally agreed definition of risk, and therefore no perfectly accurate way of measuring whether any detected excess returns have arisen out of excess risk or out of violations of the EMH.

The question of whether or not markets are efficient has important implications for investment management. Active fund managers attempt to detect exploitable mispricings, since they believe that markets are not universally efficient. Passive fund managers simply aim to diversify across a whole market, perhaps because they do not believe they have the ability to spot mispricings. If markets are inefficient, we would expect active managers with above average skill to perform better than passive managers. However, performance should be considered net of various fees and transaction costs (e.g. brokerage, market impact). To demonstrate an exploitable opportunity in the market, the opportunity should be sufficiently large to remain intact even after all these costs are taken into account.

Moreover, different stock exchanges have different levels of required disclosure of information. Hence it would be reasonable to expect different markets to have different levels of efficiency. For example, the New York Stock Exchange (NYSE), which requires a high level of disclosure, should be more efficient than a market with limited disclosure requirements.

There is also no commonly accepted definition of what constitutes publicly available information. For example, unlike professional fund managers, private investors are unable to gain access to the senior management of companies. Clearly, fund managers have an advantage in terms of being able to form an opinion on the competence of the management team and the strategy of the company. Fund managers are also increasingly utilising alternative sources of data (e.g. satellite images, web searches, social media etc) to generate excess performance. Even if information is publicly available, there is a cost involved in obtaining the information quickly and accurately. Any advantage achieved by acting on price relevant information could well be eroded by the cost of obtaining and analysing that information.

Finally, it's noted that tests of the EMH might never be conclusive, as there is no ground truth defining what is an "Efficient Market". Tests are based on premises and assumptions and conclusions are based on inductive reasoning and no matter how large the sample of empirical observations supporting or not supporting a hypothesis, this is just inferring a hypothesis from singular events.

Any conclusion drawn this way may always turn out to be false.

2.2 Behavioural Finance and Prospect Theory

The EMH has a key underlying assumption that an investor is rational and has well-defined and stable preferences regarding the outcomes of its investments. Further, rational investors are assumed risk-averse and take decisions by maximising their expected utility - this is also known as Expected Utility Theory (EUT). Von Neumann and Morgenstern [17] established a rigorous axiomatic basis for the expected utility approach, which still forms the foundation for most of the economic literature on investor behaviour under conditions of uncertainty. One of the consequences of the axioms of Von Neumann and Morgenstern is that the utility function of a rational investor is smooth and convex.

Having said that, since its inception, the Expected Utility Theory (EUT) has drawn criticism from various quarters, primarily as a result of its axiomatic characterisation of preferences. These critics include Friedman and Savage [18], who argued that an individual can have different utility functions (to be understood as attitudes to risk), depending on initial wealth. This observation was motivated by people's contrasting tendencies to buy insurance (risk-averse) and gamble (risk-seeking). In the Friedman and Savage approach, individuals are risk-seeking at low levels of wealth, and risk-averse at high levels of wealth. Similarly, Markowitz [19] also criticised the general underpinnings of the EUT, arguing that utility should be measured relative to changes from a reference point rather than in absolute values of wealth.

Perhaps the most famous critique of the EUT emerged in the 1970s from a series of papers by two psychologists, namely Daniel Kahneman and Amos Tversky, culminating in their seminal 1979 paper on Prospect Theory [20]. Prospect theory was borne out of various laboratory experiments and sought to detail how human decision-making differs systematically from that predicted by EUT, and how human beings consistently violate the rationality axioms that form its basis. The model is descriptive: it tries to model real-life choices, rather than optimal decisions.

There are two phases of decision-making described in Prospect Theory:

1. *Editing/framing phase* - where outcomes of a decision are initially appraised and ordered.
2. *Evaluation phase* - choosing among the appraised options.

Editing leads to a representation of the acts, outcomes and contingencies associated with a particular choice problem. It involves a number of basic operations that simplify and provide context for choice. The two basic operations involved in the editing process (others also exist) are:

- Acceptance: people are unlikely to alter the formulation of choices presented.
- Segregation: people tend to focus on the most 'relevant' factors of a decision problem.

Within this process, framing effects refer to the way in which a choice can be affected by the order or manner in which it is presented. Standard economic theory considers such transformations to be innocuous with no substantive impact on decisions.

Once prospects are edited, decision makers move on to the evaluation stage where they make their choice. Kahneman and Tversky observed a number of behavioural patterns in people when evaluating various alternatives. Some behavioural patterns include:

- *Reference Dependence* or *Anchoring Bias*: people derive utility from gains and losses, measured relative to some reference point, rather than from absolute levels of wealth. This emerges from the idea that people are more attuned to changes in attributes rather than absolute magnitudes. This generates utility curves with a point of inflexion at the chosen reference point.

- *Endowment effects*: The endowment effect occurs when a person's preferences depend upon what they already possess. This implies that a person's preferences depend upon a certain reference point, perhaps determined by the person's possessions.

- *Loss Aversion*: People are much more sensitive to losses (even small ones) than to gains of the same magnitude. In experiments, the pain from a loss is estimated to be twice as strong as the pleasure from an equivalent gain. Thus, utility curves are asymmetrical in the domain of gains and losses relative to some reference point and are not necessarily convex.

Prospect Theory has been used to explain some puzzles that could not be explained by the EUT by including the concept of asymmetry in the pricing of stocks. Benartzi and Thaler [21] proposed an explanation for the well-known puzzle that the long-term returns of US equities are higher than the long-term returns of US Treasury Bills by a much greater margin than predicted by the EUT ([22]). According to their theory, loss averse investors will require much higher returns (and hence much lower prices to the stocks) to compensate for potential losses arising out of investments in stocks than rational investors assumed by the EUT.

A different puzzle is noted by [23] that stocks of companies that are either distressed, bankrupt or that had just come out of an IPO have average returns that are lower than other stocks of similar or even lower risk of failure - something that is inconsistent with the idea that the market prices risk rationally, as higher risk assets in theory should yield higher returns. They formulated a model that incorporated asymmetrical outcomes and non-convex utilities and noted that stocks with return distributions that are heavily right-tailed would be priced higher than other stocks (hence yielding lower returns). In their study, they noted that distressed and bankrupt stocks on average underperform but a small minority become success stories and command extremely high returns. Similarly, a small minority of new IPOs grow fantastically well from a small company to the new Google or the new Amazon. The unlikely possibility of such strong gains is overweighted by investors that are willing to pay a high price for these assets, even it this means a low expected return. Consistent with [23], an empirical study made by [24] found that the higher the predicted skewness of an initial public offering stock, the lower is its long-term average return.

Behavioural Finance might provide some explanation to the existence of serial dependence in equity returns. For example, suppose a company has filed for bankruptcy and there is clear and public information about its asset and liabilities to make it unequivocal that its stock is worth zero. However, the stock still trades above zero based on hopes of a better reality for the company - or, maybe, based on denial of reality. These hopes/denial are not compatible with a rational investor behaviour and will fade away slowly through time, creating a situation where the price has a definite trend towards zero, yielding a serial dependence structure in the price of the said stock.

2.3 Serial Correlation in Financial Markets and Trading Strategies

It is a stylized fact that asset returns do not exhibit linear serial correlation [25]; however, a number of different tests have been proposed in the literature which have detected statistically significant evidence that asset returns exhibit some form of serial correlation which is not necessarily linear. This could be at odds with the weak form of the Efficient Market Hypothesis if the outcome of such tests can also be used to perform informed predictions about future market returns.

A simplified test of the EMH can be done using standard unit root tests such as the Augmented Dickey-Fuller test [26], the Phillips-Perron test [27] and the KPSS test [28] which look for violations to the hypothesis that asset prices follow a Random Walk, in line with that would be expected from the weak form of the EMH. This approach has been applied in an investigative study in [29]. If the price time series follows a Random Walk, then it must have a unit root while the return series must not. This approach, however, might be considered less powerful because its bound to detect only strong deviations of the EMH as it is possible that a time series does not have a unit root yet still has some form of serial correlation.

Section 3 of [30] provides an extensive review of the most well-known methods. One of the most popular tests is the variance ratio test (VR), first proposed by [31]. Under the VR test, the following statistic is calculated:

$$J(q) := \sqrt{\frac{Nq}{2(q-1)}}\left(\frac{\sigma^2(q)}{\sigma^2(1)} - 1\right) \tag{2.2}$$

where

$$\begin{aligned} \mu &= \frac{1}{N}(\log(P_N) - \log(P_0)) \\ \sigma^2(q) &= \frac{1}{N}\sum_{k=1}^{N}\left(\log\left(P_{qk}\right) - \log\left(P_{qk-q}\right) - q\mu\right)^2, \end{aligned}$$

P_t is the price of a stock at time t and N is the total number of price observations since the start time 0. The test postulates that, in the absence of serial correlation, the test statistic $J(q)$ follows a standard Gaussian distribution as $N \to \infty$ for any choice of $q > 1, q \in \mathbb{N}$.

Remark 1. *Notice that the VR statistic given by Equation 2.2 is in essence testing* *whether the variance of the q-period log-return is statistically equal to q times the variance of the one-period log-return. Also, for Equation 2.2 to be well defined, N must be a multiple of q when choosing the values of N and q to calculate the test statistic.*

In [31] it was noted that the VR statistic had optimal power when testing a random walk model against an alternative hypotheses of a random walk plus a mean-reverting process. Nevertheless, their results also indicated that such an alternative hypothesis was not a complete description of stock price behaviour. Further, several studies found that the VR statistic did not follow a standard Gaussian distribution across several global equity markets [32].

Nonetheless, we also note the following complications of the VR test:

- If q is very large in Equation 2.2, there might too few observations to calculate the VR statistic exactly as described in the equation and therefore in practical applications overlapping data is typically used to potentially improve power of the VR test; however, the use of overlapping data also creates difficulties when analyzing the exact distribution of the VR test statistic.

- Even if the VR test shows there is evidence of some form of serial correlation, on its own it is not evidence of a violation of the weak form of the EMH because the VR statistic does not provide information sufficient to make informed predictions of future market price behaviour.

Other well-known tests aiming to reject the hypothesis of the absence of serial correlations in financial markets include the automatic Box-Pierce test in [33], long-memory tests in [34], Hurst Exponent tests in [35] and tests on the frequency domain that can be found in [36]. Most of these tests are akin to Portmanteau tests where the alternative hypothesis is arguably specified in a loose manner.

An interesting model used for time series prediction is given in [7], where the authors define a linear regression of the return of an asset scaled by its ex-ante volatility against the sign of its return lagged by some amount of time and shows empirically there is strong evidence of predictability when a lagged 1-year return is used to predict the return one month ahead. It has been shown that for several markets the financial return in a particular month was positively correlated with the sign of the cumulative financial return of the previous twelve months, a phenomenom the authors called "time series momentum" and that would arguably be the source of positive returns in trend-following strategies. Empirical experiments demonstrated that, when employed together with a suitable leveraging strategy based on volatility scaling, such a phenomenom could be used to build investment strategies that offered a risk-reward ratio superior to the one offered by conventional equity investments such as buying and holding a passively managed investment fund linked to an equity index.

Similar studies of this nature can also be found in [37] where the authors establish relationships between univariate trend-following strategies in assets in futures markets and commodity trading advisors in order to examine questions of capacity constraints in trend following investing.

Additionally, [38] found evidence of negative serial dependence in daily returns of international stock, equity index, interest rate, commodity and currency markets, with such negative dependence becoming even stronger when the returns are decomposed into overnight (markets closed for trading) and intraday (markets open for trading). These studies though lacked a stronger theoretical basis, being instead more focussed on the empirical aspects.

Given that the construction of a profitable trading strategy is one of the main practical uses of a model that can predict future market returns, the performance evaluation of methods developed in this theses to test serial correlation is heavily guided by the financial aspects, such as statistical and economic significance of theoretical profits. However, we also remark that some studies claim that apparent profits from strategies based on serial correlation might actually not be arising out of serial correlation in returns but possibly out of intermediate horizon price performance ([39]) or exogenous factors such as the presence of informed trading ([40]) or imbalances in liquidity and transaction costs ([41]). In such cases, the use of time series models will not yield significant benefits to an investor in the long run, as claimed in [42].

An example of false discoveries applied to evaluating trading methods is given in [43]. In that manuscript, amongst other things, the authors evaluated a simple predictive strategy: an investment analyst would choose one stock randomly and send predictions to 50000 different people saying that, after one week, the chosen stock would go up and predictions to other 50000 different saying the same stock would go down. At the end of that week, the analyst would check if the stock went up or down and from the group of 50000 people that received the correct "prediction" a week before, the analyst would send 25000 predictions saying that, after one week, the stock would go up and 25000 predictions saying that the stock would go down. And the same process is repeated every week for ten weeks. At the end of 10 weeks, 97 people would have seen the analyst "predict" correctly the sign of the stock return one week ahead for ten weeks in a row which, from their point of view, it would have been very unlikely to be obtained by random guess ($p = 0.000976$). As a result, these 97 people conclude that the model used by the investment analyst has superior predictive ability to predict the sign of a stock return, and that is clearly a wrong conclusion. The authors used this example to warn anyone evaluating predictive market models to avoid inadvertently creating multiple testing biases which can come, for example, from model overfitting, and proposed corrective actions based on standard multiple inference tests, such as the Benjamini and Yekutieli Test [44].

Heeding this warning, the predictive methods in Chapters 3 and 4 have always been evaluated in out-of-sample environments and a minimalist number of parameters has always been chosen. In particular, the choice of a static training sample for the machine learning method of Chapter 3 meant in practice that there was no parameter calibration on new market data for more than nine years, and the model still performed well even after being left without being updated for such a long time interval, which can be seen as a good evidence against overfitting biases such as the one mentioned in [43].

2.4 Machine Learning and Predictive Models for Financial Markets

It is noted that this thesis uses cross-validation and bootstrap techniques to calibrate and test the predictive model proposed. A comprehensive review of these techniques can be seen in [45]. Numerous studies have used similar techniques to extract serial dependence and predict future behaviour in market prices: [46], [47] and [48] to name a few.

In [46], three different classifiers are used to predict the sign of the return of the Dow Jones Industrial Average index on a given day based on the returns of the previous days. These classifiers are based on an Artificial Neural Network (ANN), k-Nearest Neighbour algorithm and a Decision Tree, respectively.

All three algorithms had their parameters calibrated using a 5-fold cross validation, with 80% of the data being used to train the model and the remaining 20% of the data used to evaluate the classification error. The authors reported 60% accuracy in the prediction using the ANN classifier and 57% accuracy in the prediction using the other two classifiers, with the prediction accuracy measured by the number of days a sign was correctly predicted against the total number of days for which a prediction was made.

In [49] the authors used machine learning methods to choose individual stocks as components of equity investment portfolios and to make predictions of expected returns of stocks and equity indices. The authors investigated nearly 30,000 individual stocks over 60 years from 1957 to 2016 and used as predictors 94 characteristics for each stock, interactions of each characteristic with eight aggregate time-series variables, and 74 industry sector dummy variables, totalling more than 900 baseline signals. These predictors were used to fit, for each stock, different calibrations of Neural Networks, Random Forests and Boosted Regression Trees, with the fitted models used for prediction and selection of stocks to invest based on the highest predicted expected return. The machine learning methods were also benchmarked against a simple Ordinary Least Squares fitting. The authors reached important conclusions:

- Price return over the last month (which they referred to as one-month momentum) was the most influential predictor in all machine learning models, with the second most influential predictor being the size of the company being traded.

- Shallow learning (between one and three layers in the neural network) outperformed deep learning (more than three layers in the neural network), which the authors presumed to be a result of the low signal-to-noise ratio typical of financial returns and the fact that financial time series, however complete, will always contain substantially less observable data than other non-financial settings such as computer vision.

- Using machine learning methods to choose portfolios increased the Sharpe ratio of the investment portfolio from 0.61 (if no machine learning is applied and the investor simply buys and holds a stock) to 1.35 (if the stocks are chosen based on a rank of the expected returns predicted by a neural network algorithm).

Studies such as the one aforementioned have interesting value in the field of Computational Economics though they lack a clear link to Time Series models predominantly used in the field of Econometrics, which is the main area of research in the present thesis. Therefore, although similar machine learning and classification methods were used to guide part of our research, these methods do not provide a complete model framework needed in our research, being mostly of ancillary value.

2.5 Sign Tests and Financial Event Analysis

Stock prices can be affected by events happening in previous days - which could cause some form of return predictability if the impact of the event can be forecast. An area of active research for these effects is Event Analysis: according to [50] there are at least 500 published papers that deal with event study methodology and this number continues to grow. A widely used approach to analyse these events is based on sign tests. Such concept, brought to mainstream finance by [51] and refined by [52], is based on the assumption that if market prices are reflecting all the information available due to the occurrence of any particular event, abnormal returns should not happen after that event and therefore the frequency of positive and negative returns conditional on that event occurring should stay unchanged compared to its unconditional long term frequency.

The earliest documented use of a sign test for financial event analysis can be seen in [53], which examines the stock price reaction to stock splits by studying nominal price changes after the occurrence of the split. Using a sample of 95 splits from 1921 to 1931, the author found that the price increased in 57 of the cases but declined in only 26 instances.

In formal statistical framework, this problem has been formulated as rejecting a null hypothesis of no abnormal positive return frequencies after the occurrence of the event and is based on the following test statistic:

$$z = \frac{|p_0 - p|}{\sqrt{p(1-p)N}}$$

where p_0 denotes the observed fraction of positive returns after the event in question, p is the long-term unconditional frequency of positive returns and N is the number of observations. Under the null hypothesis, this statistic asymptotically follows a Normal distribution with zero mean and unit variance.

The test specification above has the potential to be very powerful regarding situations with a small signal-to-noise such as finance given that it removes a considerable amount of noise and, as such, is the preferred approach of the present thesis. However, the sign test aforementioned requires some improvement in order to be used in a time series context. For example, the "event" is not clearly defined as a function of a lookback window or forecasting horizon or as a function of the time series properties themselves. This limits how much the statistic can be generalised into a prediction model or how much its power can be studied. In fact, according to [50] the power and specification of sign tests based on financial events have not been documented.

2.6 Business Cycles, Trading Rules and Conditional Asymmetry

Numerous tests have been developed to identify business cycles in equity markets whereby there are periods of consistently low returns (informally known as "bear markets") and periods of consistently high returns (informally known as "bull markets").

In one of the studies, using over 160 years' worth of monthly US equity data, [54] used a variation of the Markov Switching model [55] to explain features such as the duration of these states and the volatility dynamics under each market.

Whilst studies using 160 years worth of data should be looked at with caution given that markets are very different than 160 years ago with regards to the use of automation, the trading volumes and the industry/sector of major corporations, some features are still relevant to present days. For example, the S&P 500 had its lowest first-quarter return ever in the first-quarter of 2020, even lower than the stressful record set in the first-quarter of 1938 [56], meaning that old observations representing high volatility periods are still relevant to present days.

An application of business cycle models in the UK markets can be seen in [57], also demonstrating clear periods of consistently low returns and periods of consistently high returns. The presence of such periods of continuity in the direction of the market has motivated studies attempting to detect whether at least the future sign of the market returns is predictable based on previous returns: a few recent examples can be seen at [58], [59] and [60].

Subsequent work on an extended universe of assets and over a greater historical record [61] found that in each decade since 1880 investment strategies based on time series momentum using lookback horizons of one month, three months and twelve months would have delivered positive average returns with low correlations to traditional asset classes. The authors claimed such a strategy has delivered positive average returns in each market, with an average Sharpe ratio of approximately 0.4 net of transaction costs (where the Sharpe ratio is defined as the quocient between the annualised return of the strategy and is annualised standard deviation). Based on such evidence, the authors have concluded that the presence of monthly trends in global market returns is a pervasive feature of such markets.

Another study applied simple well known trend-following trading rules by the means of different parameterisations of moving average crossovers and observed that such trading rules, when applied to 22 different commodity futures prices on a monthly frequency, yielded positive returns which were both economically and statistically significant [62].

Some studies attempted to ascertain whether a potential source of nonlinear serial dependence in financial returns arises out of conditional asymmetry. Under the presence of conditional asymmetry, the shape of the probability distribution of the returns varies in a way that can be expressed as a function of previous returns.

Using a time-varying log-gamma distribution, in [63] the authors found in the NYSE composite daily returns from 1981 to 1999 evidence of time varying skewness driven by, amongst other factors, the return in the previous day. In [64] the authors obtained skewness estimates from the risk-neutral probability distribution taken from a sample of option prices from 1996 to 2005 and verified strong correlation between risk-neutral skewness and the future returns of the underlying stocks.

2.7 Option Pricing Theory: Real-World and Risk-Neutral Distributions

The question of predictability of market returns has drawn strong interest by many academics and practicioners in the field of finance. Such level of interest does not appear to wane and recent work has brought additional evidence to be analysed that, though relatively well documented, is yet to be incorporated into risk management practices and mainstream option pricing theory.

In modern option pricing theory it is accepted that the price of an option does not depend on the actual probabilities of events happening and instead depends on an agreement between buyers and sellers of an acceptable value for these probabilities. To understand why, this can be demonstrated using a simple example.

Suppose we are back to the day just before the final of the Fifa World Cup 2018 and we know the final is France vs Croatia, so we know for sure that there are only two possible outcomes to who is going to be the champion: either France or Croatia. However, there is disagreement between the actual probabilities of who is going to be the champion. French supporters believe there is a 70% probability of France being the champion. Croatian supporters, on the other hand, believe Croatia will surprise France, and therefore for them the probability of France being the chamption is only 45%.

One French and one Croatian supporter would like to bet on the outcome of the final, but there is no market for that, so they come to Fabio Dias, a clever middleman who proposes the following deals:

- For the French supporter, Fabio offers a deal whereby if France is the champion a £100 bet returns £142.85 (with a full loss if Croatia is the champion). The French supporter sees it as a fair bet as, for him, the expected value of the £100 bet is £100 so the French supporter accepts it in principle.

- For the Croatian supporter, Fabio offers a deal whereby if Croatia is the champion a £100 bet returns £181.82 (with a full loss if France is the champion). The Croatian supporter also sees it as a fair bet given that, for him, the expected value of a £100 bet is £100 so the Croatian supporter accepts it in principle.

In this situation, Fabio would have guaranteed a profit of at least £18.18 with no risk (in case Croatia is the champion and Fabio has to pay £181.82 to the Croatian supporter out of the £200 initially paid by the two supporters), with a chance of a profit of £57.15 in case France is the champion.

The Croatian and French supporters, however, are more clever than Fabio and decide to talk to each other before signing the deal with the middleman.

They notice that if they agree between themselves a deal where both pay £100 but the winner takes all the £200 home, they will get an outcome that is better for both of them than if they had dealt with Fabio. So they tell greedy Fabio to look for another person to exploit and they agree to the alternative deal between themselves directly.

In this simple example, we can see that, irrespective of whether the supporters agreed the actual probability of France being the champion, they agreed on a price for the bet that implied that the probability of France being the champion was 50%. In the terminology of modern option pricing theory, this probability is called the risk-neutral probability. The actual probability of France being the champion did not affect directly the price of the bet. This probability is called the real-world probability. The two probabilities relate to each other via a quantity called the market price of risk. The market price of risk reflects how much market participants are willing to pay to insure themselves in case certain events occur.

Notice that real-world probablities always relate to future events and therefore there is no "ground truth" for them in the sense that any of their properties will depend on model assumptions. On the other hand, there is a "ground truth" for some properties of risk-neutral probabilities, which can be obtained by simply looking at actual transaction prices in the derivatives market. More specifically, the Fundamental Theorem of Asset Pricing [65] states that when there are no arbitrage opportinuties there is a risk-neutral measure equivalent to the real-world probability measure such that the price π at time zero for any derivative contract with maturity at time T is given by:

$$\pi = \int_{-\infty}^{\infty} \exp(-rT) q(x) f(x) \mathrm{d}x \tag{2.3}$$

where r is the continuously compounded risk-free interest rate from time zero until time T, $q(x)$ is the risk-neutral distribution of the asset price x at time T and $f(x)$ is the payoff of the derivative contract at time T given the asset price.

Obviously, in practical applications the probabilities of events are more complex than a simple binomial distribution as in the footballing example. For example, in practical applications in stock markets it is widely known that derivatives (which can be seen in the context of our footballing example as bets) that pay at equity market crashes are more expensive than derivatives that pay at equity market rallies - a phenomenom also known as the "volatility skew". The presence of a volatility skew does not necessarily mean that markets are more likely to go down than to go up. It only means that market participants are willing to pay more to insure themselves against markets going down than to insure themselves against markets going up.

The empirical findings of [7], [61], [66] and [67] are compatible with asset price dynamics models for the real-world probability distribution that incorporate an element of time dependence in the mean of the process. There are already several models that can incorporate time dependent mean, such as [68], [69] and [70]. However, these models have the drawback that they imply a non-negligible linear serial dependence in asset returns, a condition that is not expected to hold in Equity and Currency markets [25]. As such, these models are well established for use in Interest Rate and Commodity markets.

On the other hand, risk-neutral distributions that accurately reflect all market prices being observed must be able to handle somewhat complex and arbitrary distributions. Non-parametric or semi-parametric models that can be calibrated to option prices can be found in previous work such as [71], [72] and [73].

The research in [71] is considered a pioneering work in the field of option pricing under the assumption of arbitrary stochastic processes for the underlying asset price. In this case, the risk-neutral density $q_J(P_T)$, where P_T is the asset price at time T, has a specific functional form given by:

$$\begin{aligned} q_J(P_T) = a(P_T) + \frac{\kappa_2(q_J) - \kappa_2(a)}{2!}\frac{\mathrm{d}^2 a(P_T)}{\mathrm{d}P_T^2} - \frac{\kappa_3(q_J) - \kappa_3(a)}{3!}\frac{\mathrm{d}^3 a(P_T)}{\mathrm{d}P_T^3} \\ + \frac{\kappa_4(q_J) - \kappa_4(a) + 3(\kappa_2(q_J) - \kappa_2(a))^2}{4!}\frac{\mathrm{d}^4 a(P_T)}{\mathrm{d}P_T^4} + \varepsilon(P_T) \end{aligned} \tag{2.4}$$

where $\varepsilon(P_T)$ is a residual error, $a(P_T)$ is an approximating density function (the log-normal density was chosen in [71]) and $\kappa_i(q)$ is the i-th cumulant of the distribution corresponding to density q. For reference, the first three cumulants of any given distribution are the three central moments and the fourth cumulant is a measure of its kurtosis.

2.8 Polynomial Chaos Expansion

Another possible way to express complex risk-neutral distributions is the use of Polynomial Expansions.

In [74] the authors demonstrate that any arbitrary risk-neutral distribution $q(x)$ can be expressed as a polynomial expansion of an auxiliary density $w(x)$ as long as $w(x)$ dominates $q(x)$, similarly to Equation 2.4. In this case, the approximation is still made on the density function.

Polynomial Chaos Expansion (PCE), first introduced by Wiener in 1938 [75], aims to approximate a random variable that follows any arbitrary probability distribution by a weighted sum of polynomial functions of random variables following well-known, simple distributions. In the most common practical application of PCE, let $X \in \mathbb{R}$ be the random variable of interest and $Y \in \mathbb{R}^m$ be the random variable in terms of which X will be expressed, also known as the *germ*. We want to find the weights $w_j, j \in \mathbb{N}$ such that

$$X = \sum_{j=0}^{\infty} w_j \psi_j(Y) \tag{2.5}$$

where ψ_j form an orthogonal basis with respect to the probability density function $f(Y)$ of Y, i.e.

$$\int \psi_i(Y)\psi_j(Y)f(Y)\mathrm{d}Y = 0$$

for all $i \neq j$.

The most straightforward case is the Hermite Chaos Expansion, where the polynomial functions ψ_j are Hermite polynomials and the random variable Y follows a multivariate Gaussian distribution of independent marginals. A good tutorial to PCE can be found in [76].

Risk-Neutral distributions have already been expressed using PCE in [77]. In that manuscript, the authors tested three different assumptions for the stochastic process driving the evolution of the random variable X through time: a Geometric Brownian Motion, an Ornstein–Uhlenbeck process and a CIR process [78]. These processes are customarily used in financial applications to model risk-neutral distributions of equity markets and interest rates. The authors derived PCE approximations to the solutions of the Stochastic Differential Equations (SDEs) applicable to each of these processes. Their numerical simulations demonstrated that the PCE approximations obtained provided good approximations to the known closed-form solutions with a good rate of convergence, showing a promising potential for future research on the use of PCE to obtain solutions of other SDEs that do not have closed-form solutions.

However, a natural expansion of option pricing models is to develop methods that can incorporate all available market information into a model-free determination of the risk-neutral distribution by choosing an appropriate germ and polynomial basis and linking Equation 2.3 to Equation 2.5. When observed option prices are given as input, such calibration scheme would yield PCE weights that reflect observed option prices and guarantee by construction that the risk-neutral distribution obtained is a valid distribution at all times, something that cannot be guaranteed with a naive calibration of Equation 2.4 (in this Equation there is no guarantee that $q_J(P_T)$ will always be greater than zero, for example).

To the best of our knowledge, Chapter 5 is the first in the literature that has performed this link and provided a full calibration method to obtain the weights in a Hermite Chaos Expansion, testing the results in an empirical setting.

Chapter 3

A Non-Parametric Test and Predictive Model for Signed Path Dependence

This chapter provides theoretical contributions to the literature by proposing in Section 3.1 a non-parametric definition of signed path dependence and demonstrating the sufficient and necessary conditions for it to be present in covariance stationary time series. Further, Section 3.2 proposes a formal inference procedure to detect serial correlation of unknown form based on a hypothesis testing formulation of signed path dependence, which is validated on experiments on synthetic data in Section 3.3. This chapter provides empirical contributions to the literature in Section 3.5 by using the test previously defined to detect evidence of serial correlation in a number of equity index and foreign exchange markets. Additionally, based on the test statistic proposed, a predictive model is defined and used to feed trading strategies whose out-of-sample performance is analysed, gross and net of transaction costs.

This chapter contributes to the field of research of machine learning applied to financial markets predicion by developing its own classifiers which are bespoke to markets that have signed path dependence and hence will have better classification performance than models such as Support Vector Machines, Genetic Algorithms for trading rules and others, given that these aim to be more generic and hence less sensitive to peculiar market features. After all, it is expected that a test whose null hypothesis is more specific to a particular problem in financial time series is likely to have more power and more accurate estimation properties than a test based on wider machine learning tools that will consequently imply a wider scope for its equivalent null hypothesis.

This chapter focuses solely on both how to detect serial correlation in a formal statistical manner and how to model it, paving the way to build predictive models that can be used for the purposes of general market forecasting/risk management or even for the development of trading strategies. As such, technical discussions around assumptions on the price discovery process which might or might not generate price predictability are left outside of the scope of this research.

Further, as with any time series, its behaviour can abruptly change due to exogenous reasons; it is a common problem that can possibly reduce the usability of time series models, even though the tuning scheme based on a rolling observation window used in the calibration of the predictive model hereby proposed can alleviate this problem by readapting the model to most recent data and leaving obsolete data out of the fitted model.

3.1 Defining Signed Path Dependence in a Time Series Context

In this section we propose an extension of the concept of *time series momentum* introduced in [7]. In [7] it has been shown that for several markets the financial return in a particular month was positively correlated with the sign of the cumulative financial return of the previous twelve months. Empirical experiments demonstrated that, when employed together with a suitable leveraging strategy based on volatility scaling, such a phenomenom could be used to build investment strategies that offered a risk-reward ratio superior to the one offered by conventional equity investments. Our statistical framework is more general so that it provides a model-free hypothesis test for the presence of significant correlation, positive or negative, between the financial return in a given forecasting horizon against the sign the cumulative financial return over any arbitrary lookback window and use such alleged presence to predict the sign of future market returns. We also provide a method to detect the optimal lag to be used for forecasting and analyse our definition from the point of view of a number of parametric econometric models.

In most of this work, we will consider a *single* asset whose price is observed at different time points $T \subseteq \mathbb{Z}$, interpreted as unit intervals of observation. We will denote the log-price of the asset at time $t \in T$ by $p(t)$, the log-return by $r(t)$, as usual defined as $r(t) := p(t) - p(t-1)$.

We further define *cumulative log-returns* over longer time scales as follows:

(i) $r_{-n}(t) = p(t) - p(t-n) = \sum_{i=0}^{n-1} r(t-i)$,

the (n unit intervals) *lookback cumulative log-return* at time t.

(ii) $r_{+h}(t) = p(t+h) - p(t) = \sum_{i=1}^{h} r(t+i)$,

the (over a horizon of h unit intervals) *look-ahead cumulative log-return* at time t.

As it is common, by slight abuse of notation, we will denote by the above objects simultaneously random variables from which they may be obtained as observations/realisations, with arguments t, n, h being the only quantities considered fixed and not random. When expectations are taken, they are taken over the whole generative process. In the definitions below, we will condition on some observational knowledge at a time point t; at this state, $r_{-n}(t)$ is constant (not random) for any n, while r_{+h} is a random variable since the future price $p(t+h)$ is unknown. Additionally, throughout this chapter we shall be referring to $\overline{r}$ as the estimated long term mean log-return of the asset being modelled and we shall refer to $s(t+1) = \text{sgn}(r(t+1) - \overline{r})$ where sgn is the sign function. In the context of estimation, $\widehat{s}(t+1) \in \{-1, 1\}$ is the predicted realisation of $s(t+1)$ given the observed realisation of the log-return time series up until time t.

Definition 1 generalizes the concept of time series momentum by the use of conditional expectations of cumulative returns:

Definition 1 (Signed Path Dependence)**.** *We say the considered asset (or the associated process) has* signed path dependence with memory n at forecast horizon $h > 0$ *- or simply* dependence*(−n,+h) - if, for all t where $r_{-n}(t) \neq n\mathbb{E}[r(t)]$ it holds that*

$$\text{sgn}(r_{-n}(t) - \mathbb{E}[r_{-n}(t)])(\mathbb{E}[r_{+h}(t) | r_{-n}(t)] - \mathbb{E}[r_{+h}(t)]) \neq 0 \tag{3.1}$$

Definition 1 intuitively says that, in a process with signed path dependence, knowledge of the lookback cumulative innovation over n time intervals in the past allows us to guess the sign of the look-ahead cumulative innovation h intervals in the future. If the expression on the left side of Equation 3.1 is positive, this sign will be the same, and if negative, the sign will be the opposite.

We intentionally look only at two cumulative innovations and only at the sign of the future one to keep the definition intentionally parsimonious; this will allow us to check whether there is serial correlation without having to specify the exact quantitative nature of such a dependence. While being less strong for forecasting than an explicit forecasting model, we intentionally abstain from a more parametric approach as we first want to answer the question whether there is any form of dependence that could be forecast instead of right away making a forecast of a specific form. However, our definition still allows one to forecast future innovations with more accuracy than making a random guess (by accurately forecasting their signs) and such a forecasting exercise on its own might after all be the best forecast one can make about future innovations if they are highly dynamic and seemingly unpredictable.

Formally, we would like to point out that the definition of signed path dependence depends on the lookback and look-ahead horizons n and h, as well as $p(t)$ considered as a random variable. n and h are parameters of either definition, and while $p(t)$ is unobserved we will see that "having serial dependence" is a property that can be estimated from observations of $p(t)$ via predictive strategies. Further properties and basic observations pertaining to signed path dependence of a time series are listed in Remark 2 below.

Remark 2. *Some remarks about Definition 1:*

(i) The main advantage of the new dependence concept over established measures of correlation such as Pearson and Copula-based measures comes to the fact that it naturally removes from the estimation noise that is highly present in financial applications. Such noise makes impossible to determine linear serial dependence measures for equities, as documented by [25]. However, at the end of the chapter there are tests on empirical data demonstrating superior performance to the Pearson counterpart, which comes exactly from the removal of noise in the estimation framework.

(ii) The definition could also have been made as a function of prices only at three time points, namely at times t, $t+h$, and $t-n$.

(iii) The same asset can have both positive dependence *$(-n,+h)$ and negative* dependence *$(-n',+h')$ simultaneously as long as the pair (n,h) is different from the pair (n',h').*

(iv) As the definition requires the inequality to hold for all time points t, it is possible that it won't hold across the entire process; however, one can also

define a local version of dependence where the inequality would hold only in a localised version of the process.

(v) *The definitions make no assumption on the probability distribution of the returns of the asset or the rate of decay of statistical dependence of two points with increasing time interval.*

(vi) *The definitions are independent of the time scale on which the estimation is done: given three time points, it does not matter whether or not the observations have finer resolution. This means parameters for high frequency data and low frequency data can be estimated using the same algorithm and even the same time series: for example, if one has a high frequency data series but sets the forecasting horizon h to a sufficiently large number, one will be in essence making low frequency forecasts.*

In this chapter, we theoretically validate our definition of signed path dependence by verifying the conditions for its presence in the class of linear processes that admit a Wold representation.

A key assumption in the models subsequently defined is that the input series is covariance stationary. Such a property is defined as follows:

Definition 2 (Covariance Stationarity)**.** *A time series $r(t)$ is covariance stationary if $\mathbb{E}[r(i)r(i+p)] = C_p$ for all $i > 0, i+p > 0$ where C_p is a constant number.*

We now provide the necessary and sufficient conditions for positive or negative path dependence in the case of covariance stationary processes. Our characterization will build on the the explicit classification of such processes given by Wold's representation Theorem 1 [79].

Theorem 1 (Wold Representation Theorem)**.** *Every covariance stationary time series $r(t)$ can be written as the sum of two time series, one deterministic and one* *stochastic, in the following form:*

$$r(t) = \sum_{j=0}^{\infty} b_j \varepsilon(t-j) + \eta(t), \tag{3.2}$$

where ε_t *is an uncorrelated innovation process,* $b_j \in \mathbb{R}$*, and* $\eta(t)$*, is a pure predictable time series, i.e.* $P[\eta(t+s)|r(t-1), r(t-2), \ldots] = \eta(t+s), s \geq 0$ *where* $P[\eta(t+s)|x]$ *is the orthogonal projection of* $\eta(t+s)$ *on* x*.*

Remark 3. *All linear processes have a Wold representation. Furthermore, one may consider this theorem as an existence theorem for any stationary process.*

The derivation of necessary conditions for signed path dependence in several parametric models depends on the lemma below:

Lemma 1. *Let* $x_{1 \leq i \leq n}$ *be independent and normally distributed random variables with zero mean and variance* s_i^2*. For a fixed* j*,* $1 \leq j \leq n$*, given positive weights* $b_{1 \leq i \leq n}$ *we have that*

$$\mathbb{E}\left[x_j \Big| \sum_{1 \leq i \leq n} b_i x_i \right] = b_j \frac{s_j^2}{\sum_{1 \leq i \leq n} b_i^2 s_i^2} \sum_{1 \leq i \leq n} b_i x_i \tag{3.3}$$

Lemma 1 allows us to validate the definition of positive dependence in a number of well known time series model classes, as long as the innovations are Gaussian. We note that extensions can be developed for other classes of driving white noise sequences, including heavy tailed cases.

Proof. Define $z = \sum_{1 \leq i \leq n} x_i$. Since the x_i are independent, z is normally distributed with zero mean and variance $\sum_{1 \leq i \leq n} s_i^2$, hence (x_i, z) are jointly normally distributed with correlation $\rho_{x_i, z} = \frac{s_i}{\sqrt{\sum_{1 \leq i \leq n} s_i^2}}$. From this result it follows that

$$\mathbb{E}\left[x_j \Big| \sum_{1 \leq i \leq n} x_i \right] = \frac{s_j^2}{\sum_{1 \leq i \leq n} s_i^2} \sum_{1 \leq i \leq n} x_i,$$

which implies the equality that was to be proven. □

The following theorem gives the conditions for negative or positive path dependence for processes admiting a Wold representation:

Theorem 2 (Signed Path Dependence in Covariance Stationary Processes). *Let $r(t)$ be a purely non-deterministic covariance stationary process with Gaussian innovations that has a Wold representation $r(t) = \sum_{j=0}^{\infty} b_j \varepsilon(t-j)$ with uncorrelated innovations such that $\varepsilon(t) \sim \mathcal{N}\left(0, \sigma_t^2\right)$.*

If $\Psi(h,t) := \sum_{i=1}^{h} \sum_{j=h}^{\infty} b_{i+j-h} \sigma^2_{t-i-j+h+1}$ then the following characterization holds:

(i) The process $r(t)$ has positive dependence$(-n,+h)$ if and only if $\Psi(h,t) > 0$ for all t.

(ii) The process $r(t)$ has negative dependence$(-n,+h)$ if and only if $\Psi(h,t) < 0$ for all t.

All above statements are independent of n.

Proof. Observe the following identity

$$
\begin{aligned}
r_{+h}(t) &= \sum_{i=1}^{h} \sum_{j=0}^{\infty} b_j \varepsilon(t+i-j) \\
&= \sum_{i=0}^{h-1} \left(b_i \sum_{j=1}^{h-i} \varepsilon(t+j) \right) + \sum_{i=h}^{\infty} \sum_{j=1}^{h} b_{i+j-h} \varepsilon(t-i-j+h+1),
\end{aligned}
\tag{3.4}
$$

implying that

$$
\mathbb{E}[r_{+h}(t) | r_{-n}(t)] = \sum_{i=h}^{\infty} \sum_{j=1}^{h} b_{i+j-h} \mathbb{E}[\ \varepsilon(t-i-j+h+1) | r_{-n}(t)]. \tag{3.5}
$$

Similarly, the following identity holds

$$
\begin{aligned}
r_{-n}(t) &= \sum_{i=0}^{n-1}\sum_{j=i}^{\infty} b_j \varepsilon(t-i-j) \\
&= b_0\varepsilon(t) + (b_0+b_1)\varepsilon(t-1) + \cdots + \varepsilon(t-n+1)\sum_{i=0}^{n-1} b_i + \varepsilon(t-n)\sum_{i=1}^{n} b_i + \cdots \\
&= \sum_{i=0}^{n-1}\left(\sum_{j=0}^{i} b_j\right)\varepsilon(t-i) + \sum_{i=n}^{\infty}\left(\sum_{j=i-n+1}^{i} b_j\right)\varepsilon(t-i) \\
&= \sum_{i=0}^{\infty}\left(\varepsilon(t-i)\sum_{j=\max(0,i-n+1)}^{i} b_j\right)
\end{aligned}
\tag{3.6}
$$

which implies that

$$
\mathbb{E}[r_{+h}(t)|r_{-n}(t)] = \\
\sum_{i=h}^{\infty}\sum_{j=1}^{h} b_{i+j-h}\mathbb{E}\left[\varepsilon(t-i-j+h+1)|\sum_{i=0}^{\infty}\left(\varepsilon(t-i)\sum_{j=\max(0,i-n+1)}^{i} b_j\right)\right]. \tag{3.7}
$$

Now, using the result from Lemma 1 yields the following expression

$$
\mathbb{E}[r_{+h}(t)|r_{-n}(t)] = \frac{\sum_{i=1}^{h}\sum_{j=h}^{\infty} b_{i+j-h}\sigma^2_{t-i-j+h+1}}{\sum_{i=0}^{\infty}\sigma^2_{t-i}\left(\sum_{j=\max(0,i-n+1)}^{i} b_j\right)^2} r_{-n}(t). \tag{3.8}
$$

Now define $\Psi(h,t) = \sum_{i=1}^{h}\sum_{j=h}^{\infty} b_{i+j-h}\sigma^2_{t-i-j+h+1}$. As the denominator of the expression in Equation 3.8 will always be positive, it follows that the sign of $\mathbb{E}[\, r_{+h}(t)|r_{-n}(t)]$ will be entirely determined by the sign of $r_{-n}(t)\Psi(h,t)$. Thus if $\Psi(h,t) > 0$ for all t the process will have positive sign dependence and if $\Psi(h,t) < 0$ for all t the process will have negative sign dependence. □

Remark 4. *Notice that the assumption of Gaussian innovations in Theorem 2 was only critical in the proof to ensure that the distribution of the cumulative increments was still Gaussian. Therefore, this assumption can be relaxed to an assumption* *that partial sums of the innovations of the process follow a stable law, i.e. their distribution is such that a linear combination of variables with this distribution has the same distribution up to location and scale parameters, and a similar derivation of the same properties can be constructed that allows for heavy tailed and skewed innovations.*

Based on the results of Theorem 2, it becomes straightforward to infer in any covariance stationary fitted time series model whether there is positive or negative (or no) serial dependence in the sense of Definition 1 by finding its equivalent Wold decomposition and verifying if the fitted parameter values satisfy the derived conditions on the characteristic polynomial of AR and MA roots.

Moreover, notice that the definition of signed path dependence concerns only deviations around the process unconditional expectation - i.e. the deterministic part of the process. Also, if the process is linear, its innovations will have the same variance. As such, we can state the following corollary:

Corollary 1 (Signed Path Dependence in Linear Covariance Stationary Processes). *Let $r(t)$ be a linear covariance stationary process whose innovations follow a distribution such that a linear combination of variables with this distribution has the same distribution up to location and scale parameters. If $r(t)$ has signed path dependence of sign $\zeta \in \{-1, 1\}$ with lookback n_l and horizon h for some $n_l > 0$, then $r(t)$ will have signed path dependence of the same sign ζ with lookback n and horizon h for any integer $n > 0$.*

Remark 5. *Also notice that if these sufficient conditions in Theorem 2 are not satisfied for all values of $n \in \mathbb{N}^+$ but only a finite subset of these, then the presence of signed path dependence needs further refinement. In this way we comment that these conditions are sufficient but not necessary.*

To facilitate understanding, we conclude the section by enunciating two lemmas that link the concept of Signed Path Dependence to known parametric time series models. These lemmas demonstrate one-to-one correspondences between parameter ranges and the presence of signed path dependence. Beyond providing theoretical evidence and hence validation for our definition, these lemmas will allow us to generate data with known signed path dependence by simulating a parametric process with suitable parameter values and subsequently use in the synthetic validation experiment given in Section 3.3. The first lemma provides sufficient and necessary results for the simplest non-trivial case of time series models, the AR(1) processes. The lemma supresses the constant term by assuming that the variable being modelled has been measured as deviations from its mean.

Lemma 2 (Signed Path Dependence in a Stationary AR(1) Process of Constant Volatility). *Assume that $r(t)$ is a stationary* AR(1) *process given by*

$$r(t) = \varphi r(t-1) + \varepsilon(t) \tag{3.9}$$

with i.i.d. innovations $\varepsilon(t) \sim \mathcal{N}\left(0, \sigma^2\right)$ and $|\varphi| < 1$. Then:

(i) The process $r(t)$ has positive path dependence if and only if $\varphi > 0$.

(ii) The process $r(t)$ has negative path dependence if and only if $\varphi < 0$.

In particular, if $\varphi = 0$ the process has no signed path dependence, and all statements above are independent of n and h.

Proof. From Equation 3.9 we have that

$$r(t) = \sum_{j=0}^{\infty} \varphi^j \varepsilon(t-j). \tag{3.10}$$

Now, let $\Psi(h,t) = \sigma^2 \sum_{i=1}^{h} \sum_{j=h}^{\infty} \varphi^{i+j-h}$. It trivially follows that if $\varphi > 0$ then $\Psi(h,t) > 0$ and hence from Theorem 2 it follows that the process will have positive path dependence irrespective of n and h. Equally, if $\varphi = 0$ then $\Psi(h,t) = 0$ and as such the process will have no signed path dependence.

Also notice that

$$\Psi(h,t) = \sigma^2 \sum_{i=1}^{h} \frac{\varphi^i}{1-\varphi} = \sigma^2 \frac{\varphi\left(1-\varphi^h\right)}{(1-\varphi)^2}. \tag{3.11}$$

As whenever $0 > \varphi > -1$ we have that $1-\varphi^h > 0$, it follows that whenever $0 > \varphi > -1$, $\Psi(h,t) < 0$ irrespective of h and therefore, from Theorem 2 it follows that the process will have negative path dependence. □

The second lemma derives conditions for Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. These models generalize ARMA models by introducing long-memory features considering a non-integer differencing parameter. Note that the ARFIMA (=FARIMA) models include the ARIMA models (as ARFIMA(p,d,q) with d integer), the ARIMA models include the ARMA models (as ARMA(p,q) = ARIMA($p,0,q$)), and the ARMA models include the AR models (as AR(p) = ARMA($p,0$)), including the AR(1) of the first lemma.

Lemma 3 (Signed Path Dependence in an ARFIMA(p,d,q) Process)**.** *Assume that $r(t)$ is a purely non-deterministic* ARFIMA(p,d,q) *process with i.i.d. Gaussian innovations given by $\Phi(B)(1-B)^d r(t) = \Theta(B)\varepsilon(t)$, where Φ and Θ are polynomials of degree p resp. q in the backshift operator B, with i.i.d. innovations $\varepsilon(t) \sim \mathcal{N}\left(0,\sigma^2\right)$. Let $\sum_{j=0}^{\infty} b_j \varepsilon(t-j)$ be the Wold representation of the ARMA process $\frac{\Theta(B)}{\Phi(B)}\varepsilon(t)$ and define $\Psi(h) = \sum_{i=1}^{h}\sum_{j=h}^{\infty} b_{i+j-h}$. It holds that:*

- *(i) If $d > -0.5$, the process $r(t)$ has positive path dependence for all $h \in \mathbb{N}^+$ and $n \in \mathbb{N}^+$ if and only if $d\Psi(h) > 0$; and*
- *(ii) If $d > -0.5$, the process $r(t)$ has negative path dependence for all $h \in \mathbb{N}^+$ and $n \in \mathbb{N}^+$ if and only if $d\Psi(h) < 0$.*

Additionally, if $p = q = 0$, it holds that:

- *(i) The process $r(t)$ has positive path dependence if and only if $d > 0$.*
- *(ii) The process $r(t)$ has negative path dependence if and only if $-0.5 < d < 0$.*

Proof. According to [80], an ARFIMA(p,d,q) process can also be written as

$$\Phi(B)r(t) = \Theta(B)\left(\sum_{k=0}^{\infty}\frac{\Gamma(k+2d)}{\Gamma(k+1)\Gamma(2d)}\varepsilon(t-k)\right) \tag{3.12}$$

and therefore we have that

$$r(t) = \frac{\Theta(B)}{\Phi(B)\Gamma(2d)}\sum_{k=0}^{\infty}\frac{\Gamma(k+2d)}{\Gamma(k+1)}\varepsilon(t-k) = \sum_{k=0}^{\infty}\frac{\Gamma(k+2d)}{\Gamma(k+1)}\frac{\sum_{j=0}^{\infty}b_j\varepsilon(t-j)}{\Gamma(2d)}. \tag{3.13}$$

If $d > -0.5$ then $\sum_{k=h}^{\infty}\frac{\Gamma(k+2d)}{\Gamma(k+1)} > 0$ and $\operatorname{sgn}(\Gamma(2d)) = \operatorname{sgn}(d)$. Therefore, using the same derivation as per the proof of Theorem 2, it follows that given $h \in \mathbb{N}^+$ the sign of $d\Psi(h)$ will determine whether the process has positive or negative path dependence for all n.

Additionally, when $p = q = 0$, $\Phi(B) = \Theta(B) = 1$ and hence the process will have positive path dependence for all h and n whenever $d > 0$ and negative path dependence whenever $-0.5 < d < 0$. □

Notice that ARFIMA models can demonstrate signed path dependence if they are at least in fact in the sub-class of invertible processes and may or may not contain long memory. In the case of reversal properties of an ARFIMA model we learn that invertible models are also of interest and again these sub-class of models may or may not contain long memory properties. Most interestingly, we see that in the special case of only fractional differencing of the process, i.e. no AR and no MA components of the process, we can only have positive path dependence in the ARFIMA process if the process is invertible and it may or may not be stationary or contain long memory. In the case of the ARFIMA model with no AR and no MA components of the process, then negative dependence can only occur in sub-processes which have the properties that they dont have long memory, furthermore, they must be invertible and stationary.

3.2 Inference Under Signed Path Dependence

The models in the previous section assumed some structure for the serial correlation present in the series. We now derive a non-parametric hypothesis test for signed path dependence under certain regularity conditions. These regularity conditions are listed in Assumption 1. The test proposed assumes that these conditions hold over the entire series, but one can allow for a localised version of the test by segmenting the series being tested and testing each segment. The inference procedure proposed aims to test the following assumption (the single sided version of the test would swap the inequality sign for a greater than or less than sign depending on whether positive or negative path dependence is being tested):

***H0* (Absence of Linear Correlations between Signs and Returns)**: given n, for all $k \leq n$ the sign of the sum of the previous k observations has no linear correlation with the observation 1-step ahead, i.e. $\mathbb{E}[\text{sgn}(r_{-k}(t))r(t+1)] = 0$ for all $k \leq n$.

***H1* (Presence of some Linear Correlation between Signs and Returns)**: $r(t)$ are associated such that $\mathbb{E}[\text{sgn}(r_{-k}(t))r(t+1)] \neq 0$ for some value of $k \leq n$.

Before formally stating the not too restrictive classical regularity conditions, a definition is needed:

Definition 3 (Lindeberg Condition). *A time series $r(t)$ is said to satisfy the Lindeberg condition if for all $\epsilon > 0$ the following holds:*

$$lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_{|x|>\epsilon\sqrt{n}} x^2 df_i = lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(r(i))^2\mathbb{I}\{|r(i)|>\epsilon\sqrt{n}\}\right] = 0 \tag{3.14}$$

where f_i is the density function of $r(i)$ and $\mathbb{I}$ is the indicator function.

The intuition behind the Lindeberg condition is that, in the series being tested, the individual contribution of any observation to the sum of the variances of all observations should be arbitrarily small for a sufficiently large series.

The regularity conditions are given by the following assumption:

Assumption 1. *The time series being tested $r(t)$ is comprised of deviations from the long-term mean of a stochastic process with a well-defined mean, hence having zero mean by definition, is covariance stationary in the sense of Definition 2, satisfies the Lindeberg condition and weighted partial sums of the process are assumed to follow asymptotically a Gaussian law..*

The assumption that the Lindeberg condition is satisfied is very general and few processes of interest will fail to satisfy this condition so it is not overly restrictive in any sense. Also notice that, as the series being tested is assumed covariance stationary in the sense of Definition 2, Theorem 2 will guarantee that if this series has positive or negative path dependence for a given forecast horizon h, it will have the same property (i.e. it will be always positive or always negative) for all values of the memory n, and at all time points t. So, one can fix the forecast horizon $h = 1$ and test for signed path dependence at that forecast horizon by aggregating the time series over t (for example, testing with daily data if the forecast horizon h equals to one day, testing with weekly data if it equals to one week, and so on so forth).

Before proposing our test statistic, we will state a result on which we will base the derivation of the asymptotic distribution of our test statistic under the null hypothesis for the class of processes which are covariance stationary. For proof, see [81].

Theorem 3 (Central Limit Theorem for Covariance Stationary m-Dependent Variables)**.** *Let $R = r(1), r(2), \ldots, r(n)$ be random variables of zero mean and finite variances such that $r(t)$ is uncorrelated with $r(t+i)$ for all $i > m$, m fixed. If the variables are covariance stationary and satisfy the Lindeberg condition then, as $n \to \infty$, $\frac{\sum_{t=1}^{n} r(t)}{\sigma\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1)$ with $\sigma^2 = \mathbb{E}\left[(r(t))^2\right] + 2\sum_{k=1}^{m} \mathbb{E}[r(t)r(t+k)]$.*

The asymptotic distribution of our test statistic is given by the following theorem:

Theorem 4 (Critical Rejection Value of the Test Statistic)**.** *Given an arbitrary positive value of n, define* $d(t) = r(t)\omega(t)$ *where* $\omega(t) = \frac{\sum_{i=1}^{n} sgn\left(\sum_{k=t-i}^{t-1} r(k)\right)}{n}$ *and let* $\overline{d} = \frac{1}{s-n}\sum_{j=n+1}^{s} d(j)$*. Let* $r(t)$ *be a time series that satisfies Assumption 1 and define a random vector* $R := (r(1), r(2), \ldots, r(s))$ *where s is a given constant. When* $s \to \infty$*, under the null hypothesis of Absence of Linear Correlations between Signs and Returns the test statistic* $H(R) \xrightarrow{D} \mathcal{N}(0,1)$ *with*

$$H(R) = \frac{\overline{d}}{\sqrt{\frac{\sum_{\tau=-n}^{n}\widehat{\gamma}(\tau)}{s-n}}} \tag{3.15}$$

where

$$\widehat{\gamma}(\tau) = \frac{1}{s-n}\sum_{j=|\tau|+1}^{s-n}\left(d(j) - \overline{d}\right)\left(d(j-|\tau|) - \overline{d}\right). \tag{3.16}$$

Therefore, for a given significance threshold α*, the null hypothesis will be rejected if* $|H| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ *where* $\Phi^{-1}(x)$ *is the inverse standard normal cumulative distribution function of* x*.*

Proof. As $\omega(t+1) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{sgn}(r_{-k}(t))$, we have that

$$\mathbb{E}[d(t)] = \sum_{k=1}^{n} \mathbb{E}[\mathrm{sgn}(r_{-k}(t-1))r(t)] \tag{3.17}$$

and hence under the null hypothesis it follows that $\mathbb{E}[d(t)] = 0$.

Further, given that $r(t)$ is assumed covarance stationary, by Theorem 2 we know that the value of $\mathrm{sgn}(r_{-k}(t))\mathbb{E}[r(t+1)|r_{-k}(t)]$ is the same for all k and hence if

$$\mathrm{sgn}(r_{-k}(t))\mathbb{E}[r(t+1)|r_{-k}(t)] = 0 \tag{3.18}$$

for all $k \leq n$, then the equality will also hold for all $k > n$. Therefore, under the null hypothesis we have that

$$\mathbb{E}[\operatorname{sgn}(r_{-k}(t))r(t+1)] = 0 \tag{3.19}$$

for all values of k which implies that $d(t)$ is uncorrelated with $d(t+i)$ for all $i > n$, hence being n-dependent.

Additionally, as $0 < |\omega(t)| \leq 1$, we have that

$$\begin{aligned} \mathbb{E}\left[(d(i))^2 \mathbb{I}\{|d(i)| > \epsilon\sqrt{n}\}\right] &= \mathbb{E}\left[(r(i)\omega(i))^2 \mathbb{I}\{|r(i)| > \frac{\epsilon}{|\omega(i)|}\sqrt{n}\}\right] \\ &\leq \mathbb{E}\left[(r(i))^2 \mathbb{I}\{|r(i)| > \frac{\epsilon}{|\omega(i)|}\sqrt{n}\}\right] \end{aligned} \tag{3.20}$$

implying that $d(t)$ also satisfies the Lindeberg condition given that $r(t)$ is assumed to satisfy it.

Therefore, as per Theorem 3, we have that under the null hypothesis, when $s \to \infty$ the statistic defined in Equation 3.15 will converge in distribution to a standard normal normal variable as $\widehat{\gamma}$ defined in Equation 3.16 is the sample estimator of the variance of $\overline{d}$. □

Remark 6. *There are situations where one would be interested only to test for a specific sign of the signed path dependence. For example, if the aim is to test the suitability of trend-following strategies, one would be interested to test only for positive signed dependence. Equally, if the aim is to test the presence of mean reversion in a particular series, one would be interested to test only for negative signed dependence. These situations can be accommodated by making the test proposed in Theorem 4 one-sided and rejecting the null hypothesis if* $H > \Phi^{-1}(1-\alpha)$ *for a positive sign dependence test or if* $H < \Phi^{-1}(\alpha)$ *for a negative sign dependence test.*

Remark 7. *The test statistic defined in Equation 3.15 can also be interpreted as a correlation coefficient if normalised by the sum of all absolute returns one-step ahead instead of normalised by the sample variance, i.e. if calculated as*

$\rho = \frac{\sum_{j=n+1}^{s} d(j)}{\sum_{j=n+1}^{s} |r(j)|}$. *This coefficient will be bounded between -1 and 1, with its sign indicating the sign of the path dependence being estimated and its absolute value indicating the strength of such dependency. Section 3.5 calculates this coefficient for a number of financial assets and forecast horizons and compares its values and significance against the same metrics for the Pearson correlation of the future return against the average return over the previous n days. We chose comparing our measure against the Pearson correlation to provide a meaningful comparison between signed path dependence and classical dependence in correlation metrics. The results show that this coefficient was able to detect a significantly positive serial correlation in situations where its Pearson counterpart was not significantly different than zero.*

Remark 8. *Theorem 4 can be easily extended with relaxation of the Lindeburg condition to admit covariance stationary processes which admit an analogous α-stable limit theorem result. Interested readers are referred to derivations of such results in [82].*

Remark 9. *Theorem 4 is similar in nature to the asymptotic derivation of rejection criterion for signed rank tests. Interested readers are referred to derivations of such results in [83].*

As the test given by Theorem 4 relies on the asymptotic distribution of the test statistic under the null hypothesis, one might want to investigate how large the sample should be to obtain approximate convergence. In this subsection, an alternative test is proposed that has the same asymptotic distribution for the test statistic under the null hypothesis. This way, one can run both tests under the same simulation experimental conditions whilst increasing the sample size of the simulated data to assess the power of the test, also obtaining an indication of the speed of convergence of the asymptotic test given by Theorem 4. Such test is described in Algorithm 1. The test relies on an empirical percentile bootstrap to get a confidence interval for the test statistic under the null hypothesis of no signed path dependence and can therefore be used to check whether the value of test statistic calculated in the given sample is outside of this confidence interval, rejecting the null if this is the case.

Algorithm 1 Bootstrap procedure to test for signed path dependence

Input: $R = (r(1), r(2), \ldots, r(z))$ a sample of length z of observed log-returns, a significance threshold α and a number of simulations W

Output: TRUE if the null hypothesis of no signed path dependence is rejected with Type I error probability α and FALSE otherwise

1: Generate matrix $U \in [0,1]^{W \times z}$ with each $U_{w,t}$ drawn i.i.d. according to a discrete uniform distribution on the set $\{1, 2, \ldots, z\}$
2: Generate W random series $r'_w(t)$ of length s such that $r'_w(t) := r(U_{w,t})$
3: Construct an array B of length W so that $B(1 \leq w \leq W) = H\left(r'_w\right)$ where $H\left(r'_w\right)$ is the test statistic given by Equation (3.15) calculated over series r'_w
4: Construct $\widehat{F_B}$ the empirical cdf of B
5: Evaluate $\widehat{h_\alpha} \leftarrow inf\{x \in \mathbb{R} : \widehat{F_B}(x) \geq 1 - \frac{\alpha}{2}\}$
6: Return TRUE if $|H(R)| \geq \widehat{h_\alpha}$ or FALSE otherwise

The bootstrap test proposed is adequate to check for convergence of the asymptotic test statistic given by Theorem 4 because of the result given by Lemma 4:

Lemma 4 (Convergence of the Bootstrap Test Statistic Distribution under the Null Assumption)**.** *Let $\widehat{h_\alpha}$ be the $1 - \frac{\alpha}{2}$ empirical percentile of the bootstrap test statistic used to reject the null hypothesis of the test procedure given by Algorithm 1 at significance threshold α. For large samples ($s \to \infty$), $\widehat{h_\alpha} \to \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, where Φ^{-1} is the inverse CDF of a standard normal distribution. Therefore, for large samples, the critical rejection value of the bootstrap test statistic given by Algorithm 1 converges to the critical rejection value of the test given by Theorem 4.*

Proof. Let h_α be the $1 - \frac{\alpha}{2}$ quantile of the true distribution of the test statistic given by Equation (3.15) under the null hypothesis stated in the second paragraph of Section 3.2. As per [84] (p. 187), $\mathbb{P}\left[h_\alpha > \widehat{h_\alpha}\right] = \frac{\alpha}{2} + \frac{c}{\sqrt{s}}$ where c is a constant value. Hence when $s \to \infty$ we have that $\mathbb{P}\left[h_\alpha > \widehat{h_\alpha}\right] \to \frac{\alpha}{2}$.

Now, as per Theorem 4 the true distribution of the test statistic given by Equation (3.15) converges to a standard normal as $s \to \infty$. Therefore $h_\alpha \to \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ and hence $\mathbb{P}\left[\widehat{h_\alpha} < \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right] \to \frac{\alpha}{2}$, meaning that $\widehat{h_\alpha} \to \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$. □

It is worth noting that the rejection of the null hypothesis of the test given by Theorem 4 also implies that one can build a predictive model for the sign of return one step ahead based on the sign of the cumulative returns observed on the previous n steps. This is further explored in Section 3.5. The section compares the performance of a predictive model based on signed path dependence to make buy and sell decisions against the performance of an investment strategy known as Buy & Hold, which in essence is always buying and hence always estimating $\widehat{s}(t+1) = 1$. Theorem 5 gives the mathematical construction of the predictive model to be studied.

Theorem 5 (Mathematical Construction of the Predictive Model). *Obtain $\widehat{s}(t+1)$ by performing the following steps:*

1. *Given a log-return sample $r(1), r(2), \ldots, r(t)$ estimate the sample mean as $\overline{r} = \sum_{i=1}^{t} r(i)/t$.*
2. *Apply the test given by Theorem 4 to the given sample to detect if there is path dependence of any sign and let $\hat{\zeta} \in \{-1, 1\}$ be the sign of the path dependence detected.*
3. *If $\hat{\zeta} = -1$ choose n such that $\mathbb{E}[sgn(r_{-n}(t) - n\overline{r}) r(t+1) - \overline{r}] < 0$ and set $\widehat{s}(t+1) = -sgn(r_{-n}(t) - n\overline{r})$*
4. *Otherwise choose n such that $\mathbb{E}[sgn(r_{-n}(t) - n\overline{r}) r(t+1) - \overline{r}] > 0$ and set $\widehat{s}(t+1) = sgn(r_{-n}(t) - n\overline{r})$*

Also define the prediction loss of a predictive strategy as $L = -\widehat{s}(t+1) r(t+1)$. Note that the prediction loss of a Buy & Hold strategy is trivially given by $L_{B\&H} = -r(t+1)$. For the aforementioned strategy we have that $\mathbb{E}\left[L_{\widehat{s}(t+1)}\right] < \mathbb{E}[L_{B\&H}]$

if the null hypothesis of the test given by Theorem 4 is not true for the demeaned returns $r(1)-\overline{r}, r(2)-\overline{r}, \ldots, r(t)-\overline{r}$.

Proof. If the null hypothesis of the test given by Theorem 4 is not true then there is a value of $\hat{\zeta} \in \{-1,1\}$ such that $\hat{\zeta} = \mathrm{sgn}(\mathbb{E}[sgn(r_{-n}(t)-n\overline{r})r(t+1)]-\overline{r})$. In that case, we have that if $\hat{\zeta} = -1$ then $\mathbb{E}[-sgn(r_{-n}(t)-n\overline{r})r(t+1)] > \overline{r} \implies \mathbb{E}\left[L_{\widehat{s}(t+1)}\right] < \mathbb{E}[L_{B\&H}]$ as $\mathbb{E}[L_{B\&H}] = -\overline{r}$. Equally, if $\hat{\zeta} = 1$ then $\mathbb{E}[sgn(r_{-n}(t)-n\overline{r})r(t+1)] > \overline{r} \implies \mathbb{E}\left[L_{\widehat{s}(t+1)}\right] < \mathbb{E}[L_{B\&H}]$. □

We conclude the section by describing a predictive algorithm that attempts to forecast the sign of future asset returns using the method described in Theorem 5. Notice that the method given by Theorem 5 dynamically decides whether the prediction should be made expecting a continuation of sign ($\hat{\zeta} = 1$) or reversal of sign ($\hat{\zeta} = -1$). Given that the particular case of trading strategies that only assume continuation in sign (*trend-following*) receives special attention in Finance, see [62]. Our algorithm will cater for both cases (only continuation assumed or dynamic decision between continuation and reversals). To achieve that, we define two different predictors in Definition 4 and Definition 5:

Definition 4 (Moving Average (MA) Classifier). *An MA($n,\overline{r}$) predictor receives a log-return sample* $r = (r(1), \ldots, r(n))$ *and predicts that* $\widehat{s}(n+1) = \mathrm{sgn}(r_{-n}(n)-\overline{r})$.

Definition 5 (Dynamic Moving Average (DMA) Classifier). *A DMA($n,\overline{r},\hat{\zeta}$) predictor receives a log-return sample* $r = (r(1), \ldots, r(n))$ *and predicts that* $\widehat{s}(n+1) = \hat{\zeta} \times \mathrm{sgn}(r_{-n}(n)-\overline{r})$ *where* $\hat{\zeta} \in \{-1,1\}$.

In order to apply these methods for prediction, parameters n, $\overline{r}$ and $\hat{\zeta}$ have to be estimated. Naturally, $\overline{r}$ can be estimated as the sample mean log-return. The remaining parameters can be estimated by weighted least squares of an adequately defined error function. A simple suitable error weighted function can be given as $\varepsilon(t) = r(t)(\widehat{s}(t) - s(t))$. Also, it is clear from Theorem 5 that predictions made by the aforementioned predictors will not be useful in a process that has no signed path dependence. Therefore, before attempting to apply any of these methods, the test defined in Theorem 4 is applied to the return series used to fit the model to ascertain the presence of signed path dependence, given that in the absence of this property no prediction should be attempted. This yields Algorithm 2.

Remark 10. *Algorithm 2 implements the DMA predictor, but it can be easily modified to implement the MA predictor by fixing $\hat{\zeta} = 1$ and performing a single sided test in Step 1.*

As with any statistical model, there are multiple ways to estimate its parameters and weighted least squares is just one simple possibility. The empirical tests presented in Section 3.5 have been performed using a more sophisticated machine learning algorithm, whose detailed description is outside of the scope of the main text of this chapter. A description of the machine learning algorithm in the form of commented pseudocode is available in Section and the R implementation of the machine learning algorithm used in the empirical experiments of this chapter is available under request.

While the use of such a machine learning method is not required to establish the main results in the paper, we have chosen it for our empirical application because the use of Train and Test sets reduces potential overfitting that could come out of a simple weighted least squares procedure, and the use of a rolling model tuning stage produces better estimates of how accurately the predictive model will perform in a practical scenario where the model gets updated as new data arrives with the passage of time.

Algorithm 2 Simplified predictive method based on signed path dependence

Input: $R=(r(1),r(2),\ldots,r(z))$ a sample of length z of observed log-returns in chronological order, a significance threshold α and a test lookback horizon n'

Output: 0 if no prediction was made for $s(z+1)$, $\widehat{s}(z+1)$ otherwise

1: **if** $\left|\frac{\overline{d}}{\sqrt{\frac{\sum_{\tau=-n'}^{n'}\widehat{\gamma}(\tau)}{s-n'}}}\right| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ **then**

2: Compute $\overline{r} \leftarrow \frac{1}{z}\sum_{i=1}^{z} r(i)$

3: Compute $\hat{n},\hat{\zeta} \leftarrow \underset{\substack{n\in\mathbb{N}^+:n<<z\\ \zeta\in\{-1,1\}}}{\arg\min} \sum_{i=n+1}^{z}\left(r(i)(\zeta\times\text{sgn}(r_{-n}(i-1)-\overline{r})-\text{sgn}(r(i)-\overline{r}))\right)^2$

4: Return $\hat{\zeta}\times\text{sgn}(r_{-n}(z)-\overline{r})$

5: **else**

6: Return 0

7: **end if**

3.3 Validation of Correctness by Simulation

In this section we validate by simulation the properties of the test given in Section 3.2 by simulating synthetic time series where parameters are chosen so that positive or negative signed path dependence is guaranteed in the simulated series. As part of the validation exercise, the test defined in Section 3.2 is applied to a number of different synthetic series and used to detect the simulated dependence. Such results can be used to infer the robustness of the test statistic against the following factors:

1. Different assumptions in the functional form of the input time series;
2. Strength of the dependence present in the input time series, including the effect of long-range dependence;
3. Time varying variance in the input time series when the required assumption of covariance stationarity is met; and
4. Time varying variance in the input time series when the required assumption of covariance stationarity is not met.

Robustness against the aforementioned factors in a controlled environment can give assurances for the suitability of the test in empirical applications, where the actual functional form of the input time series is not known and the required assumption of covariance stationarity might not necessarily be met. Moreover, comparing different levels of Type I and Type II error probabilities across all validation scenarios can give useful insights on which properties have higher impact on the accuracy of the test and any potential biases.

The goals of the present section are to demonstrate that the previous derivations are correct and to get an idea of the power of the test under different serial correlation structures. The Type I error probabilities as function of significance threshold reported in Figures 3.1 to 3.6 are not expected to stay always near the identity line as our simulation is intentionally mixing several different strenghts of serial dependence in the cases being tested, generating very non-homogeneous samples. Under such circumstances, the test might be conservative and have actual Type I error probabilities that are much lower than the desired significance threshold, but this conservatism will be adequate if at all times the test does not drive the user to wrong conclusions by having Type I error probabilities that exceed the identity line (i.e. a proportion of false rejections of the null hypothesis greater than the desired significance threshold). In Section 3.4 it is demonstrated with homogeneous samples that the Type I error probabilities of the test here proposed stays very close to the identity line in the scenarios where it should be.

We start our synthetic examples by the simplest case of the AR(1) models. As per Lemma 2, we know that the sign of the dependence for an AR(1) process depends entirely on the sign of the parameter ϕ, and all series $r_j(i)$ have positive path dependence when $\varphi_j > 0$ while all other series will not have so. Therefore the first synthetic validation experiment performed by us is given by the steps below:

Experiment 1. *Perform the following steps to illustrate Lemma 2:*

1. *Generate* $P = 1000$ *AR(1) series of length* $z = 1000$ *so that* $r_j(i) = \varphi_j r_j(i-1) + \varepsilon_j(i)$ *with* $r_j(1) = \varepsilon_j(1)$, $0 < i \leq z$, $0 < j \leq P$, $\varphi_j = 4\left(\frac{j}{5P} - \frac{P+1}{10P}\right)$ *and* $\varepsilon_j(i) \sim \mathcal{N}(0,1)$.
2. *Fix the significance threshold* α *and apply the single sided version of the asymptotic test to detect positive path dependence as per Remark 6 to all the P series.*
3. *Use the known presence of positive dependence to calculate the type I and type II error probabilities respectively by counting how many of the series where* $j \leq \frac{P+1}{2}$ *had the null hypothesis rejected and how many of the series where* $j > \frac{P+1}{2}$ *had the test failing to reject the null hypothesis.*
4. *Perform Steps 2 and 3 using the bootstrap test defined in Algorithm 1 to detect positive path dependence (as a control case).*
5. *Repeat the procedure for different values of* α.

Figures 3.1 to 3.6 show the behaviour of type I and type II errors for the bootstrap and asymptotic tests as a function of the significance threshold for the procedure given by Experiment 1. Notice that it is not adequate to infer that the test has low power solely due to Type I errors in these figures being lower than the significance threshold. The Type I errors in these figures are lower than the significance threshold because the data supplied to the experiment is, by construction, including simulations from several different generative processes. For example, in the case of the AR(1) test, the Type I error of the test stayed nearly constant around 0.002 (as per Figure 3.1) because out of 500 AR(1) series with ϕ coefficients varying from -0.4 to 0, in only one case the test produced a false positive (i.e. failed to reject the null hypothesis that $\phi > 0$). This not a shortcoming in the test (arguably this can be taken as a positive outcome instead). The significance threshold won't necessarily be close to the ratio of false positives in the tests made in Section 3.3 because the test statistic under the null is derived considering the specific case where $\phi = 0$, and not for varying levels of ϕ. One would expect to get approximately 25 false positives out of 500 AR(1) series only if all the 500 AR(1) series supplied had $\phi = 0$. This comparison is only made in Section 3.4 and, as Table 3.1 shows, 23 false positives were obtained (0.046 Type I error for a 0.05 significance threshold).

The goal of the current section is not to ascertain if the test has low or high power but instead is to certify that the test will not have completely unexpected behaviour in extremes such as very negative values of ϕ for an AR(1) model. Table 3.1 gives a much clearer picture of test conservatism and as the fourth column of the table shows, when $\phi = 0$ the ratio of false positives was mostly in line with the significance threshold, as expected.

Figure 3.1 shows the behaviour of type I and type II errors for the bootstrap and asymptotic tests as a function of the significance threshold for the procedure given by Experiment 1. The two tests converged and the type I error probability of the asymptotic test was not statistically different than the one of the bootstrap test. The type II error probabilities of both tests are also not statistically different than each other and always lower than 50%, suggesting that even though the test is conservative, for the AR(1) series constructed it will fail to reject the null when the alternative hypothesis truly holds only in the minority of the cases.

The model has also been validated by simulating an ARFIMA(0,d,0) series. Lemma 3 gives the condition under the forecast horizon $h = 1$ where the sign of the path dependence for this model class is known for the whole series: the sign of the parameter d. To keep the simulation procedure simple, we have reused the procedure given by Experiment 1 and in Step 1 we constructed the series based on a single parameter $-0.1 \leq d \leq 0.1$, with the only modification being that the series generated in Step 1 are ARFIMA(0,ϕ,0) series (instead of AR(1)).

Figure 3.2 shows the behaviour of type I and type II errors for the bootstrap and asymptotic tests as function of the significance threshold for the simulated ARFIMA(0,d,0) series. The type I error probability of the asymptotic test in this case is also not statistically different than the one of the bootstrap test. The type II error probabilities of both tests are also not statistically different than each other and always lower than 25%, suggesting that in the large majority of the cases it will not fail to reject the null when the alternative hypothesis truly holds, which can be seen as a positive trait in the test.

The synthetic experiment for the ARFIMA(0,d,0) model yielded conclusions similar to the ones obtained for the AR(1) model, with the difference being that both tests were considerably more powerful. This suggests the long-memory feature of the ARFIMA model introduces a stronger signed path dependence to the modelled time series than the one introduced by an AR(1).

In figures 3.1 to 3.6, the solid lines represent the simulated error probabilities and the dashed lines represent a 95% confidence margin of the error probability of each test, obtained via a Poisson approximation for the count of false acceptances/rejections. The chart on the left shows the simulated type I error probabilities of the asymptotic and bootstrap statistics as functions of the significance threshold when these tests are applied to the generated series. The black straight line corresponds to the identity line, where the test type I error is exactly equal to the desired significance threshold. The chart on the right shows the simulated type II error probabilities of the two tests as functions of the significance threshold with the dashed lines being the 95% confidence margin.

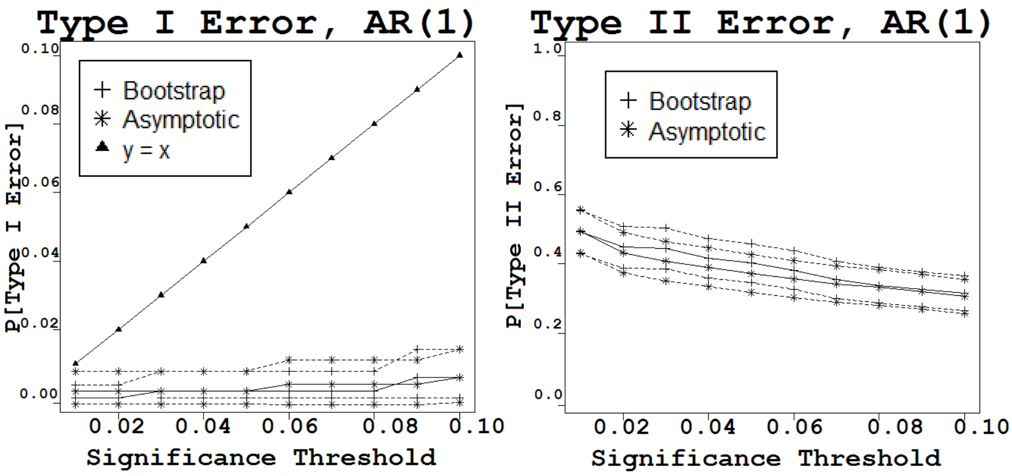

Figure 3.1: Performance of the tests in the simulated AR(1) series.

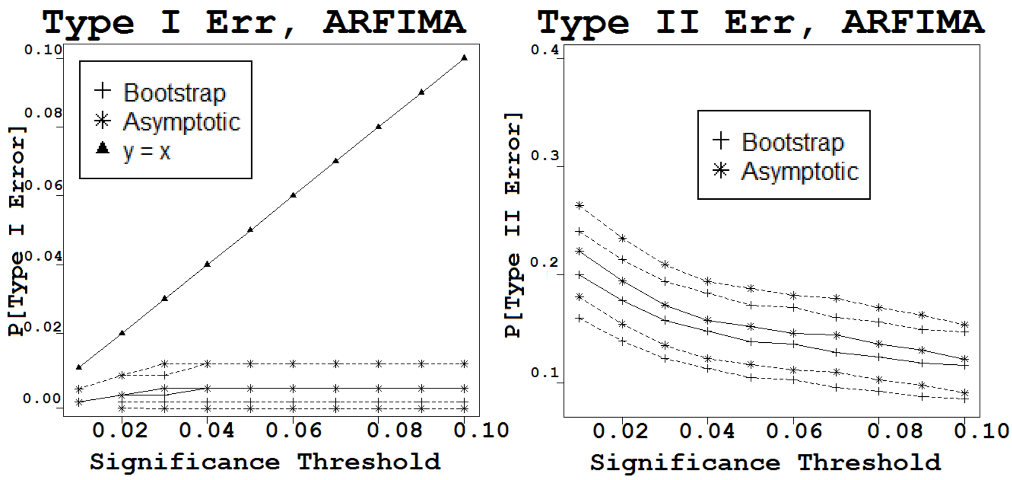

Figure 3.2: Performance of the tests in the simulated ARFIMA(0,d,0) series.

Finally, to ascertain some boundary conditions for the applicability of our tests, we have created a synthetic MA(n) series and analysed the behaviour of the test by modifying it so that it has a) a very weak correlation structure; b) a non-constant volatility that follows a stationary GARCH process; or c) a non-constant and non-stationary volatility that is driven by a two-state Markov-switching process.

To build the baseline MA(n) series we note that, as per Theorem 2, when the forecast horizon $h = 1$ we have that the sign of the sum of the MA coefficients will determine the sign of the path dependence of the model. We again have reused the procedure given by Experiment 1 but in Step 1 we generated MA(5) series based on a single parameter $-0.1 \leq \phi \leq 0.1$ so that the coefficients of all 5 lags were equal to $\frac{4}{5}\phi$. The synthetic experiment for the MA(5) model yielded conclusions similar to the ones obtained for the AR(1) model, with the difference being that the test was slightly less powerful. Figure 3.3 shows the behaviour of type I and type II errors for the bootstrap and asymptotic tests as function of the significance threshold for the simulated MA(5) series. Like in the AR(1) case the tests converged, with both the type I and type II error probabilities of these tests not being statistically different amongst themselves. Additionally, the type II error probabilities of both tests was still lower than 50% when the significance threshold was 0.05 or higher.

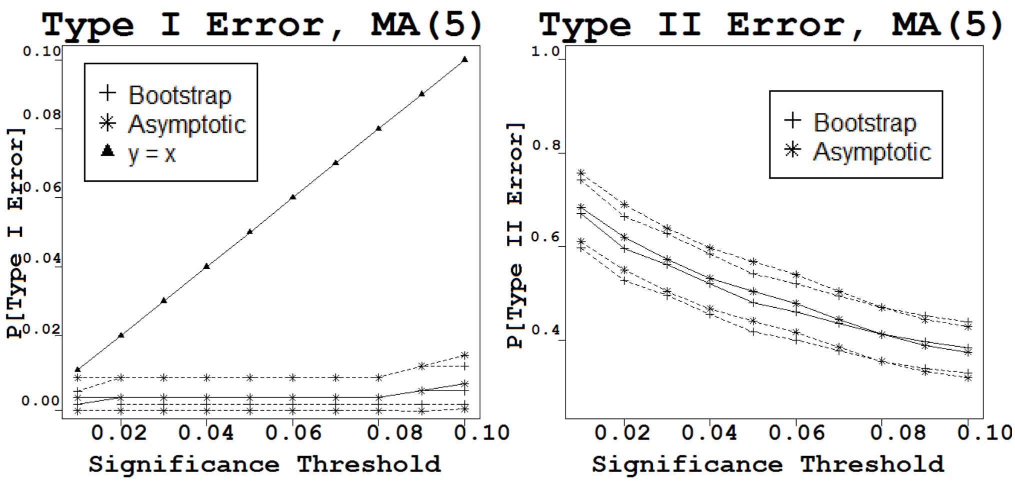

Figure 3.3: Performance of the tests in the baseline simulated MA(5) series.

After generating the baseline case, we created a stressed MA(5) simulation where series length is reduced considerably and the correlation structure is weakened. In this stressed MA(5) simulation, the procedure given by Experiment 1 was reused but in Step 1 we generated MA(5) series based on a single parameter $-0.1 \leq \phi \leq 0.1$ so that the coefficients of lags 1, 3 and 5 were equal to ϕ and the coefficients of lags 2 and 4 were equal to $-\phi$. In this situation, the average sum of all MA coefficients when the alternative hypothesis holds was still positive, equal to 0.05, though the alternating signs of the lags weaken the overall positive dependence. Further, the procedure generated series of shorter length $z = 100$.

Figure 3.4 shows the behaviour of type I and type II errors for the bootstrap and asymptotic tests as function of the significance threshold for the stressed MA(5) series. The type I error probability of the asymptotic test is always greater than the one of the bootstrap test for a given significance threshold; when the significance threshold is lower than 0.05 it is also above the identity line and for a significance threshold lower than 0.015 the identity line falls below the 95% confidence interval of the error probability for the asymptotic test though stays very close to its lower boundary. At the same time, the type II error probability of the asymptotic test is always lower than the one of the bootstrap test for a given significance threshold, though it is always higher than 90%, suggesting the test will conservatively fail to reject the null in many genuine cases that the alternative hypothesis truly holds.

It is worth noting that the reduced power in this scenario is a consequence of the intentionally weak correlation structure imposed and smaller sample. However, it is reassuring to observe in Figure 3.4 that even with a small sample (of only 100 observations) the asymptotic test did not generate incorrect results. Nevertheless, in the small sample the asymptotic test provided up to 2% more false rejections than the bootstrap version of the test and its power approached the one of the bootstrap test only at a significance threshold of 0.08 or higher. As empirical applications are expected to have a weak correlation structure, whenever possible the bootstrap version of the test should be used instead of the asymptotic version for smaller samples.

Despite being subject to series of very weak correlation structure, the asymptotic test still behaved reasonably, with adequate conservatism for all but very low significance thresholds ($\alpha \leq 0.02$), when the lower 95% confidence bound goes above the identity line. The bootstrap test remained adequate even for this low significance threshold, though at the cost of an even reduced power.

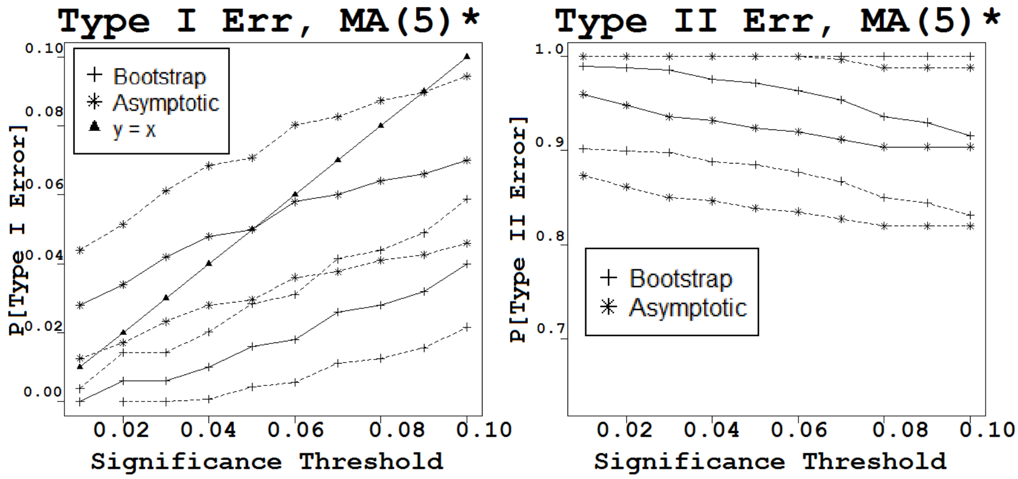

Figure 3.4: Performance of the tests in the stressed MA(5).

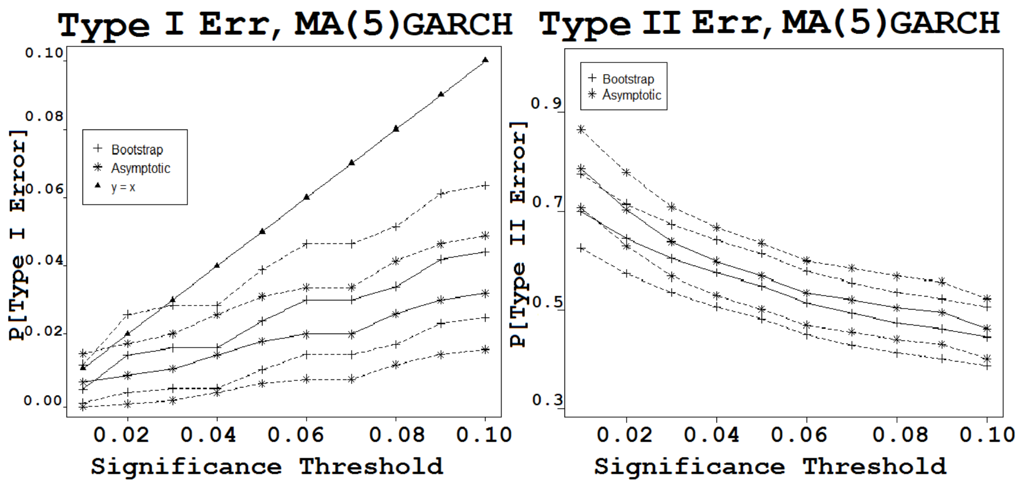

Figure 3.5: Performance of the tests in the MA(5)-GARCH(1,1) series.

All simulations so far assumed constant volatility. In empirical applications, volatility is known vary over time and to exhibit serial dependence ([25]). To verify the effect of a non-constant volatility, we have changed the baseline MA(5) case so that the error variance follows a GARCH(1,1) process with coefficients $\alpha = 1$ (the constant term), $\beta = 0.89$ (the term that multiplies the lagged squared error) and $\gamma = 0.1$ (the term that multiplies the lagged variance). The choice of

GARCH parameters ensured the series was still unconditionally covariance stationary. All other parameters of the baseline MA(5) model remained unchanged.

We note that, while the power of the test reduced slightly compared to the baseline case, comparing Figures 3.4 and 3.5 one can see that the impact of a GARCH volatility in the power of the test was smaller than the impact of reduced correlation structure. The test remained very adequate even in the presence of GARCH volatility and Type I errors stayed either within or under the identity line for all significance thresholds.

Finally, to ascertain a boundary scenario where the main assumption of covariance stationarity is not strictly satisfied, we have changed the baseline MA(5) case so that the error variance follows a Markov Switching between two equally probably GARCH(1,1) states. Such framework is widely popular in financial market research and is accepted to be a relevant deviation from time series models that has been extensively observed in empirical applications. In one of the states, the error variance follows a GARCH(1,1) process with coefficients $\alpha = 1$, $\beta = 0.89$ and $\gamma = 0.1$ and in the other state the error variance follows a GARCH(1,1) process coefficients $\alpha = 3$, $\beta = 0.95$ and $\gamma = 0.049$. All other parameters of the baseline MA(5) model remained unchanged.

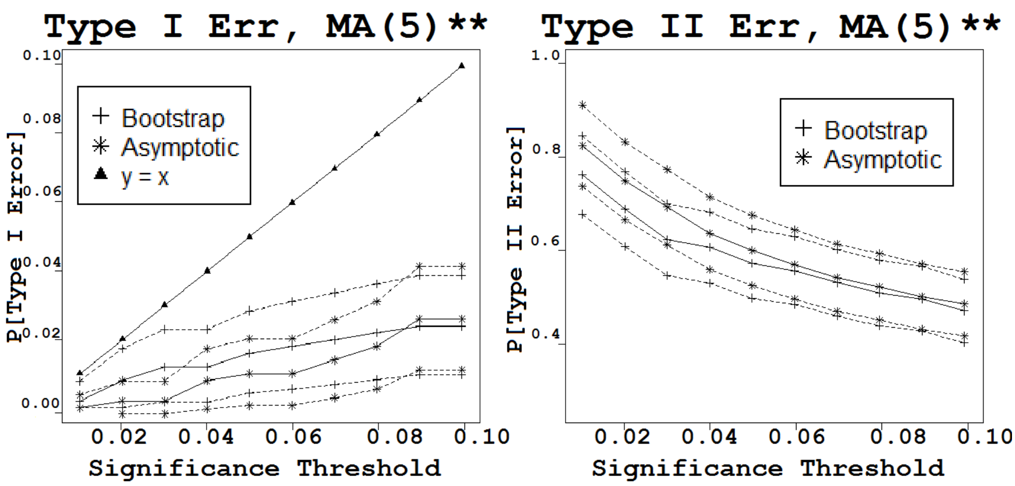

Figure 3.6: Performance of the tests in the MA(5)-RSGARCH(1,1) series.

We again note that, while the power of the test reduced slightly compared to the baseline case, comparing Figures 3.4 and 3.6 one can see that the impact of a GARCH volatility in the power of the test was smaller than the impact of reduced correlation structure. The test remained very adequate even in the presence of Regime Switching GARCH volatility and Type I errors stayed either within or under the identity line for all significance thresholds, indicating adequate test behaviour in stressed parameters.

It is noted that the testing procedure described in this section is similar to the one performed in [85]. In that paper, the author proposed a test to detect whether there was serial correlation present in a series, without inferring on the shape of that serial correlation. Section 6 of this paper shows Monte Carlo simulations of a) white noise processes of normally distributed errors, b) white noise processes of non-normally distributed errors, c) AR(1) processes and d) ARIMA(0,0.35,0) processes. The author demonstrated that the test proposed was able to successfully identify serial dependence in cases c) and d) without suffering from identification errors in cases a) and b).

The sign path dependence test, however, is different from the test in [85] in the sense that the test in [85] had a higher number of parameters and performed transformations to and from the frequency domain in order to obtain conclusions about possible serial dependence in the series being tested, while the sign path dependence test did not have to resort to the frequency domain and only required a simple parameter (a large lookback window), offering a better intuitive explanation of the results and procedures than the test proposed in[85].

The section is concluded by noting that Lemma 4 has also been tested in the synthetic experiments and the experiments demonstrated that the distribution of the test statistic under the null for the asymptotic and bootstrap tests converged as the size of the tested samples increased, as predicted by the Lemma. Figure

3.4 illustrates the two distributions with a tested sample size of 100 observations and it is noticeable that the confidence intervals of the Type I and Type II error frequencies do not coincide in this small sample. However, Figure 3.3 shows the same two tests now applied to a tested sample size of 1000 observations and the confidence intervals for the Type I and Type II frequencies are much more similar; in fact, the Type II error confidence bands for the asymptotic and bootstrap tests almost coincide.

3.4 Model Specification Error Sensitivity Analysis

While the previous section provided good information about the correctness of the test under several different generative models, its inhomogeneous nature gave less infomation about the conservatism and power of the test. These features will be analysed in the present section with the aid of the following experiment.

Let x_t be the data at time t, which is known to follow an AR(1) process given by

$$x_t = \alpha + \varphi x_{t-1} + e_t$$

where e_t follows a Standard Gaussian Distribution.

According to Lemma 2, if the underlying data being tested is known to follow an AR(1) process, the test here proposed is equivalent to testing for the sign of coefficient φ. To check if the test here proposed is excessively conservative from the perspective of the Type I error probabilities, one can compare the Type I and Type II error probability profile between the test here proposed and a Wald test applied to the maximum likelihood estimates of φ.

Let's first review the formulation of the Wald test. Let $\hat{\varphi}$ be the Maximum Likelihood Estimate (MLE) of φ and formulate the following null hypothesis:

H_0: $\varphi = \varphi_0$

Under the null hypothesis, the Wald statistic given by

$$W = \frac{(\hat{\varphi} - \varphi_0)^2}{s_\varphi^2}$$

follows a χ^2 distribution with one degree of freedom (note that s_φ^2 is an estimator of the variance of the MLE, which can be obtained using the inverse observed Fisher Information matrix).

Our goal, however, is to perform a single sided test: our null hypothesis is actually

H_0: $\varphi \leq \varphi_0$ with $\varphi_0 = 0$.

Because we are testing a single parameter, we can rely on the fact that the square root of a random variable that follows a χ^2 distribution with one degree of freedom is a random variable that follows a standard Gaussian distribution and work on the square root of the Wald statistic, i.e.:

$$\sqrt{W} = \frac{\hat{\varphi}}{\sqrt{s_\varphi^2}},$$

which will follow a standard Gaussian distribution under the null assumption.

The procedure to compare the Type I and Type II error probability profiles of the Wald test (single sided modification) and our test is the same given in Section 3.3 but now modified to have just two homogeneous samples, one that is uncorrelated (i.e. $\varphi = 0$ and one that has a single value of $\varphi = 0.1$, yielding the following procedure:

Experiment 2. *Perform the following steps:*

1. *Generate $P = 1000$ AR(1) series of length $z = 1000$ so that $r_j(i) = \varphi_j r_j(i-1) + e_j(i)$ with $r_j(1) = e_j(1)$, $0 < i \leq z$, $0 < j \leq P$, $e_j(i) \sim \mathcal{N}(0,1)$, $\varphi_j = 0$ for $j \leq \frac{P+1}{2}$ and $\varphi_j = 0.1$ for $j > \frac{P+1}{2}$.*
2. *Fix the significance threshold α and apply the single sided version of the*

asymptotic test to detect positive path dependence described in the thesis to all the P series.

3. *Use the known presence of positive dependence to calculate the type I and type II error probabilities respectively by counting how many of the series where $j \leq \frac{P+1}{2}$ had the null hypothesis rejected and how many of the series where $j > \frac{P+1}{2}$ had the test failing to reject the null hypothesis.*
4. *Perform Steps 2 and 3 using the one sided Wald test as a control case.*
5. *Repeat the procedure for different values of α.*

Experiment 2 yielded the probabilities shown in Table 3.1. As the table shows, under the homogeneous samples provided the type I error probabilities for the Wald test and for the asymptotic signed path dependence test behaved as expected, with values that are close to the nominal significance level. Also as expected, the Wald test was more powerful than the asymptotic signed path dependence test as it was designed to work with the assumption that the input data was an AR(1) process.

However, in practical situations one will not know in advance what is the type of serial dependence that the data series being tested truly exhibits. Choosing a test that works very well for an AR(1) process and using it for an input data series that series that might resemble superficially an AR(1) process could lead the user to incorrect conclusions if the true serial dependence structure in the data is not AR(1), as the following experiment will demonstrate.

Let's remember that as per Theorem 2, when the forecast horizon $h = 1$ we have that the sign of the sum of the MA coefficients will determine the sign of the path dependence of the model. We have now reused the procedure given by Experiment 1 but in Step 1 we generated MA(5) series based on a single parameter $-0.1 \leq \phi \leq 0.1$ so that the coefficients of all 5 lags are of the form $(b_1, b_2, b_3, b_4, b_5) = (-2, -1, 1, 2, 2)\phi$, where b_i is the i-th coefficient of the Wold representation of the MA(5) process.

Under those circumstances, $\sum_{i=1}^{5} b_i = 2\phi \Rightarrow \mathrm{sgn}\left(\sum_{i=1}^{\infty} b_i\right) = \mathrm{sgn}(\phi)$, meaning that the series will have positive (negative) sign dependence when ϕ is positive (negative). However, as the coefficient of the first lag is -2ϕ, an AR(1) model fit will normally have a coefficient that has the opposite sign of the true sign of the dependency, which comes from the higher lags.

Running this experiment and comparing the Type I and Type II errors of the Wald test against the asymptotic signed path dependence test yielded the results in Table 3.2. As the table shows, the Wald test performed badly, with actual Type I error probabilities greater than 65% for all significance thresholds tested - i.e. worse than if a coin toss was used as the criterion to reject the null hypothesis. At the same time, the asymptotic signed path dependence test was conservative due to the inhomogeneous nature of the given input samples of Experiment 1 (homogeneous samples were only used in Experiment 2, Table 3.1) but always lower than the nominal significance level, hence not suffering from model misspecification errors in this particular example. Moreover, Table 3.2 shows that the signed path dependence test was also more powerful than the Wold test under all significance thresholds when faced with a model specification problem.

Table 3.1: Comparison of Type I and Type II Error Probabilities given a Significance Threshold for the AR(1) Series

Signif. Threshold	Type I Err Wald	Type II Err Wald	Type I Err Sign Path Dp	Type II Err Sign Path Dp
0.01	0.010	0.208	0.010	0.890
0.02	0.018	0.134	0.018	0.842
0.03	0.022	0.104	0.030	0.806
0.04	0.036	0.080	0.034	0.742
0.05	0.044	0.070	0.046	0.704
0.06	0.056	0.060	0.050	0.682
0.07	0.066	0.054	0.060	0.654
0.08	0.074	0.046	0.072	0.630
0.09	0.078	0.040	0.078	0.604
0.10	0.092	0.036	0.088	0.580

Table 3.2: Comparison of Type I and Type II Error Probabilities given a Significance Threshold for the MA(5) Series

Signif. Threshold	Type I Err Wald	Type II Err Wald	Type I Err Sign Path Dp	Type II Err Sign Path Dp
0.01	0.658	1.000	0.002	0.976
0.02	0.700	0.998	0.004	0.958
0.03	0.722	0.996	0.004	0.932
0.04	0.748	0.994	0.012	0.926
0.05	0.756	0.994	0.016	0.906
0.06	0.768	0.992	0.020	0.894
0.07	0.774	0.988	0.028	0.878
0.08	0.788	0.986	0.030	0.866
0.09	0.794	0.986	0.036	0.862
0.10	0.806	0.986	0.038	0.854

3.5 Empirical Applications

This section applies to real financial data the test given by Theorem 4 and validates the performance of trading strategies created based on Theorem 5. The dataset used comprised of the time series of the prices for the MSCI Equity Total Return Indices (in US Dollar, price returns plus dividend returns net of withholding taxes) for 21 different countries and the nominal exchange rates of 16 different currencies, all of them traded against the US Dollar. Table **??** lists the equity indices and currencies studied.

For all the 37 financial instruments the daily log-returns (defined as the logarithm of the closing price at the end of day d+1 divided by the closing price at the end of day d) were obtained based on the price series from 31-Dec-1998 to 31-May-2017, yielding 4,803 observations per instrument (and 177,711 observations in total). To conduct further analyses at different levels of observational noise, several other sampling frequencies were studied. The same log-returns were aggregated to weekly log-returns (defined as the logarithm of the closing price at the end of the Friday of week w+1 divided by the closing price at the end of the Friday of week w), yielding 961 observations per instrument (and 35,557 observations in total); the returns were also aggregated to monthly log-returns (defined as the logarithm of the closing price at the end of the last business day of month m+1 divided by the closing price at the end of the last business day of month m), yielding 221 observations per instrument (and 8,177 observations in total). Each of the MSCI Index studied currently has at least one tradable Futures contract linked to it and many of them have also shortable ETFs linked to them, which means the any predictive model resulting out of the experiments performed in this paper can also be used to guide real equity investment strategies.

In the case of exchange rates, to accurately reflect the total returns of the investment strategies proposed, the results of the experiment include the financial returns arising out of the interest rate differential between the issuing country and the US (also known as the "rollover interest") for each open position and also include the interest paid on the cash deposit held as margin for the currency trade.

Before testing any trading strategies, the first half of the sample described in the second paragraph of this section (corresponding to observations from 31-Dec-1999 to 31-Mar-2008) was used to calculate the test statistic defined in Equation 3.15 and detect the presence of significant serial dependence in individual assets. Tables **??** and **??** show the values of the serial correlation coefficient calculated as Remark 7 for forecast horizons of one day, one week and one month ($n = 5$ in all cases).

The statistical significances in the tables were obtained using the test procedure described in Theorem 4. For control, the tables also display the Pearson correlation coefficient between the average return of the n previous periods and the 1-step ahead return. The significances of the Pearson correlation were estimated using a Fisher transform and its limiting Gaussian approximation. The rationale for applying the test only on the first half of the sample is to ensure there is no fitting bias in the trading model, whose performance is evaluated using the second half of the sample and whose equities and currencies traded are selected based on the test results.

Table 3.3: Different correlation metrics for Equity Indices in the first half of the sample

Country	Daily		Weekly		Monthly	
	Rho	Pearson	Rho	Pearson	Rho	Pearson
AUSTRALIA	0.0193	-0.0248	0.0067	-0.0108	-0.0303	-0.0747
BRAZIL	0.0851***	0.0218	0.0292	0.0072	0.1305**	-0.0567
CANADA	0.0210	-0.0185	-0.0213	0.0120	0.1414**	0.1321*
CHILE	0.1642***	0.0978***	0.1012**	0.0449	0.1274	0.1054
CHINA	0.0871***	0.0316*	0.0916**	0.0560	0.1914***	0.0459
COLOMBIA	0.2021***	0.0972***	0.2152***	0.1301***	0.0590	-0.0083
CZECH	0.0557***	0.0051	0.0989***	0.0469	0.0447	-0.0353
FRANCE	-0.0327**	-0.0710***	-0.0271	-0.0278	0.1684***	0.0308
GREECE	0.0809***	-0.0129	0.0445	0.0261	0.2160***	0.0845
HONGKONG	0.0418***	-0.0045	0.0773**	0.0326	0.1407*	0.0200
HUNGARY	0.0671***	0.0250	0.0699**	0.0262	0.0303	-0.0038
INDIA	0.1186***	0.0540***	0.1379***	0.0750*	0.1145**	0.0011
JAPAN	-0.0177	-0.0271*	-0.0190	-0.0211	0.2458***	0.2195**
MEXICO	0.0669***	-0.0045	0.0079	-0.0164	0.0253	-0.1339*
NEWZEALAND	-0.0118	-0.0358**	0.0437	-0.0388	-0.0086	0.0376
RUSSIA	0.0661***	-0.0003	0.0537**	-0.0162	-0.0572*	-0.0973
SINGAPORE	0.0628***	0.0276*	0.0774*	0.0497	0.2127**	0.0586
SOUTHAFRICA	0.0334**	0.0004	0.0266	0.0171	-0.0114	-0.1463*
TAIWAN	0.0546***	0.0255	0.0123	0.0038	0.0852**	0.0143
UK	-0.0450***	-0.1026***	-0.0249	-0.0532	0.1575**	0.0848
USA	-0.0405***	-0.0668***	-0.0456	-0.0959**	0.0603*	0.0377

* Significantly different from zero at 10% significance threshold
** Significantly different from zero at 5% significance threshold
*** Significantly different from zero at 1% significance threshold

[1] *Daily*, *Weekly*, *Monthly* represent the frequency the return was observed in order to calculate the correlation metrics

Table 3.4: Different correlation metrics for Currencies in the first half of the sample

Country	Daily		Weekly		Monthly	
	Rho	Pearson	Rho	Pearson	Rho	Pearson
AUDUSD	0.0107	-0.0125	0.0347	-0.0042	0.0108	-0.0086
BRLUSD	0.0791***	0.0539***	0.1283***	0.0866**	0.2025**	0.1113
CADUSD	-0.0291*	-0.0319*	0.0041	0.0280	0.0252	0.0033
CHFUSD	-0.0279*	-0.0156	0.0123	0.0024	-0.0169	-0.0516
CZKUSD	0.0116	0.0152	0.0407	0.0391	0.0409	-0.0537
EURUSD	-0.0112	0.0015	0.0312	0.0502	0.1393**	-0.0149
GBPUSD	-0.0193	-0.0033	0.0121	-0.0375	-0.0559*	-0.0052
HUFUSD	-0.0196	-0.0299*	-0.0239	-0.0012	0.0046	-0.0210
INRUSD	-0.0651*	-0.0657***	0.2311***	0.1995***	0.2657***	0.0942
JPYUSD	-0.0107	-0.0271*	-0.0247	-0.0245	0.0504	0.0075
MXNUSD	-0.0077	-0.0392**	0.0531	-0.0274	-0.0674	-0.1037
NOKUSD	0.0077	-0.0055	-0.0049	-0.0025	0.0069	-0.0667
NZDUSD	0.0043	-0.0199	0.0142	0.0046	0.1042	-0.0051
PLNUSD	0.0576***	0.0275*	0.0292	0.0357	-0.0802	-0.0653
SEKUSD	-0.0011	-0.0165	0.0330	0.0355	0.0411	0.0272
ZARUSD	0.0005	-0.0519***	0.0868*	0.0467	0.1157	0.0712

* Significantly different from zero at 10% significance threshold
** Significantly different from zero at 5% significance threshold
*** Significantly different from zero at 1% significance threshold

[1] *Daily*, *Weekly*, *Monthly* represent the frequency the return was observed in order to calculate the correlation metrics

It can be seen that the correlation coefficient proposed in this paper and its associated test are more powerful than the Pearson estimate. Notably, only in the case of New Zealander equities that the Pearson estimate was able to detect significant correlation without our proposed method not simultaneously finding any significant correlation, whilst for many equity indices only our proposal could find a significant coefficient when the Pearson estimate was not significantly different than zero. Another interesting finding is that even though most of the equity indices had positive dependence, this was not always the case. For example, US, UK and French equities bucked the trend and exhibited significant negative dependence at a one-day forecast horizon and Russian equities had significant negative dependence at a one-month forecast horizon. Interestingly enough, the markets that had significant negative one-day dependence had significant positive one-month dependence and the market that had significant negative one-month dependence had significant positive one-day dependence.

It can also be seen that the procedure proposed in this paper detected dependence in a mix of emerging and developed market currencies, with significant one-day dependence being found in the Brazilian Real, Indian Rupee, Polish Zloty, Canadian Dollar and Swiss Franc in addition one month-dependence being found in the Euro and British Pound. Our proposal detected dependence in more currencies than the Pearson estimate could, though the Pearson estimate detected one-day dependence on the Japanese Yen and Mexican Peso that was undetected by our proposal. However, in most cases that the dependence found is statistically significant it is of lower magnitude than with Equities and therefore less pronounced and likely to be less exploitable for economic purposes.

To validate the power of our predictive framework, the machine learning version of Algorithm 2 (described in Section 3.6) has been applied to the data described in the beginning of this section and the predictions were translated into investment decisions: buy (when the predicted sign is positive) and sell (when the predicted sign is negative). The trading strategies were tested assuming the following: (i) there are no short sale restrictions - which is most likely the case if only futures are used to trade; and (ii) if a shortable ETF is used, its borrow cost is no more than the interest on the cash received from the short sale proceeds. Initially the markets are assumed frictionless but this assumption is relaxed at the end of this Section by performing a sensitivity analysis to transaction costs.

To ascertain the out-of-sample performance of the trading models proposed, five different performance metrics were calculated for all strategies and compared against the same values of a Buy & Hold benchmark portfolio. The Buy & Hold benchmark portfolio consisted of an equal weighted long position across all 21 equity indices in the case of equity strategies and an equal weighted long position across all 16 foreign currencies in the case of currency strategies. As the rollover rates are being included in the calculation of the returns, the currencies benchmark can be interpreted as a global unhedged money market investment portfolio from a US investor's perspective.

The peformance metrics used were:

- Average monthly out-of-sample returns of a portfolio that invested equally weighted across all equities and currencies according to the underlying classification strategy, per asset class per classifier;
- Standard Deviation of the out-of-sample monthly returns of the said portfolio;
- Sharpe Ratio, not adjusted by the risk-free rate: simple average of the out-of-sample monthly returns divided by their respective standard deviations; and
- Maximum Drawdown: the maximum loss from a peak to through of the cumulative monthly portfolio returns before a new peak is attained.
- Alpha, not adjusted by the risk-free rate: the intercept of a linear regression between the returns of the strategy and the returns of the Buy & Hold

portfolio.

The returns used to calculate the Sharpe Ratio and Alpha were not adjusted by the risk-free rate to keep the implementation simple and avoid an extra choice of variable (i.e. the choice of what constitutes a risk-free rate). The lack of adjustment does not change the conclusions of the present Chapter given that the risk free rate has been very low or close to zero in most developed economies since the Great Financial Crisis of 2008, which is the vast majority of the out-of-sample period used for this experiment. All performance measures are also accompanied by their standard errors calculated using the leave-one-out Jackknife method. Additionally, the Alpha was tested for statistical significance and the Type I Error probability of the Hypothesis Test that the Alpha is greater than zero is shown on all comparative performance tables.

The trading strategy given by Theorem 5 has been empirically tested using the MA and DMA predictors to obtain out-of-sample predictions for the period corresponding to the months from April-2008 to May-2017, equivalent to the second half of the entire dataset used in this study. As our predictive method only predicts one step ahead, after a prediction is made the in-sample period gets the actual return observed one step ahead appended to it whilst the first in-sample point is discarded, so that in the next iteration the algorithm is over the new in-sample period. This is repeated until the end of the out-of-sample period. Tables 3.5 to 3.10 list the returns gross of transaction costs; however, we later incorporate transaction costs into our analysis.

The strategies proposed obtained their best performance when trading equities using a one-day forecast horizon. As Table 3.5 shows, the MA strategy was able to get a monthly return that was more than three times the return of the Buy & Hold strategy, with a monthly standard deviation that was almost one third and, more importantly, a maximum drawdown that was less than one sixth, meaning that the tail behaviour was even better than the average behaviour. The resulting Sharpe Ratio, when annualised, is greater than one, a level taken as very good for the practitioners. The DMA strategy, that takes both trend-following and contrarian positions, performed even better, with an 18bps gain in expected return over the MA and a reduced maximum drawdown of just 3.7%, a fraction of the Buy & Hold maximum drawdown. The resulting Sharpe Ratio, when annualised, is very high and close to two. This can be seen as evidence that, on a one-day trading interval, the introduction of contrarian positions in the DMA strategy can have a significant improvement over the pure trend-following MA strategy. However, when applied to foreign exchange markets, both the MA and DMA trading strategies had returns close to zero with very small standard errors. This is driven by the fact that not many currencies had displayed significant correlation (as per Table **??**) and the trading model only enters into positions when the correlation detected is statistically significant.

As demonstrated in Table 3.7, there was a reduction in performance, gross of transaction costs, when the forecast horizon lengthened from one day to one week. However, as Table 3.11 shows, there was a commensurate reduction in trading activity, meaning that depending on the level of transaction costs, it might be more advantageous to trade using a one week forecast horizon (as opposed to a one day forecast horizon). Whilst the average monthly return of the strategies is no longer statistically different than the Buy & Hold, they have considerable lower risk which translates into a significanly positive Alpha. The DMA trading strategy did not have a significant difference in risk-adjusted performance compared to the MA strategy, with both obtaining similar Sharpe Ratios and Alphas. This can be seen as evidence that, in longer forecast horizons, equity markets have mostly positive dependence and trend-following trading strategies alone will capture this dependency, with the inclusion of contrarian trading not adding economic value.

In the case of weekly currency trading, Table 3.8 shows similar performance when compared to the daily strategy (gross of transaction costs), though the DMA strategy had a slight improvement. Like in the case of the daily frequency, the Alphas generated by currency trading are not significanly different than zero, though the maximum drawdown of the DMA strategy was much lower than the one of the Buy & Hold, indicating potential use of the model for risk management and forecasting purposes.

As shown in Table 3.9, the equity strategies did not provide monthly returns statistically different than the one of the Buy & Hold though they did have much reduced risk, again translating into a significanly positive Alpha. This means all equity trading strategies applied to all forecast horizons proposed produced significantly positive Alphas. As it can be seen in Table 3.11, there was a reduction in portfolio turnover when lengthening the forecast horizon from one week to one month (from about 75% of the portfolio per month to about 25% of the portfolio per month). This reduction is most likely not enough to justify the reduction in performance for investors with access to relatively sophisticated execution methods and low execution costs, but is likely to justify the implementation of a smart-beta / semi-passive strategy, as it will yield a similar return to Buy & Hold but with a considerable reduction in market risk and limited portfolio turnover.

We remark that the currency strategies with monthly trading (shown in Table 3.10) also did not produce an Alpha statistically different than zero, though the significant reduction in maximum drawdown was another feature present in this case. Therefore, we conclude that the use of the strategies proposed in currency markets is not likely to add economic value on its own, but there is potential risk management and forecasting value for the underlying correlation model.

All the results so far are gross of transaction costs. The previous steps have avoided embedding the costs as part of the model given that the actual transaction costs an investor has will depend massively on a large numbers of factors that are particular to each individual investor: whether market or limit orders are used, the order execution algorithm, the latency between the server issuing the order and the exchange receiving the order, the size of the order, the choice of broker, the level of commissions and rebates are all factors affecting the final cost.

In [86] is presented a thorough descriptive analysis of trades executed in the US equity markets by 664 institutions that traded 100 times or more per month during the time period from January 1, 1999 to December 31, 2005. The authors found that the difference between the costs of the lowest quintile and the highest quintile was as high as 69 basis points. They also found that institutions in the lowest quintile were able to get negative execution shortfalls, meaning their technology, order execution policy and broker arrangements was able to add value to the gross returns, instead of removing it. Similar findings are made by [87] where the authors claim the lowest quartile of execution costs by institutional investors in UK equities was only 0.4 basis points whilst the top quartile had costs of 27.8 basis points. Therefore, instead of fixing a single cost number, we have chosen a set of possible costs per transaction and combined it with the out-of-sample metrics provided in Tables 3.5 to 3.10 to determine what would be return of a given strategy if it was subject to a given level of transaction costs. To make the numbers easier to compare between themselves and with Buy & Hold, we have rescaled the net returns for all strategies by the ratio of the strategy's monthly standard deviation and the Buy & Hold's monthly standard deviation - i.e. all returns, net of transaction costs, are risk adjusted so that all of them have the same market risk as the Buy & Hold strategy.

Strategy	B&H	MA	DMA
Monthly Avg Return	0.31% (0.60%)	1.07% (0.24%)	1.25% (0.22%)
Monthly Std Deviation	6.27% (0.68%)	2.48% (0.56%)	2.26% (0.59%)
Sharpe Ratio	0.0502 (0.0999)	0.4312 (0.0833)	0.5516 (0.0867)
Max Drawdown	56.2% (19.2%)	6.3% (2.9%)	3.7% (2.4%)
Alpha	N/A	0.0111 (0.0023)	0.0128 (0.0020)
p-value (Alpha $>$ 0)	N/A	<0.0001	<0.0001

[1] Column *B&H* states the financial gain of simply buying a stock and not selling it, hence holding it as a passive investment

[2] Column *MA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the MA classifier predicts a positive (negative) return one step ahead

[3] Column *DMA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the DMA classifier predicts a positive (negative) return one step ahead

Table 3.5: Comparative performance of the strategies applied to equity indices on daily data, gross of transaction costs. Standard Errors in parenthesis.

Strategy	B&H	MA	DMA
Monthly Avg Return	-0.10% (0.28%)	-0.03% (0.04%)	0.00% (0.03%)
Monthly Std Dev	2.93% (0.25%)	0.43% (0.04%)	0.33% (0.04%)
Sharpe Ratio	-0.0355 (0.0964)	-0.0776 (0.0984)	-0.0208 (0.0990)
Max Drawdown	28.5% (12.1%)	6.7% (3.5%)	3.7% (2.5%)
Alpha	N/A	-0.0004 (0.0004)	-0.0001 (0.0003)
p-value (Alpha > 0)	N/A	0.8296	0.6044

[1] Column *B&H* states the financial gain of simply buying a stock and not selling it, hence holding it as a passive investment

[2] Column *MA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the MA classifier predicts a positive (negative) return one step ahead

[3] Column *DMA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the DMA classifier predicts a positive (negative) return one step ahead

Table 3.6: Comparative performance of the strategies applied to currencies on daily data, gross of transaction costs. Standard Errors in parenthesis.

Strategy	B&H	MA	DMA
Monthly Avg Return	0.34% (0.61%)	0.35% (0.22%)	0.28% (0.17%)
Monthly Std Dev	6.33% (0.66%)	2.28% (0.41%)	1.77% (0.36%)
Sharpe Ratio	0.0533 (0.0993)	0.1539 (0.0894)	0.1595 (0.0757)
Max Drawdown	56.2% (19.2%)	8.9% (5.7%)	6.6% (2.9%)
Alpha	N/A	0.0040 (0.0020)	0.0031 (0.0016)
p-value (Alpha > 0)	N/A	0.0240	0.0245

[1] Column *B&H* states the financial gain of simply buying a stock and not selling it, hence holding it as a passive investment

[2] Column *MA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the MA classifier predicts a positive (negative) return one step ahead

[3] Column *DMA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the DMA classifier predicts a positive (negative) return one step ahead

Table 3.7: Comparative performance of the strategies applied to equity indices on weekly data, gross of transaction costs. Standard Errors in parenthesis.

Strategy	B&H	MA	DMA
Monthly Avg Return	-0.09% (0.27%)	-0.03% (0.03%)	0.01% (0.03%)
Monthly Std Dev	2.82% (0.25%)	0.52% (0.07%)	0.32% (0.05%)
Sharpe Ratio	-0.0318 (0.0969)	-0.0573 (0.1034)	0.0047 (0.0993)
Max Drawdown	28.5% (11.4%)	1.9% (1.1%)	1.6% (0.7%)
Alpha	N/A	-0.0003 (0.0005)	0.0000 (0.0003)
p-value (Alpha $>$ 0)	N/A	0.7682	0.5112

[1] Column *B&H* states the financial gain of simply buying a stock and not selling it, hence holding it as a passive investment

[2] Column *MA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the MA classifier predicts a positive (negative) return one step ahead

[3] Column *DMA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the DMA classifier predicts a positive (negative) return one step ahead

Table 3.8: Comparative performance of the strategies applied to currencies on weekly data, gross of transaction costs. Standard Errors in parenthesis.

Strategy	B&H	MA	DMA
Monthly Avg Return	0.30% (0.60%)	0.34% (0.27%)	0.34% (0.23%)
Monthly Std Dev	6.27% (0.68%)	2.36% (0.38%)	2.05% (0.23%)
Sharpe Ratio	0.0473 (0.1002)	0.1174 (0.0910)	0.1165 (0.0905)
Max Drawdown	56.2% (19.2%)	11.4% (7.9%)	8.2% (5.5%)
Alpha	N/A	0.0041 (0.0024)	0.0039 (0.0021)
p-value (Alpha $>$ 0)	N/A	0.0458	0.0305

[1] Column *B&H* states the financial gain of simply buying a stock and not selling it, hence holding it as a passive investment

[2] Column *MA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the MA classifier predicts a positive (negative) return one step ahead

[3] Column *DMA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the DMA classifier predicts a positive (negative) return one step ahead

Table 3.9: Comparative performance of the strategies applied to equity indices on monthly data, gross of transaction costs. Standard Errors in parenthesis.

Strategy	B&H	MA	DMA
Monthly Avg Return	-0.10% (0.28%)	-0.01% (0.04%)	0.00% (0.03%)
Monthly Std Dev	2.94% (0.25%)	0.53% (0.09%)	0.35% (0.07%)
Sharpe Ratio	-0.0334 (0.0970)	-0.0113 (0.1035)	-0.0065 (0.1064)
Max Drawdown	28.5% (12.1%)	5.1% (4.1%)	3.5% (2.1%)
Alpha	N/A	-0.0001 (0.0005)	0.0000 (0.0003)
p-value (Alpha > 0)	N/A	0.5856	0.5621

[1] Column *B&H* states the financial gain of simply buying a stock and not selling it, hence holding it as a passive investment

[2] Column *MA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the MA classifier predicts a positive (negative) return one step ahead

[3] Column *DMA* states the financial gain of implementing a trading strategy that buys (sells) the stock when the DMA classifier predicts a positive (negative) return one step ahead

Table 3.10: Comparative performance of the strategies applied to currencies on monthly data, gross of transaction costs. Standard Errors in parenthesis.

It is noted that the strategies above could be also run without the help of the machine learning algorithm that is used to calibrate the lookback window. As an experiment to gauge the benefit of the use of the machine learning algorithm, a simple DMA classifier was run on the test sample with no fitting of the look-back window and using a blank choice of $n = 1, \hat{\zeta} = -1$, guided by the findings of [38] that equity markets tend to be mean-reverting on daily forecasting horizons. This simple classifier had a surprisingly good monthly return of 1.07%, which compares to 1.25% of the machine learning algorithm given on Table 3.5 (and the difference is not statistically significant given the standard errors in the table). However, the simple classifier also had a monthly standard deviation of 5.82%, as opposed to the 2.26% of the machine learning classifier, meaning that the simple uncalibrated strategy indeed had a similar return profile, but was subject to a much higher variability of returns, being suboptimal from the point of view of risk-adjusted performance. Therefore, one can conclude that the main benefit of the machine learning algorithm over the uncalibrated naive strategy was better dynamic risk management of trading positions.

The level of transaction activity for each strategy was also calculated as this is an additional input needed to calculate the net returns. Table 3.11 shows how many times the portfolio was entirely rebalanced over a month. It can be seen that the Daily Equities MA strategy had the highest turnover, trading on average 11.319 times the entire portfolio in a typical month. On the other hand, the Monthly FX DMA strategy had the lowest turnover, trading only 5% of the portfolio per month. FX strategies naturally had lower turnovers as the test identified less evidence of serial correlation in currencies than in equities, so it traded them less.

Table 3.12 shows the returns of the strategies applied to equities. The daily strategies have the best returns for investors that can achieve low trading costs. In particular, if one can achieve trading costs of 5bps per trade (a level that can be achieved for the most sophisticated investors), the EQ DMA strategy would have obtained an annualised return of more than 20% for the same level of risk that the Buy & Hold would have obtained an annualised return of 3.87%. However, even with trading costs as high as 20bps per trade (a level that is attainable even for savvy retail investors), trading the DMA strategy on a monthly basis can achieve an annualised return in the region of 10% for the same level of risk of the Buy & Hold.

Table 3.13 shows the returns of the strategies applied to currency trading. Given that no strategy could have a positive Alpha, it is also seen that no strategy could have a positive annualised return net of costs, though the Buy & Hold return was also negative for the period. Therefore, these strategies when applied to currencies cannot provide absolute returns on their own. Nevertheless, it is also worth noting that the lower maximum drawdown indicates some potential use of the underlying model for Global Fixed Income portfolios, where the foreign carry interest is also a component of the return, and the strategies would be acting as currency risk mitigants.

It is noted that Tables 3.12 and 3.13 do not report significance ratios or standard errors because the transaction costs were assumed constant for each level of assumed costs. The variability of the results can be ascertained by comparing the values of the same row across different columns; for example Table 3.12 shows that there is a variability between 25% and -25% in the total annualised returns of the DMA Daily trading strategy just due to transaction costs alone. This is very strong evidence that good management of transaction costs is as important as the trading strategy itself.

Equities						Currencies					
Daily		Weekly		Monthly		Daily		Weekly		Monthly	
MA	DMA	MA	DMA	MA	DMA	MA	DMA	MA	DMA	MA	DMA
9.81	11.32	0.94	0.61	0.30	0.18	2.06	0.92	0.55	0.31	0.11	0.05

Table 3.11: Average monthly turnover per strategy (times the portfolio size)

Cost per transaction	3 bps	5 bps	10 bps	15 bps	20 bps
Daily EQ MA	24.58%	18.16%	2.71%	-11.93%	-25.80%
Daily EQ DMA	31.79%	23.55%	3.83%	-14.70%	-32.11%
Weekly EQ MA	10.96%	10.31%	8.71%	7.12%	5.53%
Weekly EQ DMA	11.50%	10.96%	9.62%	8.28%	6.94%
Monthly EQ MA	9.04%	8.88%	8.46%	8.05%	7.63%
Monthly EQ DMA	10.87%	10.75%	10.46%	10.16%	9.87%
Annualised Return of Buy & Hold gross of any transaction costs					3.87%

Table 3.12: Annualised return, net of costs, per equity trading strategy per level of costs

Cost per transaction	2 bps	4 bps	6 bps	10 bps	15 bps
Daily FX MA	-6.10%	-9.47%	-12.82%	-19.47%	-27.71%
Daily FX DMA	-2.67%	-4.60%	-6.53%	-10.37%	-15.15%
Weekly FX MA	-2.66%	-3.37%	-4.08%	-5.50%	-7.27%
Weekly FX DMA	-0.48%	-1.13%	-1.77%	-3.05%	-4.65%
Monthly FX MA	-0.55%	-0.69%	-0.84%	-1.13%	-1.50%
Monthly FX DMA	-0.33%	-0.43%	-0.53%	-0.73%	-0.98%
Annualised Return of Buy & Hold gross of any transaction costs					-1.16%

Table 3.13: Annualised return, net of costs, per currency trading strategy per level of costs

The results net of transaction costs show that even though the economic and statistical value of the signed path dependence models for currency trading is low, this value can actually be high for equity trading. The results of Table 3.12 can be compared to the results reported at Table 3 of [88], where the authors propose an equity trading strategy based on a filter model that aims to exploit serial correlation of stocks, reporting the returns of their strategy net of transaction costs and split per level of correlation. In their experiment, stocks with mean-reverting prices (i.e. negative serial dependence) obtained very negative returns, but for stocks with high levels of serial dependence they could obtain a total annualised return between 20.5% and 21%, net of transaction costs implied by the "typical bid-ask spread" observed in the fifty year window between 1964-2014. Under transaction costs of 5bps or lower, Table 3.12 shows that our proposed strategy can obtain returns marginally better than the ones reported by [88].

3.6 Machine Learning Estimation Method

This section describes a machine learning method that reads a data stream of log-returns in chronological order $t = 1, 2, \ldots$ and for each $r(t)$ read it returns either $\widehat{s}(t+1)$ or zero if it cannot make a reliable prediction for this sign, which could happen either because there is not enough observed information up until time t to make such prediction or because the test defined in Theorem 4 could not find evidence of signed path dependence in the series provided up until that point.

The section alternates between pseudocode, in a smaller font to make it distinguishable, and descriptive text. Reading this section requires some knowledge of object-oriented programming concepts as the pseudocode listed comprises of an object-oriented implementation aimed to be ported into languagues such as C++ and Python. All required variables are listed and commented in the pseudocode.

A single class object is described in this section. Such an object, once fully initialised, should be run by sequentially invoking method *Classify* every time a new data point is observed, obtaining either $\widehat{s}(t+1)$ or zero. For a practical application, the output of the method *Classify* is translated into a trading strategy with three decisions: buy, sell or take no position.

All class variables only require private visibility and are listed as follows:

$className$ the name of the classifier being implemented (either MA or DMA)

R a list that will contain a sample of observed log-returns for a given asset at a given trading frequency

P a list that will contain positive integer numbers candidate values for the parameter n of the classifier being implemented

z an integer number corresponding to the desired learning sample size, assumed even for simplicity

α the significance threshold for the application of the signed path dependence test

w an integer number corresponding to a retuning interval

v an integer number corresponding to a retraining interval

n' an integer number corresponding to a test lookback horizon

n_{max} the optimal value of the classifier lookback for prediction

$bTrade$ indicating whether the asset is worth trading or not

$\overline{r}$ the long term unconditional mean of the return process

ζ the sign of the path dependence being assumed by the predictive method

$countSinceTune$ storing how many data points have been read since the last time the classifier was tuned

$countSinceTrain$ storing how many data points have been read since the last time the classifier was trained

The class object contains four different methods: Initialise, Classify, Predict and Update.

Method **Initialise** is implemented as below:

procedure INITIALISE($P = (n_1, n_2, \ldots, n_N)$, $z > 0$, $1 > \alpha > 0$, $w > 0$, $v > 0$, $n' > 0$, $className$)

  $this.R \leftarrow \emptyset$ ▷ R is initialised as the empty set and will be built as new data is observed

$this.P \leftarrow P$; $this.z \leftarrow z$; $this.\alpha \leftarrow \alpha$; $this.w \leftarrow w$; $this.v \leftarrow v$; $this.className \leftarrow className$

$this.countSinceTune \leftarrow 0$; $this.countSinceTrain \leftarrow 0$; $this.n' \leftarrow n'$

end procedure

Description of method **Initialise**: This method is the class constructor and will be called whenever an instance of the class is created. It stores into internal memory all parameters that are required by subsequent methods and initialises other relevant internal variables. Creating an instance of this class will always require the following parameters:

- $className$: one of MA or DMA, representing the classifier being implemented by this instance of the class - notice that it is a decision *of the user* whether only sign continuation is to be assumed or if a dynamic decision between continuation and reversals is needed.
- P : a list of integers containing acceptable values for the parameter n of the classifier being implemented (the actual choice of n will be performed by method *Update*).
- z : an integer corresponding to the desired learning sample size, to be used by methods *Classify* and *Update*.
- α : the significance threshold to be used by method *Update*.
- w : an integer corresponding to a retuning interval, to be used by method *Update*.
- v : an integer corresponding to a retraining interval, to be used by method *Update*.
- n' : an integer corresponding to a test lookback horizon, to be used by method *Update*.

Method **Classify** is implemented as below:

function CLASSIFY($r(t)$)

Append $r(t)$ to the end of list $this.R$

if $length(this.R) > this.z$ **then**
  Drop the first element of list $this.R$
end if
if $length(this.R) = this.z$ **then**
  Call $this.$Update()
  if $bTrade = TRUE$ **then**
    return $this.$Predict($this.R, this.\overline{r}, this.n_{max}$)
  end if
end if
return 0
end function

Description of method **Classify**: This is the only public method of the class. It should be called every time t a new data point has become available and returns either $\widehat{s}(t+1)$ or zero, which also can be taken as a trading decision: -1 meaning *negative / sell*; 1 meaning *positive / buy*; 0 meaning *no predicted sign / take no position*. The only parameter required to run this method is:

- $r(t)$: the observed data point (log-return).

The method stores $r(t)$ in an internal class variable, in sequential order so that it can be used in subsequent iterations and updates of the model. However, if after storing $r(t)$ the internal storage has $z+1$ points (with z given during initialisation), the method will drop the oldest (first) data point of the internal storage to ensure that the learning sample size stays constant at z.

If there are fewer than z data points in the internal storage, the method simply returns zero, not attempting any prediction until more data is received to fill up the internal storage. Otherwise, it calls the internal method *Update* which will learn the optimal values for parameters n and $\overline{r}$ given the data in the internal storage and save to internal memory the results of the test defined in Theorem 4.

Finally, if the test results saved in memory inform that there is no significant path dependence in the series saved to the class internal storage, the method will again simply return zero. Otherwise, it will return the result of calling the internal method *Predict* giving the whole data in the internal storage as the parameter R and the parameters n and $\overline{r}$ as obtained in the method *Update*.

Method **Update** is implemented as below:

procedure UPDATE( )

$this.countSinceTune \leftarrow this.countSinceTune + 1$

$this.countSinceTrain \leftarrow this.countSinceTrain + 1$

if $this.countSinceTune = w$ **then** ▷ The tuning stage is performed inside this block

$this.\overline{r} \leftarrow \frac{1}{z}\sum_{i=1}^{z} this.R(i)$

if $this.className =$ "MA" **and** $\frac{\overline{d}}{\sqrt{\frac{\sum_{\tau=-this.n'}^{this.n'}\widehat{\gamma}(\tau)}{s-this.n'}}} > \Phi^{-1}(1-this.\alpha)$ **then**

$this.bTrade \leftarrow TRUE$

$this.\zeta \leftarrow 1$

else if $this.className =$ "DMA" **and** $\left|\frac{\overline{d}}{\sqrt{\frac{\sum_{\tau=-this.n'}^{this.n'}\widehat{\gamma}(\tau)}{s-this.n'}}}\right| > \Phi^{-1}\left(1-\frac{this.\alpha}{2}\right)$ **then**

$this.bTrade \leftarrow TRUE$

$this.\zeta \leftarrow 1$ **if** $\frac{\overline{d}}{\sqrt{\frac{\sum_{\tau=-this.n'}^{this.n'}\widehat{\gamma}(\tau)}{s-this.n'}}} > \Phi^{-1}\left(1-\frac{this.\alpha}{2}\right)$ **or** -1 **if** $\frac{\overline{d}}{\sqrt{\frac{\sum_{\tau=-this.n'}^{this.n'}\widehat{\gamma}(\tau)}{s-this.n'}}} < \Phi^{-1}\left(\frac{this.\alpha}{2}\right)$

else

$this.bTrade \leftarrow FALSE$ ▷ The asset is not worth trading

end if

$this.countSinceTune \leftarrow 0$

end if

if $this.countSinceTrain = v$ **then** ▷ The training stage is performed inside this block

$\overline{r_1} \leftarrow \frac{2}{z}\sum_{i=1}^{z/2} this.R(i)$

$this.n_{max} \leftarrow \underset{n\in P}{\arg\max}\frac{2}{z}\sum_{i=\frac{z}{2}+1}^{z}\exp(this.R(i) \times this.\text{Predict}((this.R(i-z/2),\ldots,this.R(i-1)),\overline{r_1},n))$

$this.countSinceTrain \leftarrow 0$

end if

end procedure

Description of method **Update**: This is an internal method that performs the learning process of the class, split into two stages:

- *Tune*: This stage decides if any reliable prediction can be made and sets other relevant internal variables of the class. To enhance runtime performance, the *Update* method runs this stage only once every w iterations (with w given during initialisation) and relies on results from previous runs (stored in internal memory) when this stage is not run. This stage uses all data points in the internal storage of the class to decide if the asset is worth trading by checking for significant path dependence as follows:

  – If the classifier being implemented by this instance of the class is the MA classifier, the test defined in Theorem 4 is applied to all stored data points to look for significant positive path dependence only, and the asset is deemed worth trading if such dependence is found.
  – Alternatively, if the classifier being implemented by this instance of the class is the DMA classifier, the test defined in Theorem 4 is applied to all stored data points to look for significant path dependence of any sign which, if found, would make the asset worth trading.

If the asset is deemed worth trading, the method estimates its long term mean return (and stores it internally for future usage) as the observed mean return over all data points in the internal storage of the class.

- *Train*: This stage decides the actual value of n that should be used for classification. To enhance runtime performance, the *Update* method runs this stage only once every v iterations (with v given during initialisation) and relies on results from previous runs (stored in internal memory) when this stage is not run. This stage splits the data points in the internal storage into two subsamples of equal length: *training* (the first chronological half of the stored list) and *validation* (the second chronological half of the stored list, henceforth referred to as $R_{validation}$). Using the mean return over the *training* sample as an input to the classifier being trained, it finds which value of n produces the highest classification gain over the *validation* sample, with the gain defined as:

$$G(n) = \exp\left(\sum_{r(t) \in R_{validation}} r(t) \times \widehat{s}(t) \right) \tag{3.21}$$

where $\widehat{s}(t)$ is obtained by method *Predict* using $\overline{r}$ equal to the mean return over the *training* sample, $R = (r(t-1), r(t-2), \ldots, r(t-z/2))$ and n as the lookback parameter.

All learned parameters are stored in the private class variables for subsequent usage.

Remark 11. *The gain function chosen has the advantage of being linear on the returns, which means it can also be used to evaluate the returns of a portfolio comprised of several assets invested according to the underlying classification strategy.*

And, finally, method **Predict** is implemented as below:

function PREDICT(R, $\overline{r}$, n)
  return $this.\zeta \times \operatorname{sgn}\left(\sum_{i=0}^{n-1} (R(length(R)-i)-\overline{r})\right)$
end function

Description of method **Predict**: This is an internal method that performs the actual prediction as a direct one step ahead implementation of Definition 4 if $className$ = "MA" or as per Definition 5 if $className$ = "DMA", returning either -1 or 1. It receives three inputs:

- R : a list of log-returns in chronological order, assumed to be the most recent returns up to the point that the method *Predict* was called.
- $\overline{r}$: the long term mean log-return of the asset.
- n : the value of the lookback parameter.

In case this instance of the class is a DMA classifier, the value of $\hat{\zeta}$ will have already been obtained and stored internally by method *Update* by the time this method is called.

In the empirical experiment of Section 3.5, the tuning stage of the algorithm was done every six months, which meant a retune was done after six new monthly data points, 24 new weekly data points and 126 new daily data points. The training stage was done monthly, which meant a retrain was done for every new point for monthly data, after 4 new points for weekly data and after 21 new points for daily data.

Chapter 4

Using Conditional Asymmetry to Predict Commodity Futures Prices

This chapter presents an empirical study of a number of commodity prices in the futures markets fitted to a novel time series model where the source of nonlinear serial dependence in returns arises out of conditional asymmetry: the shape of the probability distribution of the returns varies in a way that can be expressed as a function of previous returns. This property of the fitted time series model is used to provide predictions with respect to future returns by identifying moments in time where the future expected return is positive or negative. Moreover, the model here proposed can also be easily linked to the profitability of trend-following / time-series momentum trading strategies by interpreting the values of the coefficients obtained.

This chapter provides the following contributions to the existing literature: (i) an up-to-date empirical study demonstrating evidence of continued profitability of a trend-following strategy applied to commodity futures markets between 2009 and 2019; (ii) a parametric time series model that can model of nonlinear serial dependence in the shape of the distribution of price returns and can provide useful information about the expected future behaviour of a seemingly uncorrelated stochastic process; (iii) an additional empirical application of this model to generate of probabilistic forecasts of variables subject to nonlinear serial dependence triggered by dependent changes in the shape of their distribution.

The chapter is structured as follows: Section 4.1 provides the formulation of the model and explains how to estimate the required parameters of the model, which reparametrizations can be performed to enhance numerical stability of the computations involved and which properties can be used to extract nonlinear serial dependence of a given time series. Section 4.2 demonstrates on a number of commodity future price series the empirical performance of the model to detect nonlinear serial dependence and potentially guide investment decisions, while Section 4.3 benchmarks the probabilistic predictions of markets generated by the model proposed in this manuscript against the ones made by an ARMA(p,q)-GARCH(1,1) model.

4.1 Model Formulation and Fitting

The mathematical foundation behind the study here proposed is based on the standard formulation of time-varying regime switching models. The literature review in [89] gives a good background on such models. In our formulation, the asset return model is based on a constrained form of Gaussian mixture whose weights are time-varying and specified as a logistic regression of the asset return at lag $t-1$ and the sign of the cumulative return of the asset between times $t-n$ and $t-2$.

Let $f_{r_t|n;p_{t-1},p_{t-2},p_{t-n-1}}$ be the probability density function of the log-return r_t conditional on the prices observed at times $t-1$, $t-2$ and $t-n-1$. We propose that

$$f_{r_t|n;p_{t-1},p_{t-2},p_{t-n-1}} = w\varphi\left(\mu_H,\sigma_t^2\right)+(1-w)\varphi\left(\mu_L,\sigma_t^2\right), \tag{4.1}$$
$$w = \frac{1}{1+\exp(-z_t)},$$
$$z_t = \alpha+\beta r_{t-1}+\gamma\,\mathrm{sgn}(p_{t-2}-p_{t-n-1})$$

where sgn is the sign function, $\varphi(x,y)$ is the density function of a Normal distribution of mean x and variance y, w is the probability at time t that the asset is in a high-return regime, $\mu_H \geq 0$ is the expected return of the asset in the high-return regime, $\mu_L \leq 0$ is the expected return of the asset in a low-return regime and σ_t^2 is the overall volatility of asset returns, which is allowed to vary through time (and hence can follow any ARCH process [90] or similar variant). The other additional parameters are α, β and γ, which will determine how the probability of staying in the high-return regime behaves through time.

Notice that, as the sign function is bounded between 1 and -1, irrespective of how arbitrarily large is the choice of parameter n, the contribution of z_t to the variance of the returns will be bounded. At the same time, all returns from $t-n$ to $t-2$ will have equal weight in their contribution to z_t, hence allowing for a very slow decay in the serial dependence of the process, as documented by several empirical experiments. Finally, as will be shown later, the process z_t is a useful by-product of this model formulation, as one can establish whether the asset return is expected to be positive or negative at time t by observing this quantity, which can be known with certainty at time $t-1$.

The formulation above resembles the time-varying regime switching model proposed in [91], which is also known as a Logistic Mixture model. However, in [91] the regime weights are not self-excited, as it is the case in Equation 4.1. Another similar auto-regressive model in the literature is the Gaussian Mixture Auto-Regressive model (GMAR) where the means of each component in the Gaussian Mixture follow auto-regressive processes. A general formulation of a

GMAR model can be found in [92]. Smooth Transition Autoregressive (STAR) models [93] also aim to capture time series dynamics where there is a transition between two or more auto-regressive regimes; in STAR models though the conditional distribution of the dependent variable is typically assumed to be Gaussian instead of a Gaussian Mixture.

The choice of a two state regime in Equation 4.1 is driven by a large volume of empirical evidence that commodity returns can be accurately described using a two state regime switching model. In [94] the authors analysed a sample of 12 different commodity spot prices that included metals and energy and observed with high statistical significance the presence of two distinct regimes in all the commodities analysed, with a clear difference in volatility across both regimes. A similar finding was made in an empirical analysis of crude oil futures made in [95]. Further, in [96] the authors analysed futures contracts for 4 different metals plus crude oil and observed that a two state regime was the best fit for copper, zinc and crude oil futures and a three state regime was the best fit for gold and silver futures, though the difference in the location parameter was not statistically significant between two of the three regimes for gold and silver.

It is noted that the seminal work about informed trading described at [97] has described market order flow to be explained by a three-state regime, where there can be "good news", "bad news" and "no news". The model in Equation 4.1 does not contradict the market description of [97] and it can be easily adapted to include an extra variable corresponding to the probability of a "no news" event, though at a cost of increasing the chance of overfitting and consequently increasing the risk of lower out-of-sample predictive ability. As noted in Chapter 3, financial time series tend to have a high level of noise in proportion to the signal being detected and the main motivation of the concept of signed path dependence was to reduce the sensitivity of the model fit to such noise. Introducing an extra variable and an extra state might go against this motivation, especially in the case of a model fit to monthly returns, which will naturally be a reduced set of observations. The goal of this model is not to predict the actual order flows that occur on a daily basis but instead to predict the impact of a persistent market event on returns over longer term horizons, typically monthly. A model based on order flow data such as the one described in [98] is better suited for an application that involves understanding the actual order flow.

Remark 12. *The variable z_t can be interpreted as a latent factor that represents the persistence of an imbalance in supply and demand of a specific commodity. While it is clear that commodity spot prices change over time as a direct result of changes in supply and demand, it is not immediately clear that commodity future prices change as a direct result of changes in supply and demand. Derivatives trade in "zero net supply" markets, where each buyer has always a matching seller. Due to this zero net supply condition, changes in derivative prices might not necessarily be driven by changes in the fundamentals driving the spot prices, as some of the buyers or sellers can simply be financial agents not exposed to the supply and demand effects, trading purely as market-makers or speculators. Despite the presence of financial agents guaranteeing a zero net supply condition for commodity futures prices, a number empirical studies have demonstrated that these agents do not substantially alter the dynamics of commodity futures prices and the same supply and demand imbalances that affect the spot prices also affect the futures prices. In [99], the authors applied a Dynamic Equicorrelation to the returns of commodity futures of four different categories (energy, precious metals, industrial metals and agriculture) and verified that the level of time-varying co-movement across these contracts was very low for agricultural commodities and low for energy and industrial metals, suggesting that fundamentals specific to physical supply and demand each commodity were the main drivers of the prices of these futures contracts instead of financial conditions. A similar conclusion was reached for crude oil futures by [100], who* *analysed how structural breaks in the co-integration between WTI oil futures and Brent oil futures could be explained by events that affected oil production, inventories and international trade and verified that dates of major global fundamental events coincided with the points where these structural breaks occurred. Finally, it is also worth citing the work in [101] where the authors analysed changes over time in net futures positions by commercial firms and compared against the changes over time in net futures positions by financial agents and noted that although financial firms tended to trade as counterparties of commercial firms, there was no evidence that the changes in positions of financial firms were causing the movements in futures prices to deviate from movements in inventories, which are typically driven by supply and demand.*

As in most parametric time series models, there are several different ways to estimate the parameters of the model implied by Equation 4.1. In this manuscript the model will be estimated via maximum likelihood. This allows the estimation of volatility parameters to be done at the same stage that the estimation of all other model parameters is done by expanding the likelihood function so that it includes the volatility parameters of the chosen volatility model and estimating all parameters simultaneously in the expanded likelihood function. Further, it keeps the model estimation procedure very similar to what is already established in the field of regime switching models.

Like in ARCH-type and in regime switching models, the likelihood function can be expressed as a product of iterated conditional likelihood functions. Let r_t be the asset return at time t, σ_t^2 its variance at time t which is assumed known with certainty given the previous history of asset returns up until that point (this does not preclude it from being modelled by some conditionally heteroskedastic process, and it indeed is modelled as such in Section 4.2) and let $\theta_n = (\mu_H, \mu_L, \alpha, \beta, \gamma)$ the vector of parameters to be estimated, given the hyperparameter n. Define $\zeta_t = \mathrm{sgn}\left(\sum_{i=1}^{n-1} r_{t-i}\right) = \mathrm{sgn}(p_{t-1} - p_{t-n})$.

Notice that

$$f_{r_t,r_{t-1},\ldots|n;\theta_n} = f_{r_t|n;r_{t-1},\zeta_{t-1},\theta_n} f_{r_{t-1}|n;r_{t-2},\zeta_{t-2},\theta_n} \cdots f_{r_{n+1}|n;r_n,\zeta_n,\theta_n} f_{r_n,\zeta_n|n;\theta_n}$$

with $f_{r_t|n;r_{t-1},\zeta_{t-1},\theta_n}$ given by Equation 4.1. Therefore

$$\begin{aligned} L(n;\theta_n) &= \log\left(f_{r_t,r_{t-1},\ldots|n;\theta_n}\right) \\ &= \log\left(f_{r_n,\zeta_n|n;\theta_n}\right) + \sum_{i=n+1}^{t} \log\left(f_{r_i|n;r_{i-1},\zeta_{i-1},\theta_n}\right). \end{aligned} \tag{4.2}$$

For sufficiently large n, the second term of the sum in Equation 4.2 will dominate the sum, hence the maximum likelihood estimate of θ is given by

$$\widehat{\theta_n} = \underset{\substack{\theta \\ \text{s.t.}\, \mu_H \geq 0 \\ \mu_L \leq 0}}{\arg\max} \sum_{i=n+1}^{t} \log(f\,(r_i|n; r_{i-1}, \zeta_{i-1}, \theta_n))$$

which can be reparameterised to an unconstrained maximum likelihood estimation problem by defining $\mu_H = \tilde{\mu}_H^2$, $\mu_L = -(\tilde{\mu}_L)^2$ and $\tilde{\theta_n} = (\tilde{\mu}_H, \tilde{\mu}_L, \alpha, \beta, \gamma)$ and obtaining

$$\widehat{\tilde{\theta_n}} = \underset{\tilde{\theta_n}}{\arg\max} \sum_{i=n+1}^{t} \log\left(f\left(r_i|n; r_{i-1}, \zeta_{i-1}, \tilde{\theta_n}\right)\right) \tag{4.3}$$

with the value of $\widehat{\theta_n}$ being subsequently obtained from $\widehat{\tilde{\theta_n}}$ of Equation 4.3 by substitution.

In this manuscript, the value of n was also chosen via maximum likelihood by obtaining the maximum likelihood estimates of θ_n for all values of n between 1 and n_{max} where n_{max} is a user input representing the maximum value that n can be used to fit the model without compromising the stability of the estimates due to a low number of observations. The chosen value of n would be the value between 1 and n_{max} that yielded the highest likelihood value for the maximum likelihood estimate of θ_n. To ensure all likelihood values are comparable, all estimates were made using the same observation set, which meant that for each possible value of n, the first $n_{max}-n$ observations available to fit the model were discarded.

As mentioned before, the process $z_t = \alpha+\beta r_{t-1}+\gamma\zeta_{t-1}$ can be used to infer whether at time t the asset price $p(t)$ is expected to be positive or negative. To understand how, notice that it follows from Equation 4.1 that

$$\mathbb{E}[r_t|z_t] = \frac{1}{1+\exp(-z_t)}\mu_H + \frac{\exp(-z_t)}{1+\exp(-z_t)}\mu_L$$

which implies that

$$\begin{aligned}
\mathbb{E}[r_t|z_t] \geq 0 &\iff -\exp(z_t)\mu_L \leq \mu_H \Rightarrow \\
\mathbb{E}[r_t|z_t] \geq 0 &\iff z_t \geq -\log\left(-\frac{\mu_H}{\mu_L}\right).
\end{aligned}$$

In an empirical setting, one can define the estimator $\widehat{a} = -\log\left(-\frac{\widehat{\mu}_H}{\widehat{\mu}_L}\right)$ and predict the sign of the future return r_t as

$$\widehat{\mathrm{sgn}(r_t)} = \mathrm{sgn}\left(\widehat{\alpha}+\widehat{\beta}r_{t-1}+\widehat{\gamma}\,\mathrm{sgn}(p_{t-2}-p_{t-n-1})-\widehat{a}\right). \tag{4.4}$$

This prediction strategy is explored in Section 4.2. Additionally, one can use Equation 4.1 to make a probabilistic prediction of r_t, which can be useful for the calculation of metrics such as Value-at-Risk, Potential Future Exposure or also to guide any hedging strategy of commodity price risk. This is explored in Section 4.3.

Finally, it is noted that the standard errors for the maximum likelihood estimates of the parameters can be calculated obtaining the estimated Fisher Information by finite differences of the log-likelihood at the maximum point. Thereafter, the estimated standard errors are the square roots of the diagonal elements of the inverse of the estimated matrix. For example, if the maximum likelihood estimate is vector $\widehat{\theta_n} = \left(\widehat{\mu}_H, \widehat{\mu}_L, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}\right)$ then the standard error of, say, $\widehat{\mu}_H$ is estimated by

$$\widehat{\mathrm{se}}(\widehat{\mu}_H) = \sqrt{\left(-\left.\frac{\partial^2 L(\theta_n)}{\partial \mu_H^2}\right|_{\theta_n=\widehat{\theta_n}}\right)^{-1}} = \sqrt{\frac{(\widehat{\mu}_H \delta)^2}{2L\left(\widehat{\theta_n}\right) - L\left(\widehat{\theta_{n\mu_L^+}}\right) - L\left(\widehat{\theta_{n\mu_L^-}}\right)}}$$

$$\widehat{\theta_{n\mu_L^+}} = \left(\widehat{\mu}_H(1+\delta), \widehat{\mu}_L, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}\right)$$

$$\widehat{\theta_{n\mu_L^-}} = \left(\widehat{\mu}_H(1-\delta), \widehat{\mu}_L, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}\right)$$

where δ is an arbitrarily small real number. The standard errors of all other parameters can be calculated following a similar logic. Its also noted that, should the covariances between the estimated parameters be of interest, these can also be obtained from the inverse of the estimated Fisher Information matrix.

4.2 Forecasting Price Direction in Commodity Futures Markets

The model of Equation 4.1 has been fitted to different commodities covering six of the most liquid futures contracts for Energy and Metals. Agricultural commodities were not included in the sample due to the fact that their prices are also affected by seasonality and the model implied by Equation 4.1 cannot capture seasonality effects, with such effects being left outside of the scope of the present manuscript.

Prices were sourced from the Quandl continuous futures database using an observation dataset spanning from September-1990 to August-2019, with the exception of Brent Crude Oil that had price data only from May-1993 onwards and the returns calculated from the prices thereafter. The commodities analysed are listed on Table 4.1. The model was fitted using $n_{max} = 25$. The predictions were made in an out-of-sample set of 125 months ending in August-2019 using constant parameters fitted before the start of the out-of-sample period, with a forecast horizon of one month.

Return series have been obtained by taking the logarithm of the quotient between the month-end futures price of the contract month that is second to expire against the previous month-end futures price of the contract month that is third to expire. As all six commodities used in the elaboration of the present manuscript have consecutive monthly expires spanning at least the first three contract months, our return series corresponds in practice to the return of holding a long position in the same contract throughout a calendar month and rolling this position to the subsequent expiry month at month-end.

To include the effects of conditional heteroskedasticity in the model estimation, an exponentially weighted moving average variance method ([102]) was implemented. In such formulation, the variance at time t is given by

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2 \tag{4.5}$$

where $\sigma_1^2 = r_1^2$. The parameter λ was estimated by maximum likelihood by substituting Equation 4.5 into Equation 4.1 and maximising the likelihood function given by Equation 4.3.

Table 4.2 shows the maximum likelihood estimates for the six commodities studied. The estimated values can be interpreted as follows:

- For all commodities studied, the parameter estimate of β was positive, indicating that positive returns in the previous month $t-1$ increase the probability that a high-return regime will occur in month t, hence increasing the expected return in month t. This is consistent with the existence of profitable trend-following strategies in this asset class.

- For all commodities other than Copper, the parameter estimate of n was in the region of 12 and the parameter γ was positive, suggesting evidence of long term persistence in their returns affecting the returns of almost one year later. This is also consistent with the existence of profitable trend-following strategies that aim to extract value from long term trends.

- For Copper, the parameter n was 23 and the parameter estimate of γ was negative, suggesting evidence of long term anti-persistence in returns after about two years. This is consistent with the existence of Copper trading strategies that aim to extract value from long term mean-reversion.

The sign estimator given by Equation 4.4 was used to drive investment decisions of a hypothetical Managed Futures strategy that would trade on all six commodities, taking long positions in the associated futures contract whenever the predicted sign was positive and short positions whenever the predicted sign was negative. The overall portfolio was equally weighted on all commodities and no variance-weighting technique was applied. It is worth noting that the return series for five out of the six contracts had the same annualised standard deviation of circa 30%, with the only exception being Gold that had a standard deviation of about half of that amount. This indicates that variance-weighting would have had limited influence on the results.

Table 4.3 shows the annualised returns of the hypothetical strategy and annualised standard deviations, calculated by multiplying the standard deviation of monthly returns times the square root of twelve. The numbers are shown gross of transaction costs; however, as the strategy only involves one futures trade per month, these costs are not expected to materially erode the returns, especially if coupled with good order execution algorithms. It is noted that Brent Crude Oil, Gas Oil and Gold offered the best risk-return profile of all six but trading WTI Crude Oil did not generate excess returns.

Table 4.1: Data coverage per commodity analysed

Commodity	No. of Fitting Observations	Quandl 2nd Month Code	Quandl 3rd Month Code
Brent Crude Oil	190 months (1)	ICE_B2	ICE_B3
WTI Crude Oil	222 months (2)	CME_CL2	CME_CL3
Gas Oil	222 months (2)	ICE_G2	ICE_G3
Natural Gas	222 months (2)	CME_NG2	CME_NG3
Gold	222 months (2)	CME_GC2	CME_GC3
Copper	222 months (2)	CME_HG2	CME_HG3

[1] From May-1990 to March-2009
[2] From September-1990 to March-2009

Table 4.2: Maximum Likelihood Estimates of Logistic Mixture Parameters

Commodity	$\widehat{n}$ (1)	$\widehat{\mu}_H$	$\widehat{\mu}_L$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\lambda}$
Brent Crude Oil	11	0.025	-0.085	3.854	43.08	1.083	0.874
WTI Crude Oil	12	0.024	-0.053	1.564	15.73	1.398	0.813
Gas Oil	11	0.036	-0.029	0.786	21.35	1.463	0.801
Natural Gas	10	0.128	-0.031	-1.826	3.075	0.819	0.975
Gold	10	0.045	-0.011	-1.476	0.186	0.867	0.787
Copper	23	0.004	-0.267	10.21	27.36	-4.496	0.925

[1] Lookback horizon of the logistic mixture model, also estimated (i.e. not a fixed parameter)

Table 4.3: Investment Performance of Hypothetical Managed Futures Strategy.

Commodity	Ann Return	Ann Std Dev	Sharpe Ratio
Brent Crude Oil	5.23%	27.30%	0.192
WTI Crude Oil	-2.68%	28.74%	-0.093
Gas Oil	11.20%	25.20%	0.445
Natural Gas	5.56%	31.10%	0.173
Gold	3.42%	16.68%	0.205
Copper	0.46%	21.71%	0.021
Overall Portfolio	6.37%	12.93%	0.493

It is also noted that the overall portfolio comprising of the six commodities had a material reduction in risk against each individual commodity (measured by the portfolio's standard deviation) together with a good positive return, which can be take as evidence that the prediction errors across all of the six commodities were largely uncorrelated.

The returns on Table 4.3 can be compared to results already documented in the literature for commodity futures trading strategies based on serial dependence of returns. There are two leading studies: the cross sectional momentum [103] and the time-series momentum [7].

The time-series momentum analysis is very similar to the concept of signed path dependence introduced in Chapter 3 where it correlates the sign of the return n steps ahead against the sign of the return h steps back. In particular, the MA classifier of Chapter 3 with $n = 12$ and $h = 1$ for monthly observation frequency is equivalent to the study done in [7]. In this manuscript, the authors applied their predictive model to build a portfolio of positions in commodity futures, which would hold for one month a long (i.e. positive weight) exposure in a commodity futures contract in case the return in the previous twelve months was positive or a short (i.e. negative weight) exposure in case this return was negative. The authors did not report the Sharpe Ratio of the overall portfolio, but reported the Sharpe Ratio of individual commodities and their numbers were comparable to the ones in Table 4.3. For example, their Sharpe Ratio of 0.05 for

Natural Gas was slightly lower than our Sharpe Ratio of 0.17 for Natural Gas and their Sharpe Ratio of 0.5 for Gas Oil was similar to our Sharpe Ratio of 0.44. The authors reported clearly better results for the Brent Crude Oil, with a Sharpe Ratio of 0.7 in their study against our Sharpe Ratio of 0.2. However, as the authors did not report an overall portfolio Sharpe Ratio, the comparison cannot be done to the full extent as in a practical situation the returns of the overall portfolio are more important than the returns of individual components.

The cross-sectional momentum study aimed to build a portfolio based on relative returns by ranking the commodities over their returns in the previous 12 months and assigning weights based on the rank: so the commodity with the highest rank would have the highest positive weight in the portfolio and the commodity with the lowest rank would have the lowest weight in the portfolio, which would be a negative weight as the weights were normalised around the mean rank so as to sum zero. Using 27 different commodities and the aforementioned method, the authors in [103] obtained a portfolio Sharpe Ratio of 0.51, which is very close to our portfolio Sharpe Ratio of 0.49.

To confirm that the returns of the overall portfolio are statistically significant, a dummy investment strategy was created that would trade all six commodities over the same 125 out-of-sample months and take long or short positions by uninformed guessing (random number generation), with both cases taking equal probability. This strategy was simulated 5000 times so as to generate an empirical distribution for the average return of this dummy strategy over the 125 out-of-sample months, which was used as a null distribution for a hypothesis test against the null hypothesis that the average return of the overall portfolio (on Table 4.3) is equal to or less than the one obtained by random uninformed guessing. The critical value for this test at 5% Type I error probability was obtained as 4.85% and, as the overall portfolio returns (6.37%) were greater than the critical value, the null hypothesis can be rejected and the returns can be considered statistically significant at such level of confidence.

This section is concluded by analysing the behaviour of the predictions made for specific commodity contracts to ascertain the boundary conditions under which the Managed Futures strategy based on such predictions can be profitable and the conditions that cause such strategy to generate negative excess returns. To achieve that, two commodities are analysed: Gas Oil (the one whose trading yielded the best returns as per Table 4.3) and WTI Crude Oil (the one yielding the worst returns).

Figure 4.1 shows the evolution of a hypothetical index whose baseline value of 100 is set at the end of Dec-2008 and for each month thereafter the index value changes in the same proportion that the price has changed in a calendar month for the Gas Oil futures contract that had become the second month to the expire at the end of that month, in line with the futures expiry logic that was used to choose which contracts were going to be traded in the hypothetical Managed Futures strategy analysed on Table 4.3. In other words, this index corresponds to the evolution of an initial investment of 100 US Dollars in a long-only position in Gas Oil futures, rolling on a monthly basis.

The index values are then superseded with vertical dashed lines corresponding to the dates where the sign of the one month-ahead return predicted by the model has changed: for example, the dashed line with an upwards arrow at the end of July-2010 means that the model started to predict a positive monthly return for Gas Oil futures; thereafter, the model stayed predicting a positive return until the end of May-2012, when it started to predict a negative monthly return, illustrated by the subsequent dashed line with a downwards arrow.

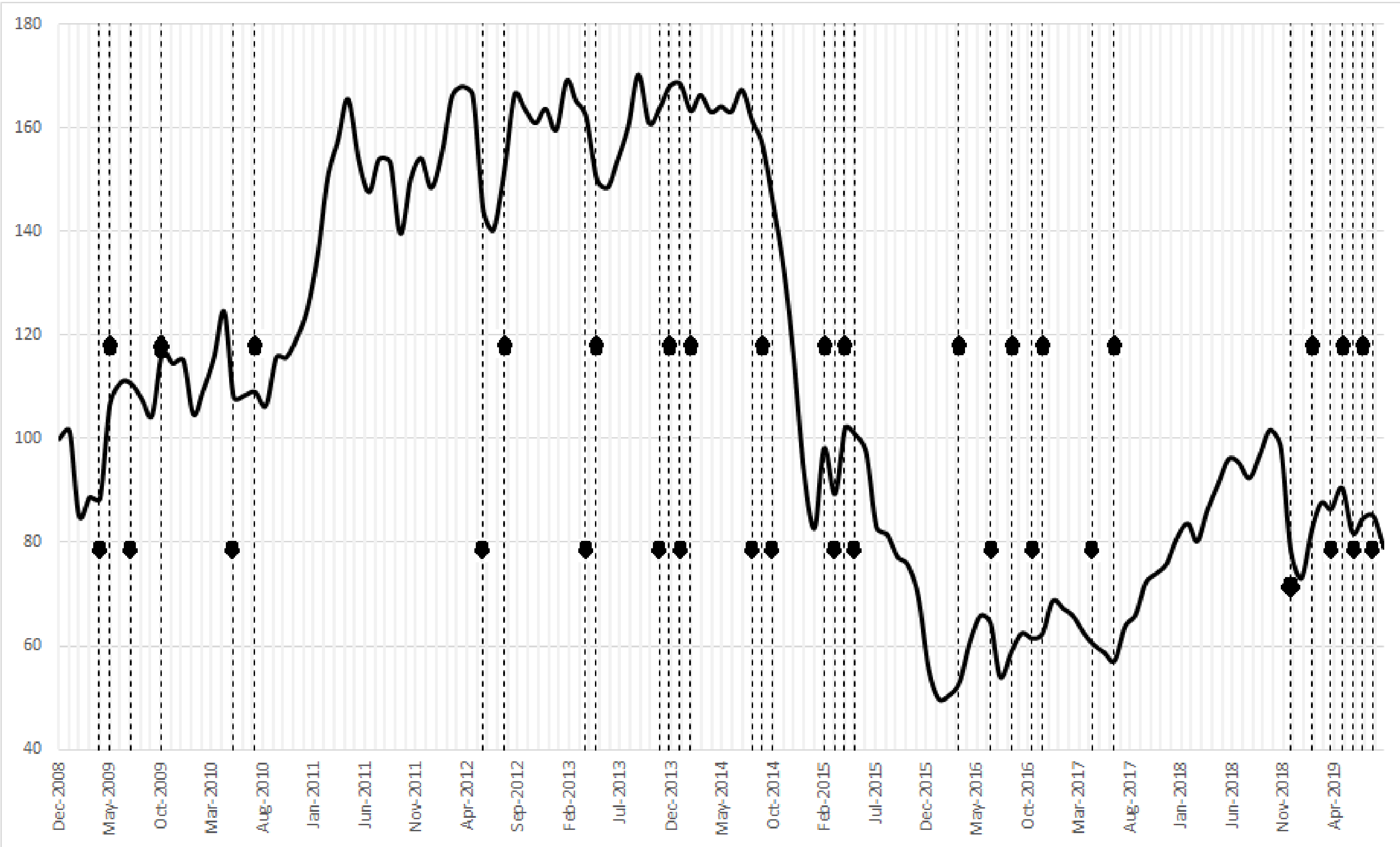

Figure 4.1: Predicted start dates of Gas Oil Futures uptrends (dotted vertical line with upwards arrow) and downtrends (dashed vertical line with downwards arrow). Baseline 100 at 31-Dec-2008.

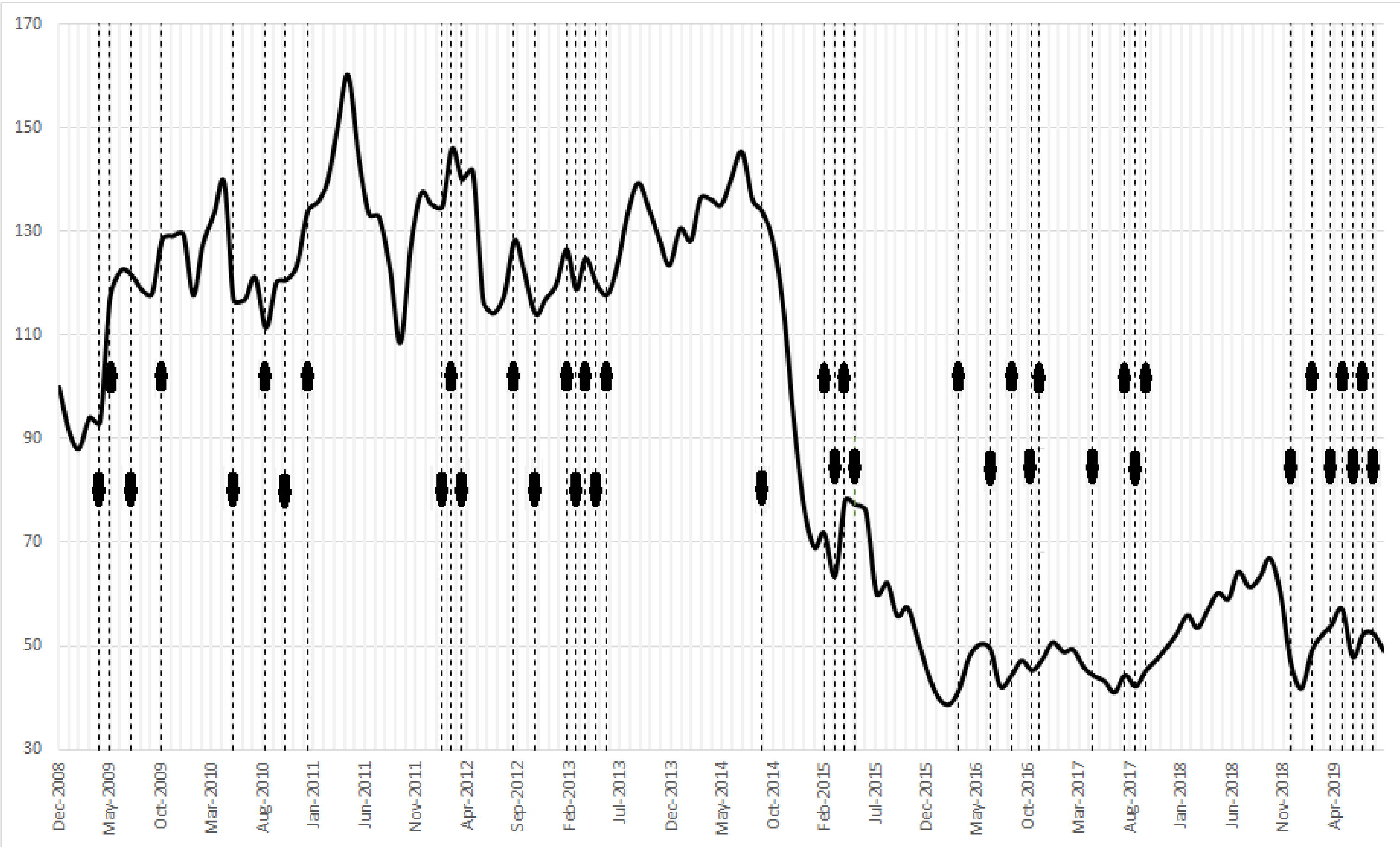

Figure 4.2: Predicted start dates of WTI Crude Oil Futures uptrends (dotted vertical line with upwards arrow) and downtrends (dashed vertical line with downwards arrow). Baseline 100 at 31-Dec-2008.

It can be seen that there were four times when the model correctly anticipated large moves/trends in Gas Oil futures prices (July-2010, Sep-2014, May-2015 and June-2017). Such correct predictions generated substantial value and were able to offset the relatively larger number of instances where the model anticipated a change of sign but such change did not materialise - i.e. false positives. As Table 4.3 demonstrates, Gas Oil trading was able to obtain excess profits despite having more false positives than otherwise due to the magnitude of the correctly anticipated trends.

Figure 4.2 shows the evolution of a hypothetical index constructed in the same way as in Figure 4.1 but using WTI Crude Oil futures instead. It can be seen a similar pattern to the one of Gas Oil futures, with large moves/trends in WTI Crude Oil futures being correctly anticipated in May-2013, Aug-2014, May-2015 and Sep-2017; however, there were too many false positives at times of high volatility, which meant that incorrect predictions generated large losses that eroded the returns in correctly anticipated trends.

It is worth noting that the strategy trading WTI Crude Oil futures was accumulating positive returns until Sep-2018 and these cumulative returns turned negative after extraordinary trendless volatility in the last 11 months out-of-sample, with a drawdown of almost 70% peak to through. Given the high variability of the returns of the proposed trading strategy applied to a single commodity, it is recommended that an overall commodity portfolio should be formed (as done in this study and in many others in the literature) and, ideally, this commodity portfolio should sit within a well balanced investment portfolio.

4.3 Economic Probabilistic Forecasts in Commodity Futures Markets

There is another rather important empirical application for the model implied by Equation 4.1: probabilistic forecasts of economic variables. Probabilistic forecasts have an additional information as opposed to point forecasts, providing the level of uncertainty around a forecast. Though the changes in the level of variability in market variables due to recent returns can be easily accounted for by conditionally heteroskedastic variance models, Equation 4.1 allows an econometrician to obtain a full probability density for the variable being forecasted that can take into account changes in the tail behaviour and return asymmetry due to recent returns in addition to the information also given by variance models.

As a comparative experiment, Equation 4.1 with an exponentially weighted moving average conditional variance as per Equation 4.5 (with the same fitted parameters that are shown on Table 4.2) was used to compute 125 out-of-sample probabilistic predictions (one per out-of-sample month) for each of the six commodities analysed in this study. As a benchmark, the same months and commodities had probabilistic predictions made using an ARMA(p,q)-GARCH(1,1) model with Gaussian errors. Such model is commonly used in the forecasting of Value-at-Risk models for investment banks and other fund management institutions and assumes the log-returns follow the dynamics given by the following equation:

$$
\begin{aligned}
r_t &= \mu + \sum_{i=1}^{p} A_i r_{t-i} + \sum_{i=1}^{q} M_i \varepsilon_{t-i} + \varepsilon_t \\
\varepsilon_t | r_{t-1}, r_{t-2}, \ldots &\sim N\left(0, \sigma_t^2\right) \\
\sigma_t^2 &= \omega + \tau \varepsilon_{t-1}^2 + \upsilon \sigma_{t-1}^2 .
\end{aligned}
\tag{4.6}
$$

The estimated parameters of the ARMA(p,q)-GARCH(1,1) model are given on Table 4.4. These parameters were computed using the exact same sample that was used to fit the Logistic Mixture model. The hyperparameters p and q were chosen using the stepwise procedure of [104].

In order to have meaningful conclusions for this experiment, one needs to have a way to objectively determine which probabilistic predictions had the best out-of-sample performance. This can be challenging given that, even though the predictions are full probability distributions, the actual observations comprise always of a single point. Probabilistic predictions are normally assessed for quality of fit and predictive power using scoring rules [105]. An empirical application of probabilistic predictors and scoring rules in the field of Finance, with additional illustrations, can be seen in [106], where different predictions made for the probability of a lender defaulting are ranked and compared for out-of-sample predictive ability.

Table 4.4: Maximum Likelihood Estimates of ARMA(p,q)-GARCH(1,1) Parameters

Commodity	p	q	$\widehat{\mu}$	$\widehat{A}_1$	$\widehat{A}_2$	$\widehat{A}_3$	$\widehat{M}_1$	$\widehat{\omega}$	$\widehat{\tau}$	$\widehat{\upsilon}$
Brent Crude Oil	0	1	0.010				0.215	0.0006	0.150	0.784
WTI Crude Oil	1	0	0.004	0.191				0.0003	0.165	0.811
Gas Oil	1	0	0.001	0.162				0.0003	0.166	0.815
Natural Gas	0	0	-0.003					0.0005	0.062	0.914
Gold	0	0	-0.004					0.0001	0.137	0.817
Copper	3	1	0.002	-0.775	0.236	0.209	0.854	0.0001	0.033	0.966

Table 4.5: Comparison of Out-of-Sample Probabilistic Predictions

Commodity	ARMA-GARCH Average CRPS	Logistic Mixture Average CRPS	Paired T-Test p-Value	Reject Logistic $\leq$ ARMA-GARCH
Brent Crude Oil	0.044	0.046	0.01038	Yes
WTI Crude Oil	0.047	0.049	0.06882	No
Gas Oil	0.040	0.046	0.003153	Yes
Natural Gas	0.052	0.109	< 0.0001	Yes
Gold	0.026	0.038	< 0.0001	Yes
Copper	0.043	0.035	> 0.9999	No

In this manuscript, to determine which probabilistic predictor had the best out-of-sample performance, the Continuous Ranked Probability Score (CRPS) was used, based on the implementation of [107]. The CRPS is given by [105]:

$$CRPS(F,x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, dy$$

where $F(x)$ is the cumulative distribution function of x and $\mathbb{1}$ is the indicator function.

The CRPS rule used in the present thesis has the advantage that the expected score will be maximised if and only if the true distribution of the generative process is used as the probabilistic predictor being evaluated, provided that this distribution has a finite first moment. A scoring rule that satisfies such a property is also known as being a "strictly proper" scoring rule. The CRPS rule is not the only rule to have been used already in a financial application. In [108] the Logarithmic Score is used to compare density forecasts generated using three different probabilistic models for the United States monthly inflation rate. The authors noted that their choice of score was due to its simplicity (as the Logarithmic Score is simply the logarithm of the probability estimate for the actual outcome) and ease of interpretation given that the aggregation of the difference between several scores for different predictions can be easily interpreted as a weighted likelihood test.

For each of the six commodities, each of the 125 out-of-sample probabilistic predictions were assigned a CRPS when the prediction was based on the Logistic Mixture model and another CRPS for the prediction based on the ARMA(p,q)-GARCH(1,1) model. The two sets of 125 scores were then compared using a two sample paired T-test to detect with a Type I Error probability of 5% whether the ARMA(p,q)-GARCH(1,1) average score was equal to or greater than the average score for the Logistic Mixture model. Rejection of this hypothesis implies that the Logistic Mixture model provided significantly better forecasts than the ARMA(p,q)-GARCH(1,1) model. Results are shown on Table 4.5. It can be seen that the Logistic Mixture model provided a higher average score for five out of the six commodities, and in four of them the difference was statistically significant.

Finally, a few specific cases are shown to illustrate why the model implied by Equation 4.1 provides a better probabilistic predictor than an ARMA(p,q)-GARCH(1,1) predictor. Figure 4.3 shows a probabilistic prediction for the price at the end of Jul-2011 of the Gas Oil Futures contract maturing in the month of Sep-2011, as if performed at the end of Jun-2011.

At that time, the Gas Oil Futures Prices had come from a sequence of monthly appreciations in value. This caused the entire predicted distribution by the Logistic Mixture model to shift to the right. However, the immediately preceding month actually had a fall in the price of Gas Oil futures, which caused the ARMA(p,q)-GARCH(1,1) distribution to shift to the left, as there are no long term effects in this model. As the chart shows, the long term trend continued and the price of the commodity appreciated, as opposed to the prediction made used by the Logistic Mixture model. It is noted that the Logistic Mixture model also has a short term memory component; however, in the case of this particular situation, the positive long term effect was stronger than the negative short term effect. It is not always like this, though: in other cases in the same sample the opposite combination occurred, where the short term effect had a stronger impact in the prediction than the long term effect.

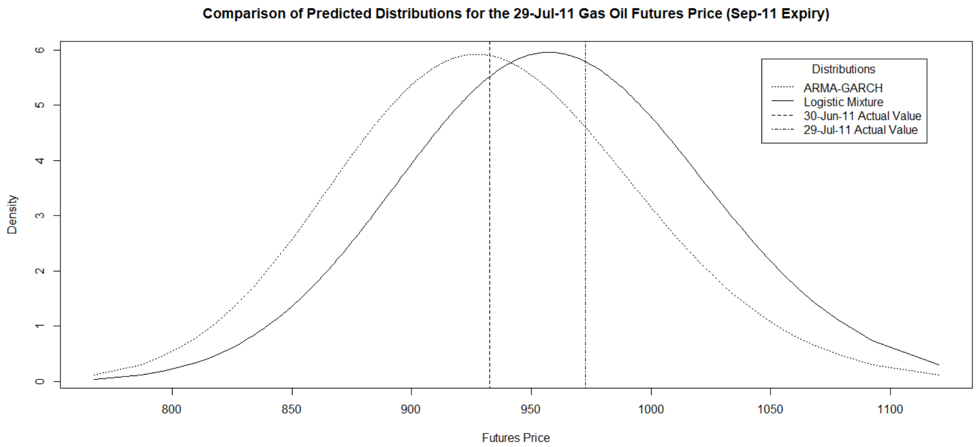

Figure 4.3: Illustration of Probabilistic Prediction for the Gas Oil Futures

It is noted that whilst the trading strategy for WTI Crude Oil did not generate excess profits, the model still can be adequately used to elaborate probabilistic forecasts of this commodity too. Trading strategies are mainly predictions of a single value (in this case the sign) whilst probabilistic predictions attempt to provide a more complete view of the uncertainty, predicting an entire distribution. Figure 4.4 compares the distributions between the prediction made with the use of the Equation 4.1 and the ARMA(p,q)-GARCH(1,1) model for the price of the WTI Crude Oil futures contract maturing in the month of Jan-2016, as if made in Oct-2015 for the price level at the end of Nov-2015.

It can be seen that, whilst both predicted distributions were relatively centered around the same location, the prediction made with the use of the Equation 4.1 had a left tail that was heavier than the right tail - this was due to previous falls in the price of WTI Crude Oil. It is worth noting that the shape of the distribution given by Equation 4.1 varies dynamically and the left tail is not always heavier than the right tail (for example during predicted uptrending markets), something that could not be achieved using the ARMA(p,q)-GARCH(1,1) model even if a skewed distribution is assumed for the error term, as this model cannot cope with time-varying skewness. As it can see in the picture, the futures price that actually materialised in Nov-2015 turned out to be very low indeed, in line with the increased left tail risk predicted by the model.

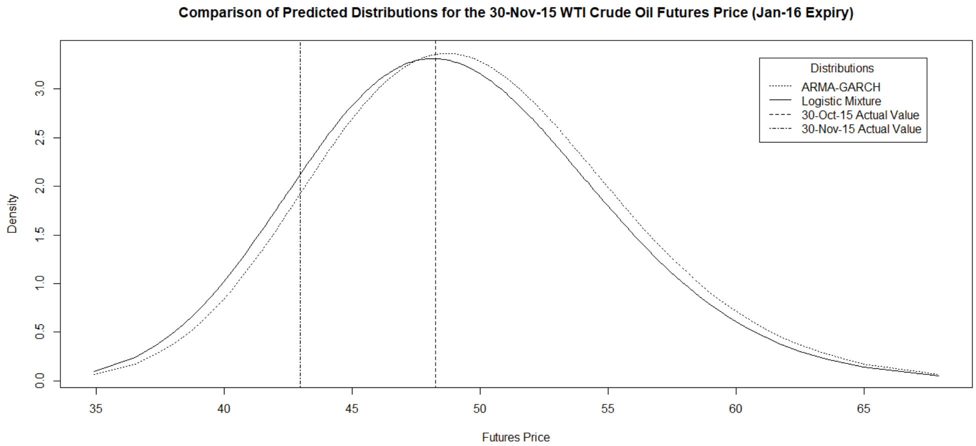

Figure 4.4: Illustration of Probabilistic Prediction for the WTI Crude Oil Futures

The previous figures illustrated that the model proposed in this chapter provides an enhanced accuracy in probabilistic forecasts for commodity futures prices against a simple ARMA-GARCH predictor. Therefore, the use of the model proposed in this chapter can provide additional information to economic planning around commodity price risk exposure.

Chapter 5

Option Pricing with Polynomial Chaos Expansion and Signed Path Dependence

In the previous chapters the concept of signed path dependence was introduced and fit to empirical observations of market returns in the real-world measure with a view to enhance the accuracy of time series models used for trading and risk management of assets with no optionality (i.e. cash assets and futures contracts). One might be interested to know if the previous findings can provide any contribution to the literature on assets with optionality. This chapter aims to provide this contribution by proposing a model to gauge to what extent the claimed presence of signed path dependence affects the behaviour of the market price of risk.

In order to reduce model specification risk and in accordance with the main approach followed by this thesis, we propose a non-parametric model for the extraction of the risk-neutral density with little assumptions over the underlying real-world process. Parametric models tend to underestimate the tails of the risk-neutral distribution in part due to an empirical property of asset returns called stochastic volatility [109] where the arrival of new information at a random arrival time causes a large temporary but persistent shock to the asset return distribution, making the presence of outlier returns much higher than most parametric models can reach. In fact, even parametric models for stochastic volatility tend to underestimate the tails of the risk-neutral distribution [110]. The risk-neutral distribution is also known to be asymmetric [72], which increases the challenge for parametric models.

One of the most well-known methods that still is widely used in the industry is based on an Edgeworth series expansion of the risk-neutral density [71], but this method has a drawback that it is not guaranteed to always provide a valid risk-neutral density as the resulting density is not guaranteed to be non-negative in all points or to integrate to one. In [73] a modified version of the Edgeworth expansion is provided where constraints are placed to ensure the resulting risk-neutral density is a valid density at all times. In [111] the risk-neutral distribution is obtained as a convolution of a fixed kernel and some arbitrary density function and a similar approach has been followed in [112] for estimating the stochastic volatility effect in a standard option pricing model.

The methods in the cited literature tend to have less power to explain relatively complex risk-neutral distributions and with the recent market and technological developments a vast amount of tradable option price data is becoming increasingly available, raising the question on whether it is valuable to follow approaches that expand the risk-neutral distribution into a Polynomial Chaos Expansion (PCE) that in theory can fit any distribution with infinite accuracy as long as there is enough data. In [74] a method is proposed to expand the density function using PCE and in [77] a method is proposed to obtain a risk-neutral distribution by expanding the return process via PCE. In [113] and [114] PCE is applied to the problem of risk-neutral density estimation in the context of stochastic volatility.

In this work we also use a PCE expansion based on the return process, as this ensures by construction that the risk-neutral density obtained is always a valid density, which will not necessarily be the case on an unconstrained expansion of the density function. Having said that, we contribute to the literature on proposing a data-driven method similar to the one described in [115] for a different problem. Our approach is the first, to the best of our knowledge, that calibrates a Polynomial Expansion of the random variable itself directly on option prices and simultaneously embeds the Fundamental Theorem of Asset Pricing directly into the derivation of the weights for the Polynomial Expansion.

In this chapter we develop a three stage process to obtain a full description of the market price of risk:

(i) First we fix a maturity date and assume a Hermite-Gaussian Polynomial Chaos representation for the risk-neutral distribution of the underlying asset returns between the measurement date and the maturity date. For each maturity, the polynomial weights are obtained by minimising the option pricing error via a least-squares procedure and subsequently all different maturities are connected via a stochastic interpolation scheme based on a stochastic bridge interpolator. Note that this step is independent of whatever functional form the underlying asset price process might have. We call it the model-free empirical risk-neutral distribution. We then couple the PCE calibrations at each time together by a stochastic bridge interpolator that allows us to obtain a consistent dynamic for the model free PCE option surface calibrations over time as a stochastic process.

(ii) Then we choose a model for the real-world asset price dynamics that can be calibrated to historical observations for the purposes of risk management, or can be embedded into an SDE model for the purposes of option pricing. In this thesis the model is proposed so as to include the effects of signed path dependence in the asset price behaviour. In this regard we develop an novel extension of a non-markovian non-parameteric dynamical process to capture features of signed path depedendence whilst preserving uncorrelated return structures that can arise in markets such as equities and currencies.

(iii) Finally, we estimate the functional form of the market price of risk by modelling the Radon-Nikodym derivative of the empirical risk-neutral distribution with respect to the modelled real-world distribution. Our model is able to capture a non-constant market price of risk using a natural cubic spline basis calibrated to the quantiles of the risk-neutral distribution.

The chapter is structured as follows: Section 5.1 describes the non-parametric method used to fit the risk-neutral distribution via PCE; Section 5.2 provides a comparision between our method and the Edgeworth expansion method, still considered by many in the industry as the gold standard for arbitrary derivation of the risk-neutral density; Section 5.3 gives a brief pause on risk-neutral work to focus on the impact of signed path dependence on the distribution of the real-world returns at any given European option expiry; Section 5.4 joins the work of the previous sections and Section 5.5 shows an empirical application that clarifies all the theoretical stages of our procedure. Section 6 concludes the work.

5.1 An Alternative Representation of the Risk-Neutral Distribution

We start by proposing a model-free methodology to obtain a risk-neutral distribution that can acommodate any arbitrary set of call prices of common maturity that can be calibrated to the observed market prices with an arbitrary degree of accuracy. We note that the applicability of such a method is not limited to any specific model for the asset price; instead, it creates a model-free link between the actual option prices observed and the models for the real-world probability distribution of asset prices and the respective market price of risk, which can be calibrated to historical data and reflect empirical properties of asset prices. So as to ensure there is always at least one risk-neutral measure which is equivalent to the real-world measure, markets are assumed arbitrage-free in the formulation of this work. This can be relaxed in subsequent works that extend these ideas.

The current section details the first stage of the process given in the beginning of the present chapter, starting with a key definition:

Definition 6. *Let $\psi_q\left(\xi_{i_1},\xi_{i_2},\ldots,\xi_{i_m}\right)$ an m-dimensional Hermite polynomial of order q, where $\xi_{i_j\in\{1,2,\ldots,\infty\}}$ is a set of i.i.d. standard Gaussian variables. The* Hermite Polynomial Chaos Expansion *of a random process $r(t)$ is given by*

$$r(t)=\tilde{w}_0(t)\psi_0+\sum_{i_1=1}^{\infty}\left(\tilde{w}_{i_1}(t)\psi_1\left(\xi_{i_1}\right)+\sum_{i_2=1}^{i_1}\left(\tilde{w}_{i_1,i_2}(t)\psi_2\left(\xi_{i_1},\xi_{i_2}\right)+\sum_{i_3=1}^{i_2}\left(\tilde{w}_{i_1,i_2,i_3}(t)\psi_3\left(\xi_{i_1},\xi_{i_2},\xi_{i_3}\right)+\ldots\right)\right)\right) \tag{5.1}$$

where $\tilde{w}_{\{\cdot\}}(t)$ are scalar coefficient functions to be estimated.

Equation 5.1 comprises an infinite sum, so in order to apply it to practical circumstances the maximum polynomial order must be truncated at a maximum order q and maximum dimension m. To simplify notation, the truncated version of Equation 5.1 can be written in the following form:

$$r(t)=\sum_{j=1}^{J}\tilde{w}_j(t)\psi_j\left(\boldsymbol{\xi}\right)+\varepsilon(m,q) \tag{5.2}$$

where $\boldsymbol{\xi}=(\xi_1,\xi_2,\ldots,\xi_m)$ is an m-dimensional Gaussian random vector with i.i.d. marginals of zero mean and unit variance, $\varepsilon(m,q)$ is a truncation error at a maximum order q and maximum dimension m, $\tilde{w}_0(t),\tilde{w}_1(t),\ldots,\tilde{w}_J(t)$ are the estimated weights at time t of the Hermite polynomials of this truncated expansion and $J = \sum_{i=0}^{q} \frac{(m+i-1)!}{(m-1)!i!}$ is the number of terms in this expansion after the truncation.

Remark 13. *We remark that numerous methods exist to guide the degree and truncation selection, see for instance [116]*

Now, let $(\pi_1, \pi_2, \ldots, \pi_M)$ be the prices (observed at time t) of M European call options on the same underlying asset of different strike prices $(K_1, K_2, \ldots, K_M)$ but equal time to maturity T and let $q(r,T)$ be the risk-neutral probability density of the log-return of that asset between times t and T, with S_t being the asset price at time t. The log return process $r(t)$ represented above in a PCE formulation itself is capturing the dynamics of the log return transformed from the the underlying asset price S_t at time t. Note, we assume in this work without loss of generality that there is a single underlying asset of interest.

As the markets are assumed arbitrage-free, it follows from the Fundamental Theorem of Asset Pricing [65] that there is a risk-neutral measure equivalent to the real-world probability measure and that the European call price for any combination of strike and maturity is given by the expected value of the discounted payoff of the option.

Therefore, if λ is the continuously compounded risk-free rate, as the payoff at time t of an European call option of strike K_v is given by $\max(0, S_t \exp(r(t)) - K_v)$, we have that the following holds for all $v \in \{1, 2, \ldots, M\}$ at time t:

$$\pi_v = \int_{-\infty}^{+\infty} \exp(-\lambda(T-t)) \max(0, S_t \exp(x) - K_v) q(x,T) \mathrm{d}x. \tag{5.3}$$

Let $\pi_v\left(\tilde{W}(t)\right)$ be the price of option v obtained by substituting Equation 5.2 into Equation 5.3. Given that the Hermite polynomials form an orthogonal basis with respect to the probability density function of a standard Gaussian variable, we have that:

$$\pi_v\left(\tilde{W}(t)\right)=\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\exp(-\lambda(T-t))\max\left(0,S_t\exp\left(\sum_{j=1}^{J}\tilde{w}_j(t)\psi_j\left(\xi\right)\right)-K_v\right)\varphi_m\left(\xi\right)\mathrm{d}\xi_1\ldots\mathrm{d}\xi_m \quad (5.4)$$

where $\varphi_m\left(\xi\right)$ is the density function of an m-dimensional Gaussian variable with i.i.d. marginals of zero mean and unit variance. Note that any arbitrary set of option prices can be fit with an arbitrary accuracy by taking $m\to\infty$ and $q\to\infty$.

It is known that Wiener first defined 'Homogeneous Chaos' as the span of Hermite polynomial functionals of a Gaussian process and then the class of Polynomial Chaos can be defined as a member of that set of representations. According to the Cameron-Martin theorem, the Fourier Hermite series converge to any L2 functional in the L2 sense.

In the context of stochastic processes, this implies that the homogeneous chaos expansion converges to any processes with finite second-order moments. Therefore, such an expansion provides a means of representing a stochastic process with Hermite orthogonal polynomials.

Other names such as 'Wiener chaos', 'Wiener-Hermite chaos', etc., have also been used in the literature.

In this thesis we are working in the non-heavy tailed finite quadratic variation context, see further details in [117].

We remind a well known fact that if $\psi_j\left(\xi\right)$ is an m-dimensional Hermite polynomial of order q, then $\psi_j\left(\xi\right)$ can also be expressed as a tensor multiplication of m unidimensional Hermite polynomials, each of order q in the product cross space of dimension m, yielding the following relationship:

$$\psi_j\left(\xi\right)=\prod_{i=1}^{m}\psi_{q(i,j)}(\xi_i) \quad (5.5)$$

where $\sum_{i=1}^{m}q(i,j)=q$ and $q(i,j)\in\{0,1,2,\ldots,q\}$. Therefore, Equation 5.2 becomes

$$r(t) = \sum_{j=1}^{J} \left(\tilde{w}_j(t) \prod_{i=1}^{m} \psi_{\hat{q}(i,j)}(\xi_i) \right) + \varepsilon(m,q) \tag{5.6}$$

where $\hat{q}$ is an $m \times J$ matrix containing all possible combinations of non-negative integers so that each column of $\hat{q}$ is unique and adds up to an integer between 0 and q.

The relationship given by Equation 5.6 is very useful in practical applications as it allows a simplified implementation of multidimensional Hermite polynomials. Moreover, the following theorem gives an important property of Equation 5.4:

Theorem 6 (Convexity of Implied Pricing Function). *$\pi_v\left(\tilde{W}(t)\right)$ is convex on $\tilde{w}_j(t)$ for all j.*

Proof. Given that $\exp(x)$ is convex on x and $\max(0,x)$ is also convex on x, as S_t is positive and K_v is a constant it follows that $\max(0, S_t \exp(x) - K_v)$ is convex on x and $\max\left(0, S_t \exp\left(\sum_{j=1}^{J} \tilde{w}_j(t)\psi_j(\xi)\right) - K_v\right)$ is convex on $\sum_{j=1}^{J} \tilde{w}_j(t)\psi_j(\xi)$.

Further, as the sum is a composition with an affine function, which preserves convexity, it is therefore also convex on $\tilde{w}_j(t)$ for all j. Finally, note that Equation 5.4 implies that $\pi_v\left(\tilde{W}(t)\right)$ is an expectation of a convex function under a multivariate Gaussian law. As the expectation is an operator that preserves convexity, it follows that $\pi_v\left(\tilde{W}(t)\right)$ is convex on $\tilde{w}_j(t)$ for all j. $\square$

Let $\widehat{W}(t) = (\widehat{w}_1(t), \widehat{w}_2(t), \ldots, \widehat{w}_J(t))$ be the least-squares estimator of $\tilde{W}(t) = (\tilde{w}_1(t), \tilde{w}_2(t), \ldots, \tilde{w}_J(t))$, i.e:

$$\widehat{W}(t) = \underset{\tilde{W}(t) \in \mathbb{R}^J}{\arg\min} \sum_{v=1}^{M} \left(\pi_v - \pi_v\left(\tilde{W}(t)\right)\right)^2. \tag{5.7}$$

Due to Theorem 6, there is a unique global optimum $\widehat{W}(t)$. However, an analytical closed-form expression for $\widehat{W}(t)$ cannot be found because the improper integral in Equation 5.4 does not have an analytical closed-form expression when $J > 2$ (when $J = 1$ or $J = 2$ the solution can be obtained using the Black-Scholes formula [118] as the underlying asset will simply follow a lognormal distribution) and therefore the partial derivatives required to be calculated in order to generate a closed-form expression for the solution of Equation 5.7 can only be calculated numerically. Having said that, also due to Theorem 6, a standard nonlinear optimisation procedure applied to Equation 5.7 is expected to converge to the global optimum weights provided that the integration error is small when calculating the improper integral in Equation 5.4 during the iterations of the nonlinear optimisation method being applied.

Remark 14. *Notice that the errors minimised by Equation 5.7 are dollar weighted pricing errors. In [119] the authors analyse a number of option pricing model error metrics, one of them being the dollar weighted pricing errors, which are given in Equation 4.1 of their paper. Equation 5.7 in this thesis is equivalent to Equation 4.1 in [119].*

In this manuscript we propose a Monte Carlo-based method to calculate the improper integral in Equation 5.4, which is particularly efficient in this setting since the target distribution can be sampled exactly and there is a tensor cross space formulation. Algorithm 3 lists the steps of the method. While due to Monte Carlo simulation noise the subsequent nonlinear optimisation step is not guaranteed to converge to the global optimum, Monte Carlo methods are better suited for the calculation of the integral in Equation 5.4 due to the high dimension of integration, which can make a quadrature based integral evaluation of the objective function passed to the optimisation algorithm computationally unfeasible as the integral will have to be calculated multiple times (possibly hundreds of times).

Algorithm 3 Monte Carlo integration procedure to calculate Equation 5.4

Input: A number of simulations N, a Gaussian dimension m, a polynomial truncation order q, a set of candidate PCE weights $\tilde{W}(t) = (\tilde{w}_1(t), \tilde{w}_2(t), \ldots, \tilde{w}_J(t))$ where J is the same as in Equation 5.2, the underlying asset price S_t, the annualised (continuously compounded) risk free rate λ and an option price π_v of strike prices K_v and time to maturity $T-t$

Output: The value of $\pi_v\left(\tilde{W}(t)\right)$ according to Equation 5.4

1: Generate matrix $\xi \in \mathbb{R}^{N\times m}$ with each $\xi_{u,w}$ drawn i.i.d. according to a standard Gaussian distribution.

2: For each row u of matrix ξ, calculate $\tilde{r}_u(t) = \sum_{j=1}^{J}\left(\tilde{w}_j(t)\prod_{i=1}^{m}\psi_{\hat{q}(i,j)}\left(\xi_{u,i}\right)\right)$ with $\hat{q}$ defined as in Equation 5.6.

3: Return $\frac{\exp(-\lambda(T-t))}{N}\sum_{u=1}^{N}\max(0, S_t\exp(\tilde{r}_u(t)) - K_v)$

Any standard nonlinear optimisation method can be used with an initial guess of $\tilde{W}(t) = 0$ and $\pi_v\left(\tilde{W}(t)\right)$ calculated as in Algorithm 3 to obtain the optimum weights that minimise Equation 5.7.

Remark 15. *In our experiments, we noted that regenerating the random matrix of Step 1 of Algorithm 3 at each iteration of the nonlinear optimisation method converged to weights that generated a risk-neutral distribution that fitted better the market prices, though the optimisation took longer to converge. Loosely speaking, one can expect that if the random matrix of Step 1 is generated only once and then kept the same throughout the entire optimisation, the weights will be fitted in part to the simulation noise of that particular matrix generation, whilst if this matrix is regenerated at each iteration of the optimisation method, no overfitting of this kind can occur.*

The results of this section are independent of the assumed price dynamics for the underlying asset. In fact, they could be used on their own if the only interest is to price other derivatives whose payoff function only depends on the asset price at the observed maturity. Having said that, the model can also be extended to cater for multi-period dynamics in a stochastically consistent framework that is again model free, through the use of a stochastic interpolation method based on Brownian bridges. We note, that generalised forms of stochastic interpolator are also possible here w.l.o.g. such as the work of [120].

In order to perform such a extension, we assume the following:

Assumption 2. *Let $t_1 > 0, t_2, \dots, t_T$ be the maturities for which there are quoted option prices available. For any intermediate time t such that $t_{i-1} \le t \le t_i$, under the risk-neutral distribution the following relationship holds:*

$$r(t) = \frac{t - t_{i-1}}{t_i - t_{i-1}} r(t_{i-1}) + \frac{t_i - t}{t_i - t_{i-1}} r(t_i) + B(t) \tag{5.8}$$

where $B(t)$ is a Brownian bridge between points t_{i-1} and t_i such that $B(t_{i-1}) = B(t_i) = 0$.

In other words, we assume that under the risk-neutral measure the cumulative asset return up to an unobserved (unquoted) option expiry date is given by a time-weighted average of the two adjacent observed maturities plus a random noise, which can be expressed using a Brownian bridge between the two observed maturities.

Now, as the weights $\tilde{w}_j(t_i), 1 \le j \le J, 1 \le i \le T$ are obtained at each t_i by minimising Equation 5.7 independently of $t_{j \neq i}$, under Assumption 2 Equation 5.8 yields

$$\begin{aligned} r(t) &= \frac{t - t_{i-1}}{t_i - t_{i-1}} r(t_{i-1}) + \frac{t_i - t}{t_i - t_{i-1}} r(t_i) + \frac{(t_i - t_{i-1}) - (t - t_{i-1})}{\sqrt{t_i - t_{i-1}}} \sqrt{\frac{t - t_{i-1}}{(t_i - t_{i-1}) - (t - t_{i-1})}} \xi_B + \varepsilon(m,q) \\ &= \sum_{j=1}^{J} \left(\frac{t - t_{i-1}}{t_i - t_{i-1}} \tilde{w}_j(t_{i-1}) \psi_j(\boldsymbol{\xi}_{\boldsymbol{i-1}}) + \frac{t_i - t}{t_i - t_{i-1}} \tilde{w}_j(t_i) \psi_j(\boldsymbol{\xi}_{\boldsymbol{i}}) + \right) + \sqrt{\frac{(t_i - t)(t - t_{i-1})}{t_i - t_{i-1}}} \xi_B + \varepsilon(m,q) \end{aligned} \tag{5.9}$$

where $\boldsymbol{\xi}_{\boldsymbol{i-1}}$ and $\boldsymbol{\xi}_{\boldsymbol{i}}$ are m-dimensional standard Gaussian variables of independent marginals and ξ_B is an independent standard Gaussian variable. Therefore, the risk-neutral distribution of $r(t)$ can be fully determined by the risk-neutral distributions of $r(t_{i-1})$ and $r(t_i)$.

Finally, we remark that as we are calibrating our estimation directly on observable option prices and ξ_B has expected value of zero, the Martingale restriction required on parameterisations of a risk-neutral distribution ([121]) is implicitly satisfied.

5.2 Comparative Analysis Against Existing Literature

In the literature review chapter and in the introductory section of the present chapter we noted the existence of previous work ([71], [72] and [73], list is not exhaustive) aiming to generalise the risk-neutral distribution so as to capture features present in observable option prices. Our work is similar in nature to the literature cited in the introductory section; however we claim our work contributes to the existing literature by performing a path-based expansion instead of a density-based expansion.

In density-based expansions, the modelled risk-neutral distribution obtained is not guaranteed by construction to be non-negative at all times and to always integrate to 1. To cater for that, the skew and kurtosis of the approximated distribution must be constrained. Chapter 18 of [122] gives specific details of the required restrictions on skew and kurtosis to ensure these expansions can produce a density. These requirements can be rather restrictive and we contribute to the literature by proposing an alternative approach based on a path-based polynomial expansion for the returns in Equation 5.2 that avoids the restrictions that would be needed in the conventional case where the risk-neutral distribution itself is approximated.

To illustrate the difference in concepts, we cite the work in [71], which is considered a pioneering work in the field of option pricing under the assumption of arbitrary stochastic processes for the underlying asset price, and compare it against our work. In our work, the functional form of the risk-neutral density

$q(r,T)$ is implicit, following from the Fundamental Theorem of Asset Pricing which ensures that the resulting implicit density function obeys the relationship given in Equation 5.3. Thereafter, the PCE weights are obtained based on the observed option prices and using the stochastic bridge interpolation for the path-based expansion given by Equation 5.2. If the risk-neutral density $q(r,T)$ is needed, it can be obtained by sampling the paths of Equation 5.2 given the estimated values of the weights.

The alternative approach to our path based stochastic path expansions is to instead expand a representation of the density such as that proposed in [71]. In this case the risk-neutral density $q_J(S_T)$, where S_T is the asset price at time T, has a specific functional form given by:

$$\begin{aligned} q_J(S_T) = a(S_T) + \frac{\kappa_2(q_J) - \kappa_2(a)}{2!} \frac{\mathrm{d}^2 a(S_T)}{\mathrm{d}S_T^2} - \frac{\kappa_3(q_J) - \kappa_3(a)}{3!} \frac{\mathrm{d}^3 a(S_T)}{\mathrm{d}S_T^3} \\ + \frac{\kappa_4(q_J) - \kappa_4(a) + 3(\kappa_2(q_J) - \kappa_2(a))^2}{4!} \frac{\mathrm{d}^4 a(S_T)}{\mathrm{d}S_T^4} + \varepsilon(S_T) \end{aligned} \tag{5.10}$$

where $\varepsilon(S_T)$ is a residual error, $a(S_T)$ is an approximating density function (the log-normal density was chosen in [71]) and $\kappa_i(q)$ is the i-th cumulant of the distribution corresponding to density q. For reference, the first three cumulants of any given distribution are the three central moments and the fourth cumulant is a measure of its kurtosis.

The model in [71] has the following potential challenges:

- $q_J(r,T)$ given by Equation 5.10 is not guaranteed to be a valid density function for a finite number of terms in the expansion and the magnitude of the residual error $\varepsilon(S_T)$ normally is analysed using numerical analysis given that only in a limited number of cases it's possible to derive a closed-form expression for the boundaries of such error.

- As a consequence of the above, the theoretical option prices derived in [71] are not directly related to the risk-neutral density implied by Equation 5.10 and, for a finite number terms in the expansion, may not be equal to the expected value of the discounted option cash flow under the risk-neutral measure.

- Equation 5.10 comprises of an expansion of the true density function around its central moments. This might not necessarily yield good behaviour at the tail of the distribution and, in some applications, the tail of the distribution is of high importance. In such cases it is better to use quantile based expansions such as the expansions proposed in [123].
- Equation 5.10 does not provide a way to ensure different maturities are connected to each other in a consistent way.

We believe that, by applying a path-based expansion calibrated directly on observed option prices, our model addresses the aforementioned challenges by construction.

Remark 16. *In Equation 12 of [71], fitted option prices are expressed as the standard Black-Scholes formula plus adjustments according to each of the central moments of the* true *risk-neutral probability density. We acknowledge that our formulation does not permit such interpretation for option prices, though we believe that in practical pricing applications this should not be a material drawback. On the other hand, the process structure of our Equation 5.9 used to obtain a full process provides a substantial contribution to pricing accuracy by generating a consistent connection between hedging points, analogous to consistent yield curve prediction methods such as [124]. Further, in our formulation the option prices will be directly related to the risk-neutral density and equal to the expected value of the discounted option cash flow under the risk-neutral measure.*

We conclude this section by noting the following differences between the other works cited in the introductory section and our own work:

- [72] is very similar in nature to our work but expands the distribution itself via PCE instead of the returns and as such is not guaranteed to provide a valid distribution at all times. In their application, this issue was not relevant as they only wanted to obtain estimates of central moments such as skewness and kurtosis but not the entire distribution; and

- [73] provides a good solution to the tail estimation via a multipoint Padé approximant (which is different in nature than a direct PCE expansion). Their work also expanded the distribution instead of the paths; however, an optimisation constraint was added to ensure the estimated density integrated to 1.

5.3 The Asset Price Process Under Signed Path Dependence

We now demonstrate the second stage of the three stage process mentioned in the beginning of the present chapter. In fact, any asset price process could be used to fulfill this stage; however, we would like to focus on a process that embeds the empirical findings of [7], [66] and [67] as motivated in the discussion in our literature review.

Let $p_t \in \mathbb{R}_+$ be the price of an asset at time $t \in \mathbb{N}_+$ and $r_t = \log(p_t/p_{t-1})$ be its log-return. To create a simple generative process for the log-returns that is compatible with the empirical findings listed in Chapter 2 and simultaneously does not generate an explicit functional form for its linear autocorrelation structure we postulate that such log-returns have a time dependent mean that oscillates between two states:

- a **high** state when it is positive, occurring whenever $p_{t-1} > p_{t-n-1}$; and

- a **low** state when it is negative, occurring otherwise.

The value $n \in \mathbb{N}_+$ corresponds to a lookback parameter, which can be either fixed as an input or estimated using the procedure described in [66].

Remark 17. *The above structure implies a positive correlation between* $\text{sgn}(p_{t-1} - p_{t-n-1})$ *and* r_t*, in line with the time series momentum effect reported in [7]. A dynamic correlation that changes over time between positive and negative can be easily acommodated by introducing a time varying correlation parameter, in line with the DMA classifier described in [66].*

In this section two models will be presented, the first will have a dynamic drift but homogeneous volatility structure; the second model will incorporate a form of dynamic stochastic volatility captured by a GARCH structure. Both models are consistent with the empirical findings listed in Chapter 2.

The aforementioned assumptions can be written in mathematical notation as follows:

Assumption 3. *At time t, the log-return r_t obeys the following distribution:*

$$r_t \sim \begin{cases} N(\mu_H, \sigma_t), & \textit{if } p_{t-1} > p_{t-n-1} \\ N(\mu_L, \sigma_t), & \textit{otherwise} \end{cases} \tag{5.11}$$

where $N(x, y)$ is a normally distributed variable with mean x and standard deviation y, $\sigma_t \in \mathbb{R}_+$, $\mu_H \in \mathbb{R}_+$ and $\mu_L \in \mathbb{R}_-$.

Assumption 4. *The variance of the process is constant through time, i.e. $\sigma_t = \sigma \in \mathbb{R}_+, \forall t$.*

Figure 5.1 illustrates the dynamics of the probability distribution of the asset log-returns under Assumptions 3 and 4 with $n = 2$ and so that the returns up to time $t = 2$ are assumed known. Notice that the distribution of $r_{t>2}$ is a

Gaussian mixture with 2^{t-3} components having non-zero weights equal to the probabilities of the sequences of positive and negative log-returns occuring in the order given by the tree and 2^{t-3} components having zero weights given that the mean of the distribution at time $t = 3$ is known with certainty.

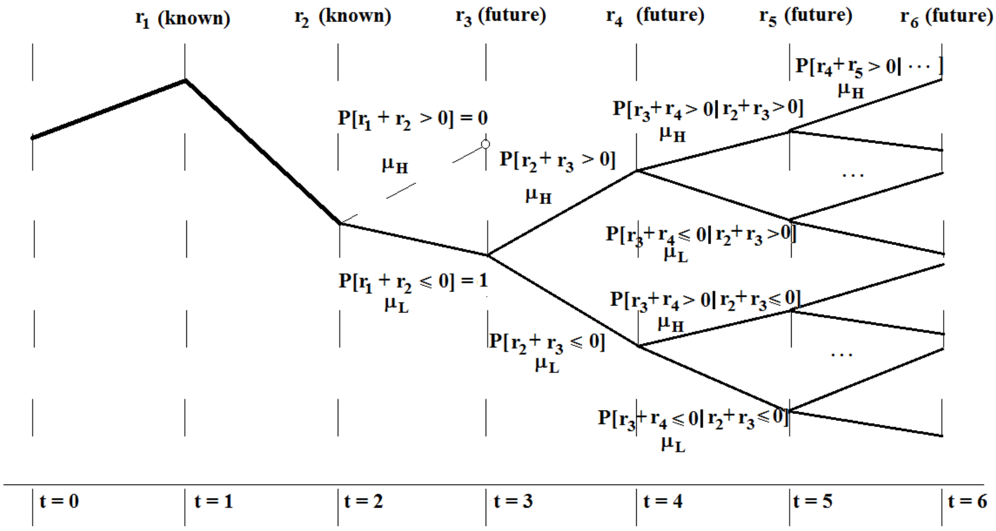

Figure 5.1: Illustration of the dynamics for the conditional mean of the process under Assumption 3

Throughout the remainder of the manuscript, the following notation is used: $\mathbb{1}\{\cdot\}$ is the indicator function, $\lfloor\cdot\rfloor$ is the greatest integer less than or equal to $\cdot$ and the operator r_{-n} is defined so that

$$r_{-n}(t) = \begin{cases} \sum_{i=1}^{n} r_{t-i}, & if\ n > 0 \\ 0, & \text{otherwise.} \end{cases}$$

We now derive results that will allow us to obtain the real-world distribution of asset returns under Assumptions 3 and 4. Knowing such distribution we can subsequently use the information from the empirical risk-neutral distribution to assess the impact of our assumptions on the risk premium.

The following theorem provides the base for our results:

Theorem 7 (Conditional Probability Density Function of Single Period Return). *Denote the probability space of the process r_t by $(\Omega, \mathcal{F}_t, P)$, where $\mathcal{F}_t$ is the natural filtration generated by this process. Let $f(r_T|\mathcal{F}_{t-1})$ be the conditional probability density function of r_T for $T \geq t$. If Assumptions 3 and 4 hold, then*

$$
\begin{aligned}
f(r_T|\mathcal{F}_{t-1}) = \mathbb{1}\{r_{-n}(t) \leq 0\} \sum_{i=0}^{2^{T-t}-1} (\Pi(i,t,T-t+1)\phi(r_T,i)) \quad + \\
\mathbb{1}\{r_{-n}(t) > 0\} \sum_{i=2^{T-t}}^{2^{T-t+1}-1} (\Pi(i,t,T-t+1)\phi(r_T,i))
\end{aligned}
\tag{5.12}
$$

where

$$
\Pi(i,t,w) = \prod_{j=2}^{w} \left(S_{B(i,j,w),\bar{B}(i,j,w)}\left(r_{-(n-j+1)}(t)\right)\right)^{b_{w-j+1}(i)} \left(F_{B(i,j,w),\bar{B}(i,j,w)}\left(r_{-(n-j+1)}(t)\right)\right)^{1-b_{w-j+1}(i)},
$$

$$
b_j(i) = \left\lfloor \frac{i}{2^{j-1}} \right\rfloor - 2\left\lfloor \frac{i}{2^j} \right\rfloor,
$$

$$
B(i,j,w) = \sum_{k=\max(0,j-n-1)}^{j-2} b_{w-k}(i), \qquad \bar{B}(i,j,w) = \min(w-1,n) - B(i,j,w)
$$

$$
\phi(r_T,i) = \frac{1}{\sqrt{2\pi\sigma^2}} \left(b_1(i)\exp\left(-\frac{(r_T-\mu_H)^2}{2\sigma^2}\right) + (1-b_1(i))\exp\left(-\frac{(r_T-\mu_L)^2}{2\sigma^2}\right)\right),
$$

$$
F_{u,w}(r) = \Phi\left(\frac{r-(u\mu_H+w\mu_L)}{\sigma\sqrt{u+w}}\right), \qquad S_{u,w}(r) = 1 - F_{u,w}(r)
$$

with $\Phi(\cdot)$ being the cumulative distribution function of a standard normal variable.

Remark 18. *The intuition behind Theorem 7 is that $r_T|\mathcal{F}_{t-1}$ follows a Gaussian mixture distribution of 2^{T-t+1} components, whose weights can be obtained by listing all integers from 0 to $2^{T-t+1}-1$ in binary form and, for each binary representation listed, mapping the i-th bit from right-to-left to realisations of $r_{-n}(T-i+1)$ so that a value of one corresponds to a positive realisation, as the following diagram illustrates. Each component will have a weight equal to the probability of that binary sequence occurring.*

$$\begin{array}{rlcl}
0: & (00\ldots00) & \rightarrow & (r_{-n}(t)\leq 0, r_{-n}(t+1)\leq 0,\ldots,r_{-n}(T-1)\leq 0, r_{-n}(T)\leq 0)\\
1: & (00\ldots01) & \rightarrow & (r_{-n}(t)\leq 0, r_{-n}(t+1)\leq 0,\ldots,r_{-n}(T-1)\leq 0, r_{-n}(T)> 0)\\
 & & \vdots & \\
2^{T-t+1}-2: & (11\ldots10) & \rightarrow & (r_{-n}(t)> 0, r_{-n}(t+1)> 0,\ldots,r_{-n}(T-1)> 0, r_{-n}(T)\leq 0)\\
2^{T-t+1}-1: & (11\ldots11) & \rightarrow & (r_{-n}(t)> 0, r_{-n}(t+1)> 0,\ldots,r_{-n}(T-1)> 0, r_{-n}(T)> 0)
\end{array}$$

Proof. From Equation 5.11 it follows that r_t is normally distributed with standard deviation σ and mean $\mathbb{1}\{r_{-n}(t)\leq 0\}\mu_L+\mathbb{1}\{r_{-n}(t)>0\}\mu_H$. It also follows that

$$f(r_{t+1}|\mathcal{F}_{t-1}) = \mathbb{1}\{r_{-n}(t)\leq 0\}\left(\Phi\left(\frac{r_{-(n-1)}(t)-\mu_L}{\sigma}\right)g_L(r_{t+1})+\left(1-\Phi\left(\frac{r_{-(n-1)}(t)-\mu_L}{\sigma}\right)\right)g_H(r_{t+1})\right) + \mathbb{1}\{r_{-n}(t)>0\}\left(\Phi\left(\frac{r_{-(n-1)}(t)-\mu_H}{\sigma}\right)g_L(r_{t+1})+\left(1-\Phi\left(\frac{r_{-(n-1)}(t)-\mu_H}{\sigma}\right)\right)g_H(r_{t+1})\right)$$

where $g_L(\cdot)$ is the density function of a normally distributed variable with mean μ_L and standard deviation σ and $g_H(\cdot)$ is the density function of a normally distributed variable with mean μ_H and standard deviation σ. Expanding to r_T we have that $f(r_{t+1}|\mathcal{F}_{t-1})$ will have the following form:

$$f(r_T|\mathcal{F}_{t-1}) = g_L\sum_{i=0}^{2^{T-t}-1}P[\Lambda_i]+g_H\sum_{i=2^{T-t}}^{2^{T-t+1}-1}P[\Lambda_i] \tag{5.13}$$

where Λ_i is a sequence of $T-t+1$ events in the probability space $(\Omega,\mathcal{F}_{t+j},P), 0\leq j\leq T-t$, each event of being either $r_{-n}(t+j)>0$ or $r_{-n}(t+j)\leq 0$ so that

$$P\left[\bigcup_{i=0}^{2^{T-t}-1}\Lambda_i\right]=1.$$

As such, $f(r_T|\mathcal{F}_{t-1})$ is a Gaussian mixture of 2^{T-t+1} components, half of them equal to $g_L(r_T)$ and half of them equal to $g_H(r_T)$. Now, to obtain the weights of each component, consider the sequence of binary digits $D=\{s_0,s_1,\ldots,s_{2^{T-t+1}-1}\}$

where each element of the sequence has $T-t+1$ digits so that:

$$s_i = (b_1(i), b_2(i), \ldots, b_{T-t+1}(i))$$
$$b_j(i) = \left\lfloor \frac{i}{2^{j-1}} \right\rfloor - 2\left\lfloor \frac{i}{2^j} \right\rfloor.$$

Consider the mapping $m : D \to \mathcal{F}$ where

$$m(s_i) = \mathcal{F}_{t-1} \bigcup_{j=1}^{T-t+1} \begin{cases} \left(\sum_{k=t+j-n-1}^{t+j-1} r_k > 0\right), & \text{if } b_j(i) = 1 \\ \left(\sum_{k=t+j-n-1}^{t+j-1} r_k \leq 0\right), & \text{if } b_j(i) = 0. \end{cases}$$

By construction, this mapping is a bijection between D and $\mathcal{F}_{T-1}$ and therefore $P[D] = 1$.

Now, let $\Lambda_i = m(s_i)$. As r_t is continuous, $P[r_k = 0] = 0$ and therefore we have that

$$P[\Lambda_i] = \prod_{j=2}^{T-t+1} P\left[\left(2b_{T-t-j-1}(i)-1\right)\sum_{k=t+j-n-1}^{t+j-1} r_k \geq 0 \middle| \bigcup_{w=2}^{j} \left((2b_{w-1}(i)-1)\sum_{k=t+w-n-2}^{t+w-2} r_k \geq 0\right)\right] \tag{5.14}$$

The conditional probabilities in Equation 5.14 will follow from Assumption 3 as the distribution of $\sum_{k=t+j-n-1}^{t+j-1} r_k$ conditional on knowing the sequence of events prior to time k is a sum of normally distributed variables of same standard deviation σ and means equal to either μ_H or μ_L.

Namely, if $X_1, X_2, \ldots, X_u$ are i.i.d. variables following a normal distribution with mean μ_H and standard deviation σ and $Y_1, Y_2, \ldots, Y_w$ are i.i.d. variables following a normal distribution with mean μ_L and standard deviation σ, defining

$$F_{u,w}(r) = P\left[\sum_{i=1}^{u} X_i + \sum_{i=1}^{w} Y_i \leq r\right] = \Phi\left(\frac{r-(u\mu_H + w\mu_L)}{\sigma\sqrt{u+w}}\right)$$

we have for any $j > 2$ that

$$P\left[\sum_{k=t+j-n-1}^{t+j-1} r_k \geq 0 \middle| \bigcup_{w=2}^{j} \left((2b_{T-t-w-2}(i)-1)\sum_{k=t+w-n-2}^{t+w-2} r_k \geq 0\right)\right] = S_{B(i,j,w),\bar{B}(i,j,w)}\left(r_{-(n-j+1)}(t)\right) \tag{5.15}$$

where

$$S_{u,w}(r) = 1 - F_{u,w}(r),$$

$$B(i,j,w) = \sum_{k=\max(0,j-n-1)}^{j-2} b_{w-k}(i)$$

and

$$\bar{B}(i,j,w) = \min(w-1,n) - B(i,j,w).$$

Therefore, applying Equation 5.15 to Equation 5.14 and subsequently to Equation 5.13 yields Equation 5.12, proving the main result. □

Theorem 7 gives the distribution of single period returns. As the price of an European option depends on the distribution of the price of the underlying asset at maturity and such price evolves based on the cumulative returns over mutiple periods, Theorem 7 needs to be extended so that the distribution of the cumulative returns is derived.

Nevertheless, such a distribution follows directly from Equation 5.12. To understand why, let $r_{t;T} = \sum_{i=t}^{T} r_t$ be the cumulative return between times t and T. Notice that the same reasoning to obtain Equation 5.13 will also hold for $r_{t;T}$ and hence $f\left(r_{t;T}|\mathcal{F}_{t-1}\right)$ can also be written in the following form:

$$f\left(r_{t;T}|\mathcal{F}_{t-1}\right) = \sum_{i=0}^{2^{T-t+1}-1} \phi\left(r_{t;T}, i\right) P[\Lambda_i]$$

where $P[\Lambda_i]$ has the same values and meaning as in Equation 5.13. To obtain $\phi\left(r_{t;T}, i\right)$, notice that for each event Λ_i the cumulative return conditional on this event $r_{t;T}|\Lambda_i$ is a sum of Gaussian variables, hence also Gaussian, meaning

$\phi\left(r_{t;T}, i\right)$ must be the density function of a Gaussian variable with some mean and variance.

Under Assumption 4 the variance of this Gaussian density is simply $(T-t+1)\sigma^2$. To obtain the mean of such Gaussian function, one just need to use the sequence $b_j(i)$ given in the body of Theorem 7 as this sequence contains the information required to compute how many times a Gaussian of mean μ_L was included in the sum $r_{t;T}$ and how many times a Gaussian of mean μ_H was included. This yields the following corollary:

Corollary 2 (Conditional Probability Density Function of Multi Period Cumulative Return). *Let $(\Omega, \mathcal{F}_t, P)$ be the probability space of the process r_t, $\mathcal{F}_t$ be the natural filtration generated by this process and $f\left(r_{t;T}|\mathcal{F}_{t-1}\right)$ be the conditional probability density function of $\sum_{i=t}^{T} r_t$ for $T \geq t$. If Assumptions 3 and 4 hold, then*

$$
\begin{aligned}
f\left(r_T|\mathcal{F}_{t-1}\right) = & \mathbb{1}\left\{r_{-n}(t) \leq 0\right\} \sum_{i=0}^{2^{T-t}-1} \left(\Pi(i,t,T-t+1)\phi\left(r_{t;T}, i\right)\right) \quad + \\
& \mathbb{1}\left\{r_{-n}(t) > 0\right\} \sum_{i=2^{T-t}}^{2^{T-t+1}-1} \left(\Pi(i,t,T-t+1)\phi\left(r_{t;T}, i\right)\right)
\end{aligned}
\tag{5.16}
$$

where

$$
\phi\left(r_{t;T}, i\right) = \Phi\left(\frac{r_{t;T} - \left(\mu_H \sum_{j=1}^{T-t+1} b_j(i) + \mu_L\left(T-t+1-\sum_{j=1}^{T-t+1} b_j(i)\right)\right)}{\sigma\sqrt{T-t+1}}\right)
\tag{5.17}
$$

and all other variables have the same values and meaning as in Theorem 7.

Assumption 4 can be relaxed to the case where σ_t follows a GARCH(1,1) process. To obtain this extension, we rely on the following theorem, whose proof can be found in [125]:

Theorem 8 (Time Aggregation of GARCH(1,1) Variables). *Let $\varepsilon(t)$ be a random* *variable following a Gaussian distribution with mean zero and conditional variance at time t given by*

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon^2(t-1).$$

Given an integer m, the random variable

$$\varepsilon_m(t) = \sum_{i=0}^{m-1} \varepsilon(t-i)$$

follows a Gaussian distribution with mean zero and conditional variance at time t given by

$$\sigma_{m,t}^2 = \omega_m + \beta_m\sigma_{m,t-m}^2 + \alpha_m\varepsilon_m^2(t)$$

where

$$\omega_m = \omega\frac{1-(\beta+\alpha)^m}{1-(\beta+\alpha)},$$

$$\alpha_m = (\beta+\alpha)^m - \beta_m$$

and $0 < \beta_m < 1$ is the solution of the quadratic equation

$$\frac{\beta_m}{1+\beta_m^2} = \frac{\beta(\beta+\alpha)^{m-1}}{1+\alpha^2\frac{1-(\beta+\alpha)^{2m-2}}{1-(\beta+\alpha)^2}+\beta^2(\beta+\alpha)^{2m-2}}.$$

Now, note that Equation 5.11 in Assumption 3 can be rewritten as the following sum

$$r_t = \mathbb{1}\{r_{-n}(t) \leq 0\}\mu_L + \mathbb{1}\{r_{-n}(t) > 0\}\mu_H + \varepsilon_t.$$

Therefore we can obtain an approximation of the conditional probability density function of r_t under GARCH(1,1) variance by substituting the time-scaled standard deviation in the denominator of Equation 5.17 with the time-aggregated standard deviation expression given by Theorem 8 and keeping unchanged the terms involving the location parameters. In other words, if $\varepsilon(t)$ is a random variable following a Gaussian distribution with mean zero and conditional variance at time t given by $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon^2(t-1)$, then we have that:

$$
\begin{aligned}
f\left(r_T|\mathcal{F}_{t-1}\right) \approx & \mathbb{1}\left\{r_{-n}(t) \leq 0\right\} \sum_{i=0}^{2^{T-t}-1} \left(\Pi\left(i,t,T-t+1\right)\phi\left(r_{t;T},i\right)\right) \quad + \\
& \mathbb{1}\left\{r_{-n}(t) > 0\right\} \sum_{i=2^{T-t}}^{2^{T-t+1}-1} \left(\Pi\left(i,t,T-t+1\right)\phi\left(r_{t;T},i\right)\right)
\end{aligned}
\tag{5.18}
$$

where

$$
\phi\left(r_{t;T},i\right) = \Phi\left(\frac{r_{t;T} - \left(\mu_H \sum_{j=1}^{T-t+1} b_j(i) + \mu_L\left(T-t+1-\sum_{j=1}^{T-t+1} b_j(i)\right)\right)}{\sqrt{\sigma^2_{(T-t),t}}}\right),
$$

$$
\sigma^2_{(T-t),t} = \omega_{(T-t)} + \beta_{(T-t)}\sigma^2_{(T-t),2t-T} + \alpha_{(T-t)}\varepsilon^2_{(T-t)}(t),
$$

$$
\omega_{(T-t)} = \omega\frac{1-(\beta+\alpha)^(T-t)}{1-(\beta+\alpha)},
$$

$$
\alpha_{(T-t)} = (\beta+\alpha)^(T-t) - \beta_{(T-t)},
$$

$0 < \beta_{(T-t)} < 1$ is the solution of the quadratic equation

$$
\frac{\beta_{(T-t)}}{1+\beta^2_{(T-t)}} = \frac{\beta\left(\beta+\alpha\right)^{T-t-1}}{1+\alpha^2\frac{1-(\beta+\alpha)^{2(T-t)-2}}{1-(\beta+\alpha)^2} + \beta^2\left(\beta+\alpha\right)^{2(T-t)-2}}
$$

and all other variables have the same values and meaning as in Theorem 7.

We note that the expression in Equation 5.18 is an approximation of the density (and not an exact equality) because the variances across intermediate time steps are different (as opposed to the case where they were constant, yielding an exact equality in Equation 5.16). If the GARCH(1,1) variance is assumed unconditionally stationary, the approximation error is expected to reduce as the terminal time horizon T increases due to decreasing differences over the variances across intermediate time steps.

5.4 The Risk Premium

In this section we demonstrate how to combine the two models from stage 1 and stage 2 of the three stage process introduced in the beginning of the present chapter to obtain a non-parametric representation of a time-varying risk premium under the assumption of signed path dependence given by the functional form introduced in Section 5.3. In order to do that and for the sake of clarity, we first enunciate the theorem supporting our results:

Theorem 9 (Radon-Nikodym Theorem)**.** *Let Ω be a set, $\mathcal{A}$ a σ-algebra on this set and P and Q two measures on $(\Omega, \mathcal{A})$. If P is σ-finite and absolutely continuous with regard to Q, then there exists a non-negative measurable function $\mathcal{D} : \Omega \leftarrow \mathbb{R}^+$ such that for any measurable set $\mathcal{S} \in \mathcal{A}$*

$$P(\mathcal{S}) = \int_{\mathcal{S}} \mathcal{D} \mathrm{d}Q. \tag{5.19}$$

The function $\mathcal{D}$ is also known as the Radon-Nikodym derivative of P with regard to Q.

Remark 19. *It's a well known result that if an asset price follows a Geometric Brownian Motion under the real-world measure, it will also follow a Geometric Brownian Motion under the risk-neutral measure, and the only change between the two measures is the drift of the process. In this special case, the Radon-Nikodym derivative is a constant function that expresses the amount of expected returns above the risk-free rate per unit of volatility - i.e. the risk premium. In our formulation, we will follow a similar rationale, but the final shape of the risk premium will not be a constant function. Instead, it will be a function that reflects the implied volatility structure observed in market option prices and its interaction with a real-world asset price process exhibiting signed path dependence.*

Now let L be the Lebesgue measure and, for any Borel set $\mathcal{S} \subseteq \mathbb{R}$, define the probability measure P as

$$P(\mathcal{S}) = \int_{\mathcal{S}} f(r_T|\mathcal{F}_{t-1})\mathrm{d}L, \tag{5.20}$$

where $f(r_T|\mathcal{F}_{t-1})$ is given by Equation 5.18. Also define the probability measure Q as

$$Q(\mathcal{S}) = \int_{\mathcal{S}} \hat{f}(r_T)\mathrm{d}L, \tag{5.21}$$

where $\hat{f}(r_T)$ is the probability density function of r_T given by Equation 5.9, whose weights are estimated by solving Equation 5.7 using the procedure in Section 5.1.

Both measures P and Q are absolutely continuous with regard to each other. Therefore, Theorem 9 guarantees the existence of the Radon-Nikodym derivative of P with regard to Q. Let $\mathcal{D}_{P,Q}$ be such derivative. The following assumption is required for our estimation of the functional form of the Radon-Nikodym derivative:

Assumption 5. *Let $\mathcal{D}_{P,Q}$ be the Radon-Nikodym derivative of measure P with regard to Q. The following equality holds:*

$$\mathcal{D}_{P,Q}(r) = \lim_{K\to\infty} \sum_{k=1}^{K} c_k \ell_k(r_T)$$

where c_k are unique coefficients and $\ell_k(\cdot)$ form a natural cubic spline basis, i.e.

$$\lim_{r\to-\infty} \ell_k(r) = \lim_{r\to+\infty} \ell_k(r) = 0.$$

Assumption 5 implies that the function $\mathcal{D}_{P,Q}$ can be approximated by a linear combination of a finite number of cubic spline basis functions, i.e.

$$\mathcal{D}_{P,Q}(r) = \sum_{k=1}^{K} c_k \ell_k(r_T) + \varepsilon(K) \tag{5.22}$$

where $\varepsilon(K)$ is an error term.

In a practical application, each natural cubic spline ℓ_k can have only one non-zero internal knot chosen by taking the probability distribution implied by Q and obtaining a percentile proportional to k. The boundary knots should be set to zero at very extreme percentiles of the distribution.

Remark 20. *While Assumption 5 implies that the modelled Radon-Nikodym derivative is zero at extremely rare events, this should not be a cause of concern given that the probability of extremely positive or extremely negative asset returns is nearly zero in either the risk-neutral or the real-world measures, so even if the "true" value of the Radon-Nikodym derivative at those points is different than zero, this will not invalidate the conclusions of our subsequent derivations. We also note that the quantile levels selected can be chosen arbitrarily on the support so numerically and in practice this compact support approximation is not at all consequential. Furthermore, if this was a concern one could readily construct a splice model where tail components were incorporated.*

Another required assumption is given as follows:

Assumption 6. *Let $\mathcal{D}_{P,Q}$ be the Radon-Nikodym derivative of measure P with regard to Q and $\mathbb{1}\{\cdot\}$ be the indicator function. $\mathcal{D}_{P,Q}(r)$ is independent of $\mathbb{1}\{r \leq x\}$ for all $x \in \mathbb{R}$.*

In other words, we assume that the likelihood of a future log-return exceeding or not any given threshold level is independent of the market price of risk for the return at that specific threshold. This assumption doesn't necessarily mean that the market price of risk is independent of the level of returns - quite the contrary, in our model there is a very explicit dependency between the market price of risk and the level of returns, as Section 5.5 will illustrate.

Now, notice that the following relationship holds for any given x:

$$P(r \leq x) = \int_{-\infty}^{+\infty} \mathbb{1}\{t \leq x\} f(r|\mathcal{F}_{t-1}) \mathrm{d}r = \int_{-\infty}^{+\infty} \mathbb{1}\{t \leq x\} \mathcal{D}_{P,Q}(r) \hat{f}(r) \mathrm{d}r$$

Therefore, one can find the adequate coefficients c_k by minimising some distance metric between the cumulative distribution function of r under the measure P and the cumulative distribution function of r under the measure Q. Moreover, let $Q^{-1}(r)$ be the inverse cumulative distribution of r under the measure Q. Defining a probability weighted distance metric as

$$d(P,Q,N) = \frac{\sum_{i=1}^{N} \left| \int_{-\infty}^{+\infty} \mathbb{1}\{t \leq Q^{-1}(\frac{i}{N+1})\} f(r|\mathcal{F}_{t-1}) \mathrm{d}r - \int_{-\infty}^{+\infty} \mathbb{1}\{t \leq Q^{-1}(\frac{i}{N+1})\} \mathcal{D}_{P,Q}(r) \hat{f}(r) \mathrm{d}r \right|}{N}$$

we have, due to Assumption 6 that

$$d(P,Q,N) = \frac{1}{N} \sum_{i=1}^{N} \left| P\left(r \leq Q^{-1}\left(\frac{i}{N+1}\right)\right) - Q\left(r \leq Q^{-1}\left(\frac{i}{N+1}\right)\right)\left(1 + \sum_{k=1}^{K} c_k E^Q[\ell_k(r)]\right) \right| \tag{5.23}$$

where $E^Q[\cdot]$ is the expectation taken under the probability measure Q. Therefore due to Equation 5.22 $d(P,Q,N)$ is linear on c_k and hence finding the values of c_k that minimise such distance can be solved very quickly and efficiently using linear programming solvers.

To formulate the linear optimisation problem of interest, let's define the approximated Radon-Nikodym derivative as

$$\mathcal{D}_{P,Q}(r;K) = \sum_{k=1}^{K} c_k \ell_k(r) \tag{5.24}$$

and substitute Equation 5.24 into Equation 5.23.

Also notice that $\mathcal{D}_{P,Q}(r;K)$ will only be a valid Radon-Nikodym derivative if its expected value under measure P is equal to 1, therefore this constraint has to be added to the linear optimisation problem as well, yielding the following problem:

$$
\begin{aligned}
\text{minimise} \quad & d(P,Q,N) \\
\text{subject to} \quad & \int_{-\infty}^{\infty}\left(\sum_{k=1}^{K} c_k \ell_k(r)\right) f(r|\mathcal{F}_{t-1})\mathrm{d}r = 1 \\
& c_k \in \mathbb{R}, k = 1,\ldots,K
\end{aligned}
\tag{5.25}
$$

where K and N are input parameters.

Let's now reduce the problem above to the general linear programming form.

First, let d_i be defined so that, for $i \in \{1,\ldots,2N\}$

$$d_{i+N} - d_i = P\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right) - Q\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right)\left(1+\sum_{k=1}^{K} c_k E^Q[\ell_k(r)]\right).$$

Note that if d_i is constrained to be always non-negative, we have that

$$d_{i+N} + d_i = \left| P\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right) - Q\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right)\left(1+\sum_{k=1}^{K} c_k E^Q[\ell_k(r)]\right)\right|. \tag{5.26}$$

Now, substituting Equation 5.26 into Equation 5.25 and constraining d_i to be always non-negative we obtain the optimum c_k by solving the linear minimisation problem given by Equation 5.27.

Such a minimisation problem can be easily and quickly solved with any linear programming package freely available. In the case of our own sythetic experiments, the solution was obtained in a single-core computer after less than one minute of calculation.

$$
\begin{aligned}
\text{minimise} \quad & \sum_{i=1}^{2N} d_i \\
\text{subject to} \quad & \sum_{k=1}^{K} c_k E^P[\ell_k(r)] && = 1, \\
& \sum_{k=1}^{K} c_k E^Q[\ell_k(r)] - \frac{d_i - d_{i+N}}{Q\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right)} && = \frac{P\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right)}{Q\left(r \le Q^{-1}\left(\frac{i}{N+1}\right)\right)}, && i = 1, \ldots, N, \\
& c_k && \in \mathbb{R}, && k = 1, \ldots, K, \\
& d_i && \ge 0, && i = 1, \ldots, 2N.
\end{aligned}
\tag{5.27}
$$

5.5 PCE Model Validation and a Full Empirical Example

In this section we calibrate the models of the previous sections to market prices of two weekly expiries of the S&P500 E-mini Call Options on Futures (European Style Exercise) observed on the Chicago Mercantile Exchange, demonstrating how all models interact with each other in an empirical setting. We also validate the Polynomial Chaos Expansion (PCE) model of Section 5.1 against an hypothetical scenario where the underlying asset follows a Geometric Brownian Motion (GBM), under which scenario the Black-Scholes formula provides the exact "correct" price of an option and under which the PCE expansion should also match nearly exactly the Black-Scholes prices. Further, we complete our PCE model validation by comparing its results to the ones of a competing model that follows the Edgeworth series expansion derived in [71] and discussed in Section 5.2.

The full list of market prices used to fit all models is given in Table 5.1. In the case of a GBM assumption, the volatility parameter was set to 10% per annum and the same maturities, strikes and underlying spot price were used as in the empirical setting. In both cases, for simplicity the risk-free rate was assumed 1.5% per annum, which was approximately the Overnight Indexed Swap rate at the time.

Table 5.1: E-mini S&P 500 Observed Prices, Value Date (t): 09-Jan-2020

Maturity Date (T):		10-Jan-2020	17-Jan-2020
Index (v)	Strike Price (K_v)	Option Price (π_v)	Option Price (π_v)
1	3200	76.25	80.50
2	3220	56.50	61.25
3	3235	41.75	47.75
4	3240	37.00	43.75
5	3245	32.25	39.50
6	3250	27.50	35.50
7	3255	23.00	31.50
8	3260	18.50	28.00
9	3265	14.50	24.25
10	3270	10.75	21.00
11	3275	7.50	18.00
12	3280	4.85	15.25
13	3285	2.95	12.50
14	3290	1.70	10.00
15	3295	0.90	7.75
16	3300	0.45	5.75
Spot Price (S_t):			3276

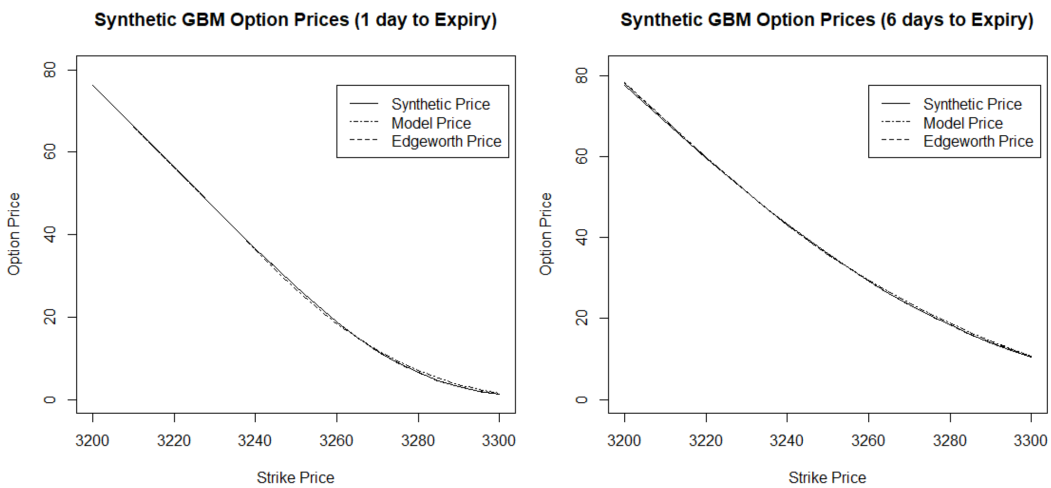

Figure 5.2: Illustration of fitted option prices under a hypothetical GBM assumption

Figure 5.2 shows the fitted Edgeworth series call prices and the fitted PCE call prices against the GBM assumption. It can be seen that both models provided a very close fit, as expected.

Figure 5.3 shows the fitted Edgeworth series call prices and the fitted PCE call prices against the market prices listed in Table 5.1. Small but noticeable differences can be seen on the 6-day expiry fit. The Edgeworth series model did not capture the volatility skew as well as the PCE model, with the Edgeworth series prices slightly underestimating the deeply in-the-money options and slightly overestimating the deeply out-of-the-money options. The PCE model provided a sligthly lower error overall as measured by the sum of errors squared over all strikes. However, for at-the-money strikes, the Edgeworth series prices had slightly better fit.

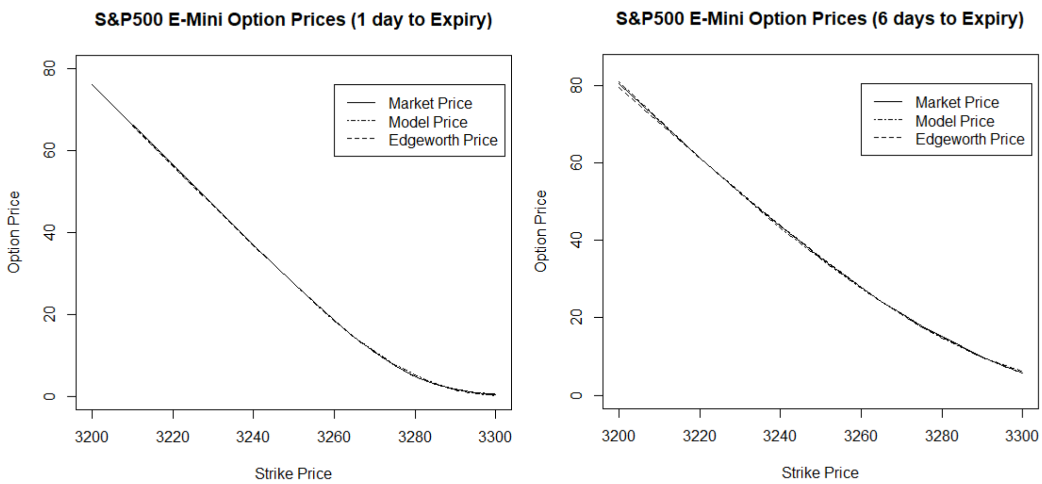

Figure 5.3: Illustration of different models fitted to S&P E-mini Options on Futures prices

The behaviour of the two models can be better compared if the option Greeks are also analysed under each model. The first two derivatives of the call option price against the underlying stock price are shown across all strike prices in Figure 5.4 (showing the first derivative, also known as "Option Delta") and Figure 5.5 (showing the second derivative, also known as "Option Gamma").

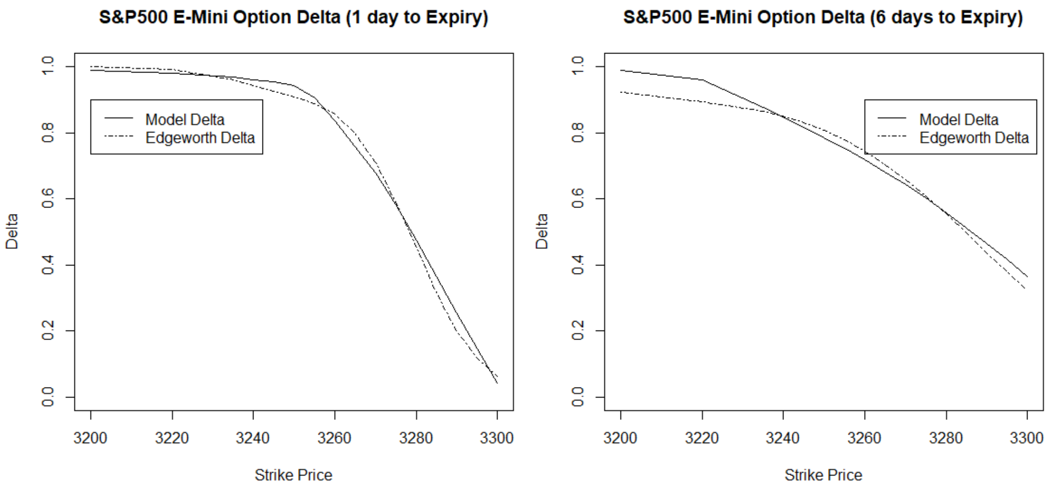

Figure 5.4: Illustration of S&P E-mini Options on Futures Delta under different models

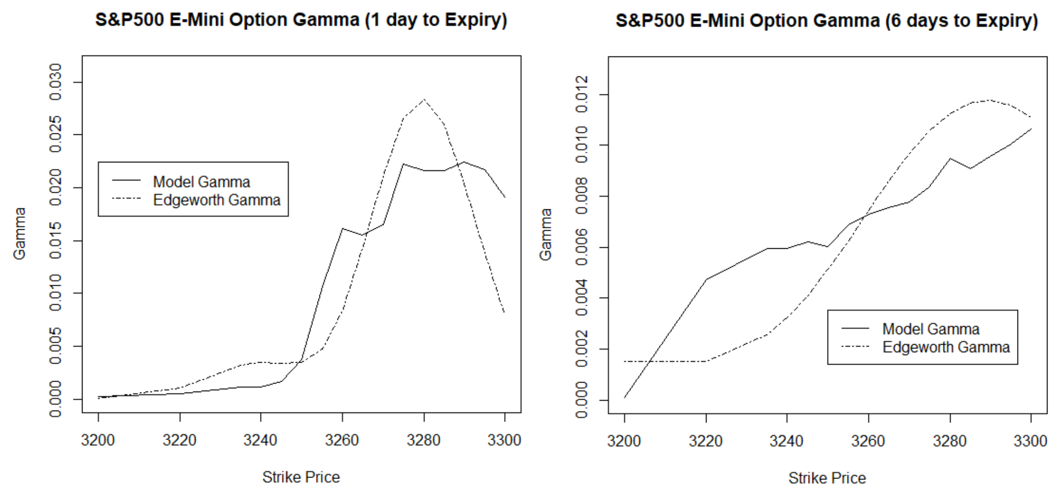

Figure 5.5: Illustration of S&P E-mini Options on Futures Gamma under different models

The option Gamma reveals the strongest feature differences between both models. The peak Gamma was lower in the case of the PCE model for both expiries being analysed, though in the case of the 1-day expiry the difference was much more pronounced. As Gamma peaks at at-the-money strikes, this feature can be seen as evidence that the option prices under the PCE model are less sensitive to small changes in the underlying asset than under the Edgeworth series model. This might be a desirable feature for hedging in markets of low volatility, as under the PCE model there would be less delta-hedging required given small offsetting market moves.

We also remark that the risk-neutral distribution under the PCE model was used directly to estimate the fitted prices, while the estimation of the Edgeworth series prices did not use the risk-neutral distribution. We also analysed the risk-neutral distribution generated by the Edgeworth series model and we noticed that it was not a valid risk-neutral distribution: it did not integrate to one and had negative values at the tails of the distribution. Therefore, adjustments would be needed to use this distribution to price more complex derivatives, rendering the risk-neutral distribution inconsistent with the market prices even if the model seemed to have a good fit.

Continuing with the empirical exercise, we have estimated a real-world model for the S&P500 assuming signed path dependence. We have accepted as valid Assumptions 3 and 4. Under such assumptions, three parameters are required: μ_H (the expected drift for the day after a positive daily return on the S&P500), μ_L (the expected drift for the day after a negative daily return on the S&P500) and σ. While Assumption 4 is known not to hold in an empirical setting, in the case of our particular application this will not be a cause of concern because we will be modelling a short time span, from 1 to 6 days, and use the at-the-money 6-day implied volatility as a proxy for σ, therefore taking into account the market views for the heteroskedasticity of market returns and the projected dispersion of the market returns' distribution at the time of the estimation. Such procedure yielded the value of $\sigma = 0.0865$.

The parameters μ_H and μ_L, were estimated based on the time series of the 20 years of S&P500 daily returns from 10-Jan-2000 to 09-Jan-2020, yielding 5032 observations. These returns were partitioned into two samples, one where the previous day's return was negative and another one when this condition was not true. The first return was not included in any sample (as it had no previous return). The parameters of interest were estimated as the sample mean in each of the two partitions. A two-sample t-test was made and confirmed a significant difference between the two means, with the null hypothesis of equal means between the two samples being rejected at a Type I error probability lower than 0.2%. Table 5.2 shows the results.

Table 5.2: S&P500 Signed Path Dependence Drift Parameters, 1 day interval

Parameter	Estimated Value	Estimate Std. Error
μ_H	-0.000337	0.000206
μ_L	0.000735	0.000271
Two sample t Test p-value		0.001648

We noticed that μ_H was significantly negative and μ_L was significantly positive, suggesting a reversal effect in the S&P500 returns (days of positive returns tending to be followed by days of negative returns and vice-versa). This is consistent with the findings of [38].

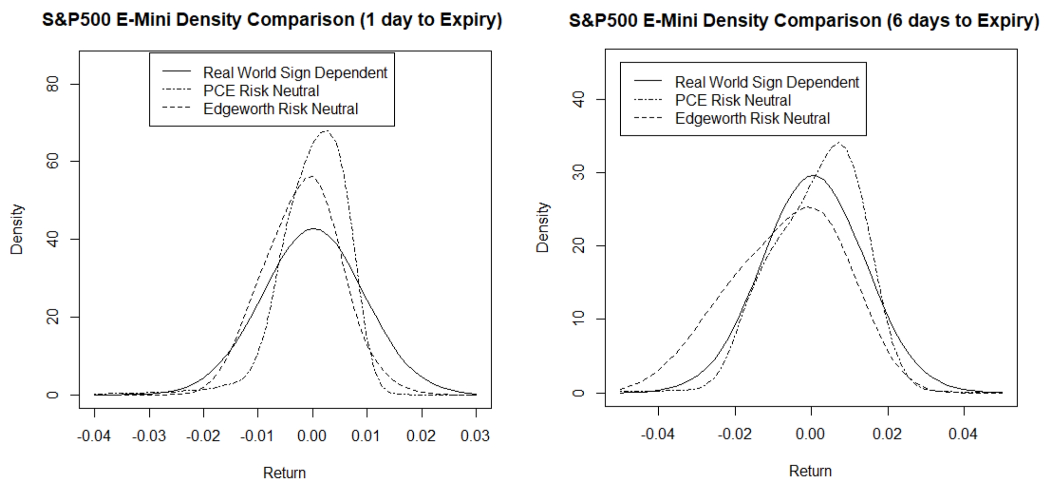

Figure 5.6: Illustration of S&P E-mini Risk-Neutral and Real-World Probability Densities

Figure 5.6 shows a comparative chart of the two probability measures estimated by our models together with the Risk-Neutral distribution obtained using the model given by [71]. It can be easily seen that the risk-neutral density is very asymmetric in both models, with a left tail that is much heavier than the one of the real-world density. Such effect is much more pronounced on options with one day to expire, which can be seen as evidence that the market prices embed large crash risks.

Figure 5.7 shows a comparative chart of the survival functions (one minus the CDF) implied by the distributions illustrated in Figure 5.6. This comparison is useful to illustrate the tail effects. It can be seen that the left tail is heavier in our formulation compared to the one in the model given by [71] as the survival function of our model approaches the value of 1 at much lower returns.

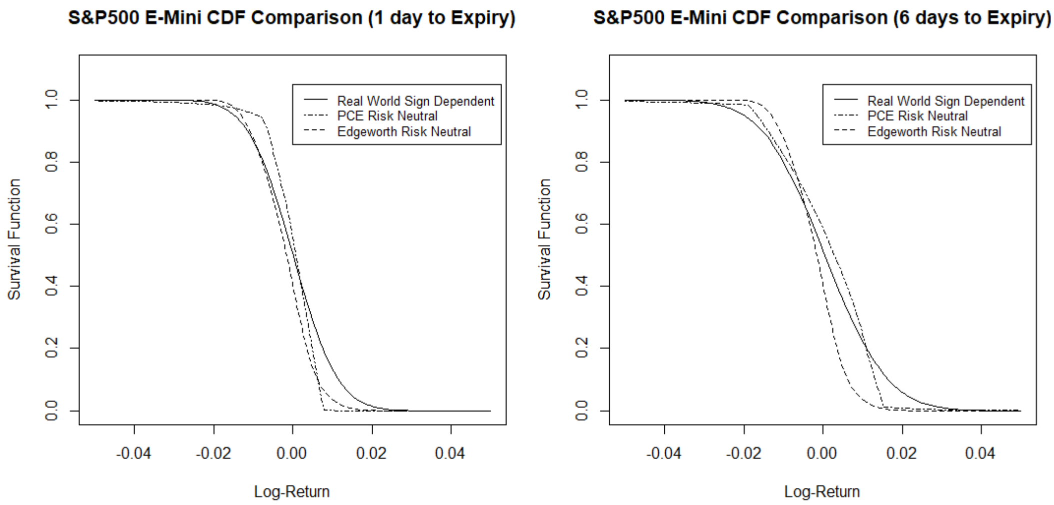

Figure 5.7: Illustration of S&P E-mini Risk-Neutral and Real-World Survival Functions

Table 5.3 lists the quantiles of the estimated densities. The 1-day 0.5% lower quantile of the risk-neutral density has almost twice the magnitude of the same quantile for the real-world density. Clearly this also reflects the thinner tails of the Gaussian distribution of Assumption 3, but the magnitude of the effect suggests a the risk-neutral density would exhibit a higher crash risk than the real-world density even if another distribution was assumed for the real-world stochastic process. Arguably this higher crash risk has been better captured by the PCE model, given that it had good fit for deeply in-the-money and out-of-the-money options.

Table 5.3: S&P500 Futures Risk-Neutral and Real-World Probability Distribution Quantiles

	1 day to Expiry		6 days to Expiry	
Quantile	Risk-Neutral	Real-World	Risk-Neutral	Real-World
0.005	-0.0374	-0.0226	-0.0483	-0.0321
0.01	-0.0270	-0.0208	-0.0308	-0.0288
0.05	-0.0089	-0.0148	-0.0168	-0.0200
0.1	-0.0063	-0.0115	-0.0139	-0.0154
0.15	-0.0052	-0.0092	-0.0114	-0.0124
0.2	-0.0042	-0.0074	-0.0089	-0.0100
0.25	-0.0033	-0.0059	-0.0065	-0.0079
0.3	-0.0023	-0.0046	-0.0043	-0.0060
0.35	-0.0014	-0.0033	-0.0023	-0.0042
0.4	-0.0006	-0.0021	-0.0004	-0.0025
0.45	0.0001	-0.0010	0.0013	-0.0010
0.5	0.0008	0.0002	0.0029	0.0005
0.55	0.0015	0.0013	0.0046	0.0021
0.6	0.0022	0.0024	0.0061	0.0036
0.65	0.0029	0.0036	0.0076	0.0053
0.7	0.0036	0.0048	0.0089	0.0071
0.75	0.0043	0.0062	0.0101	0.0091
0.8	0.0050	0.0077	0.0110	0.0112
0.85	0.0057	0.0094	0.0121	0.0136
0.9	0.0064	0.0117	0.0133	0.0165
0.95	0.0071	0.0149	0.0145	0.0210

To complete our empirical validation, we have implemented Equation 5.27 and solved it with values $K = 25$ and $N = 100$, which meant the natural cubic spline basis had 25 knots spaced over each 3.85% percentile (1/26) of the risk-neutral distribution.

We noted that in both expiries the optimal solution contained a positive value of high magnitude for the coefficient c_K and negative values of small magnitude for all other coefficients. To reduce parameterisation and avoid overfitting, we manually re-ran the model with only one non-zero coefficient (c_K) and obtained a very similar fit, meaning that a single natural cubic spline of one intermediate knot at the 96.15% percentile of the risk-neutral distribution was enough to provide good results.

Figure 5.8 shows the value of the fitted Radon-Nikodym derivative for the S&P 500 Futures for both expires considered in this manuscript. The results presented in this figure are based on the model with a single non-zero weight.

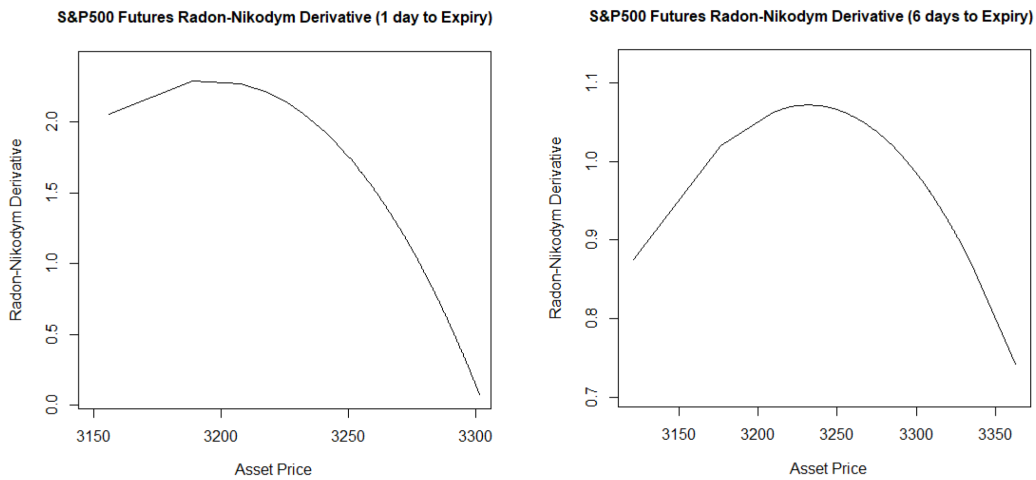

Figure 5.8: Illustration of the S&P E-mini Radon-Nikodym Derivative

In the 6-day expiry, it can be seen that the risk-neutral measure gives more weight to the center of the distribution than the real-world measure - though the left tail gets more weight than the right tail, which is consistent with the volatility skew present in equity markets. It is remarkable, however, that in the case of the 1-day expiry, there is such a heavy weight given to the left tail of the distribution that the value of the fitted derivative for positive returns decays very quickly and reaches almost zero at the 99.5% quantile, while it is still much higher than 1 at the 0.5% quantile, which is consistent with a very high crash risk being priced into the risk-neutral measure.

Chapter 6

Conclusions and Future Work

This research has proposed a novel framework to deal with unstructured serial dependence by introducing econometric models that are mainly focussed on the behaviour of the sign of innovations in the stochastic process rather than the innovations themselves. In this concluding chapter we list the main findings of our research and propose future steps.

6.1 General Conclusions

In Chapter 3 we proposed a new definition of dependence, which we called "Signed Path Dependence", that allowed a very structured statistical inference framework to detect unstructured serial dependence by focussing only on the sign of innovations in the stochastic process. In this way, irrespective of what is the true functional form of serial dependence present in a series, as long as it manifests itself in a consistent way over given lookback and forecast horizons, such a dependence can be detected by our test. This definition also has the power to simplify the interpretation of models with several parameters like ARFIMA by establishing boundaries in the values of the parameters that can be interpreted as simply positive or negative dependence.

Our statistical inference framework was validated both via simulation in synthetic models and in empirical data. In the case of empirical data for equity indices, the framework was able to detect significant dependence that could not be detected using a linear correlation model and was also able to feed a quantitative trading model that provided risk adjusted returns which were significantly superior to the ones of a Buy & Hold strategy and also robust to the presence of transaction costs.

In Chapter 4 a new model was proposed that embeds the information contained in previous returns into the shape of the probability distribution of future returns of risky assets in a way that is similar to a conditionally heteroskedastic variance model. This phenomenom, referred to as "Conditional Asymmetry", was shown to be statistically significant in several commodity future prices.

We also demonstrated the conditions necessary to predict the sign of future returns using proposed framework for conditional asymmetry and used these predictions to guide investment decisions in Commodity Futures. In our empirical experiment we obtained investment returns that were significantly superior to the ones that would have been obtained if no framework was used to guide investment decisions in the same financial instruments.

The chapter concluded by demonstrating another empirical application of our model for conditional asymmetry: probabilistic predictions. We compared the out-of-sample probabilistic predictions based on our model for Commodity Futures prices against predictions based on an ARMA(p,q)-GARCH(1,1) model. Using the Continuous Ranked Probability Score as a measure of out-of-sample performance for both models, we were able to determine that our model had superior performance than the benchmark ARMA(p,q)-GARCH(1,1) for all commodities other than Copper.

Chapter 5 provided an a novel framework to derive a risk-neutral distribution based on Polynomial Chaos Expansion that is consistent with option market prices across expiries and strikes and that provides consistent hedging points over time via a Brownian Bridge. This model has been compared against a well known Edgeworth series expansion. We have observed some empirical advantages in our approach such as a better fit for the extremes of the distribution and lower Gamma peaks providing a more stable hedging strategy.

We have also confirmed the findings of previous work in empirical finance literature showing a reversal effect in the real-world distribution of S&P500, where days of positive returns tended to be followed by days of negative returns and vice-versa. Moreover, comparing the real-world distribution with the risk-neutral distribution obtained with the PCE model, we found strong evidence that market prices embed very high crash risks - and such risks can be better captured by our PCE proposal.

Subsequently we have derived a novel way to explain the behaviour of the market price of risk based on such effect and applied it to the S&P500 call prices used to calibrate our PCE option pricing model, providing a valuable contribution to theoretical and empirical asset pricing.

6.2 Future Work

The work on each of three main chapters of this thesis can be expanded in future work.

The properties of test for signed path dependence introduced in Chapter 3 can be further explored. For example, it may be of interest to research what are the boundary conditions for parametric models under which the test might not work well. It may also be of interest to research about market conditions where the test can produce better predictive performance.

The work on conditional asymmetry introduced in Chapter 4 can be expanded by considering the inclusion of exogenous variables into Equation 4.1 and developing a more robust framework to deal with a multivariate envirnonment where multiple assets exhibit some form of joint conditional asymmetry. This can be accomplished by expanding Equation 4.1 so that it becomes

$$f_{r_t|n;p_{t-1},p_{t-2},p_{t-n-1}} = w(q_t)\varphi\left(\mu_H,\sigma_t^2\right) + (1-w(q_t))\varphi\left(\mu_L,\sigma_t^2\right)$$

where q_t is some functional form of exogenous variables or lagged returns of other assets and all other variables have the same meaning as in Equation 4.1. In this case, provided that q_t and the functional form of $w(q_t)$ are chosen appropriately, $f_{r_t|n;p_{t-1},p_{t-2},p_{t-n-1}}$ will be a well-defined probability density function.

From a practical point of view, our work on conditional asymmetry can be further explored in the fields of market risk management and portfolio optimisation, especially if the model is expanded to a multivariate envirnonment. Several financial institutions are legally required to control their market risk exposure by performing predictions of what would be their financial losses at a given quantile (also known as Value-at-Risk) and predictions of what would be the expected value of the financial loss beyond that given quantile (also known as Expected Shortfall). These metrics require knowledge of the probability density function of future returns. Therefore, any gain in predictive accuracy of the probability density function being used will be a potentially material contribution to the overall financial system.

Finally, this thesis is concluded by pointing out that the work on Polynomial Chaos Expansion and Stochastic Bridge Interpolators introduced in Chapter 5 can also be expanded to include other features such as stochastic discount factors, fatter-tail distributions for the bridge interpolation and Generalized Polynomial Chaos Expansion to expansions in polynomials orthogonal with respect to non-Gaussian probability measures.

As with any research, there is always scope for further refinement and improvement.

Bibliography

[1] Adam Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. William Strahan, London, 1776.

[2] Jules Regnault. *Calcul des Chances et Philosophie de la Bourse*. Mallet-Bachelier et Castel, Paris, 1863.

[3] Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, Paris, 1900.

[4] Benjamin Graham. *The Intelligent Investor*. Harper & Row Publishers Inc, New York, 1949.

[5] Eugene Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25:383–417, 1970.

[6] Eugene F. Fama and Kenneth R. French. Size, Value, and Momentum in International Stock Returns. *Journal of Financial Economics*, 105:457––472, 2012.

[7] Tobias J. Moskowitz, Yao H. Ooi, and Lasse Heje Pedersen. Time Series Momentum. *Journal of Financial Economics*, 104:228–250, 2012.

[8] William M. K. Trochim. *Research Methods Knowledge Base*. `https://socialresearchmethods.net/kb/introduction-to-evaluation/`, 2020. Accessed on March 31, 2020.

[9] Eckhard Platen and David Heath. *A Benchmark Approach to Quantitative Finance*. Springer Finance, 2006.

[10] Eugene Fama. Mandelbrot and the Stable Paretian Hypothesis. *Journal of Business*, 36:420–429, 1963.

[11] Eugene Fama. The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105, 1965.

[12] Eugene Fama. Random Walks in Stock Market Prices. *Journal of Business*, 21:55–59, 1965.

[13] Nabeel Al-Loughani and David Chappell. On the Validity of the Weak-form Efficient Markets Hypothesis Applied to the London Stock Exchange. *Applied Financial Economics*, 7:173–176, 1997.

[14] Ioana-Andreea Boboc and Mihai-Cristian Dinica. An Algorithm for Testing the Efficient Market Hypothesis. *PLoS ONE*, 8, 2013.

[15] Andrew Urquhart and Frank McGroarty. Are Stock Markets Really Efficient? Evidence of the Adaptive Market Hypothesis. *International Review of Financial Analysis*, 47, 2016.

[16] Leslie A. Jeng, Andrew Metrick, and Richard Zeckhauser. Estimating the Returns to Insider Trading: A Performance-Evaluation Perspective. *Review of Economics and Statistics*, 85:453–471, 2003.

[17] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944.

[18] Milton Friedman and L. J. Savage. The Utility Analysis of Choices Involving Risk. *The Journal of Political Economy*, 56:279–304, 1948.

[19] Harry M. Markowitz. Portfolio Selection. *The Journal of Finance*, 7:77–91, 1952.

[20] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47:263–292, 1979.

[21] Shlomo Benartzi and Richard Thaler. Myopic Loss Aversion and the Equity Premium Puzzle. *The Quarterly Journal of Economics*, 110:73–92, 1995.

[22] Rajnish Mehra and Edward C. Prescott. The Equity Premium: A Puzzle. *Journal of Monetary Economics*, 15:145–161, 1985.

[23] Nicholas Barberis and Ming Huang. Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review*, 98:2066–2100, 2008.

[24] Clifton Green and Byoung Hwang. Initial Public Offerings as Lotteries: Skewness Preference and First-Day Returns. *Management Science*, 58:432–444, 2012.

[25] Rama Cont. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 1:223–236, 2001.

[26] Wayne A. Fuller. *Introduction to Statistical Time Series*. John Wiley and Sons, 1976.

[27] Peter C.B. Phillips and Pierre Perron. Testing for Unit Root in Time Series Regression. *Biometrika*, 75:335–346, 1988.

[28] Denis Kwiatkowski, Peter Phillips, Peter Schmidt, and Yongcheol Shin. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics*, 54:159–178, 1992.

[29] Mohamed Khaled Al-Jafari. Variance Ratio Test and Weak-form Efficiency of Bahrain Bourse.

[30] Kian-Ping Lim and Robert Brooks. The Evolution of Stock Market Efficiency Over Time: A Survey of the Empirical Literature. *Journal of Economic Surveys*, 25:69–108, 2011.

[31] Andrew W. Lo and A. Craig MacKinlay. Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. *The Review of Financial Studies*, 1:41–66, 1988.

[32] J. M. Griffin, P. J. Kelly, and F. Nardari. Do Market Efficiency Measures Yield Correct Inferences? A Comparison of Developed and Emerging Markets. 23, 2010.

[33] Kian-Ping Lim and Weiwei Luo and Jae H. Kim . Are US Stock Index Returns Predictable? Evidence from Automatic Autocorrelation-based Tests. *Applied Economics*, 45(8):953–962, 2013.

[34] Ching-Fan Chung. Estimating a Generalized Long Memory Process. *Journal of Econometrics*, 73:237–259, 1996.

[35] Bo Qian and Khaled Rasheed. Hurst Exponent and Financial Market Predictability. Paper presented at the 2004 IASTED conference on Financial Engineering and Applications, 2004.

[36] Jin Lee and Yongmiao Hong. Testing for Serial Correlation of Unknown Form Using Wavelet Methods. *Econometric Theory*, 17:386–423, 2001.

[37] Akindynos-Nikolaos Baltas and Robert Kosowski. Momentum Strategies in Futures Markets and Trend-following Funds. Paper presented at the 2012 Finance Meeting EUROFIDAI-AFFI, 2013.

[38] Pasquale D. Corte, Robert Kosowski, and Tianyu Wang. Market Closure and Short-Term Reversal. Paper presented at the Asian Finance Association (AsianFA) 2016 Conference, 2016.

[39] Robert Novy-Marx. Is Momentum Really Momentum? *Journal of Financial Economics*, 103:429–453, 2012.

[40] Yifan Chena and Zhao Huainan. Informed Trading, Information Uncertainty, and Price Momentum. *Journal of Banking & Finance*, 36:2095–2109, 2012.

[41] David A. Lesmonda, Michael J. Schill, and Chunsheng Zhou. The Illusory Nature of Momentum Profits. *Journal of Financial Economics*, 71:349–380, 2004.

[42] Anurag N. Banerjee and Chi-Hsiou D. Hung. Active Momentum Trading Versus Passive Naive Diversification. *Quantitative Finance*, 13:655–663, 2013.

[43] Campbell R. Harvey and Yan Liu. Evaluating Trading Strategies. *The Journal of Portfolio Management*, 40:108–118, 2014.

[44] Yoav Benjamini and Daniel Yekutieli. The Control of the False Discovery Rate in Multiple Testing Under Dependency. *Annals of Statistics*, 29:1165–1188, 2001.

[45] Ron Kohavi. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 12:1137–1143, 1995.

[46] Bo Qian and Khaled Rasheed. Stock Market Prediction with Multiple Classifiers. *Applied Intelligence*, 26:25–33, 2007.

[47] Rohit Choudhry and Kumkum Garg. A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology: International Journal of Computer and Information Engineering*, 2:689–692, 2008.

[48] Wei Huang, Yoshiteru Nakamori, and Shou-Yang Wang. Forecasting Stock Market Movement Direction with Support Vector Machine. *Computers & Operations Research*, 32:2513–2522, 2005.

[49] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33:2223–2273, 2020.

[50] Anupam Dutta. Parametric and Nonparametric Event Study Tests : a Review. *International Business Research*, 7:136–142, 2014.

[51] Stephen J. Brown and Jerold B. Warner. Measuring Security Price Performance. *Journal of Financial Economics*, 8:205–258, 1980.

[52] Charles J. Corrado and Terry L. Zivney. The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns. *The Journal of Financial and Quantitative Analysis*, 27:465–478, 1992.

[53] James Clay Dolley. Characteristics and Procedure of Common Stock Split-ups. *Harvard Business Review*, 11:316–326, 1933.

[54] John M. Maheu and Thomas H. McCurdy. Identifying Bull and Bear Markets in Stock Returns. *Journal of Business and Economic Statistics*, 18:100–112, 2000.

[55] James D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384, 1989.

[56] Keris Lahiff. *Netflix, Others are Market Bright Spots After S&P500's Worst First Quarter Ever*. CNBC Trading Nation, `http://alturl.com/sxir6`, 2020. Accessed on April 04, 2020.

[57] Massimo Guidolin and Allan Timmermann. Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns. *Journal of Business and Economic Statistics*, 115:111–143, 2005.

[58] Henri Nyberg. Predicting Bear and Bull Stock Markets with Dynamic Binary Time Series Models. *Journal of Banking & Finance*, 37:3351–3363, 2013.

[59] Henri Nyberg. Forecasting the Direction of the US Stock Market with Dynamic Binary Probit Models. *International Journal of Forecasting*, 27:561–578, 2011.

[60] Harri Ponka. Predicting the Direction of US Stock Markets Using Industry Returns. *Empirical Economics*, 52:1451–1480, 2017.

[61] Brian Hurst, Yao Hua Ooi, and Lasse Heje Pedersen. A Century of Evidence on Trend-Following Investing. *Journal of Portfolio Management*, 44:15–29, 2017.

[62] Andrew C. Szakmary, Qian Shen, and Subhash C. Sharmac. Trend-following Trading Strategies in Commodity Futures: A Re-examination. *Journal of Banking & Finance*, 34:409–426, 2010.

[63] Kurt Brannas and Niklas Nordman. An Alternative Conditional Asymmetry Specification for Stock Returns. *Applied Financial Economics*, 13:537–541, 2003.

[64] Jennifer Conrad, Robert F. Dittmar, and Eric Ghysels. Ex Ante Skewness and Expected Stock Returns. *The Journal of Finance*, 68:85–124, 2013.

[65] John Michael Harrison and David M. Kreps. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, pages 381—-408, 1979.

[66] Fabio S. Dias and Gareth W. Peters. A Non-Parametric Test and Predictive Model for Signed Path Dependence. *Computational Economics*, 2019.

[67] Fabio S. Dias. Using Conditional Asymmetry to Predict Commodity Futures Prices. Working Paper, 2020.

[68] Oldrich Vasicek. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5:177–188, 1977.

[69] Andrew W. Lo and Jiang Wang. Option Pricing When Asset Returns Are Predictable. *The Journal of Finance*, 1995.

[70] Lin Chen. Stochastic Mean and Stochastic Volatility - A Three-Factor Model of the Term Structure of Interest Rates and Its Application to the Pricing of Interest Rate Derivatives. *Financial Markets, Institutions & Instruments*, 5:1–88, 1996.

[71] Robert Jarrow and Andrew Rudd. Approximate Option Valuation for Arbitrary Stochastic Processes. *Journal of Financial Economics*, 10:347–369, 1982.

[72] Charles J. Corrado and Tie Su. Skewness and Kurtosis in S&P500 Index Returns Implied by Option Prices. *The Journal of Financial Research*, 19:175–192, 1996.

[73] Guillaume Bagnarosa, Emmanuel Jurczenko, and Bertrand Maillet. An Implicit Martingale Restriction in a Closed-form Higher Order Moments Option Pricing Formula based on Multipoint Pade Approximants. *24ème Journées de Microéconomie Appliquée*, 2007.

[74] Damien Ackerer and Damir Filipovic. Option Pricing with Orthogonal Polynomial Expansions. *Swiss Finance Institute Research Paper Series, issue 17-41*, 2017. Swiss Finance Institute.

[75] Norbert Wiener. The Homogeneous Chaos. *Am. J. Math.*, 60:897—-936, 1938.

[76] Anthony O'Hagan. Polynomial Chaos : A Tutorial and Critique from a Statistician ' s Perspective. *SIAM/ASA Journal on Uncertainty Quantification*, 2013.

[77] Luca Di Persio, Gregorio Pellegrini, and Michele Bonollo. Polynomial Chaos Expansion Approach to Interest Rate Models. *Journal of Probability and Statistics*, 2015. Article ID 369053.

[78] John C. Cox, Jonathan E. Ingersoll, and Steve A. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53:385—-407, 1985.

[79] Herman Wold. *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell Book Co., 1954.

[80] Clive W. J. Granger and Roselyne Joyeux. An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1:15–29, 1980.

[81] Palahenedi Hewage Diananda. The Central Limit Theorem for m-dependent Variables Asymptotically Stationary to Second Order. *Mathematical Proceedings of the Cambridge Philosophical Society*, 50:287–292, 1954.

[82] Gareth W. Peters and Pavel V. Shevchenko. *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. John Wiley & Sons, 2015.

[83] Georgia Lee Thompson. Asymptotic Distribution of Rank Statistics under Dependencies with Multivariate Application. *Journal of Multivariate Analysis*, 33:183–211, 1990.

[84] Bradley Efron and Rob Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.

[85] Yongmiao Hong. Consistent Testing for Serial Correlation of Unknown Form. *Econometrica*, 64:837–864, 1996.

[86] Amber Anand, Paul Irvine, Andy Puckett, and Kumar Venkataraman. Performance of Institutional Trading Desks: An Analysis of Persistence in Trading Cost. *The Review of Financial Studies*, 25:557–598, 2012.

[87] Jonathan Brogaard, Terrence Hendershott, Stefan Hunt, and Carla Ysusi. High-Frequency Trading and the Execution Costs of Institutional Investors. *The Financial Review*, 49:345–369, 2014.

[88] Andrew W. Lo and Alexander Remorov. Stop-loss Strategies with Serial Correlation, Regime Switching, and Transaction Costs. *Journal of Financial Markets*, 34:1–15, 2017.

[89] Andrew Ang and Allan Timmermann. Regime Changes and Financial Markets. *Annual Review of Financial Economics*, 4:313–337, 2012.

[90] Robert F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50:987–1008, 1982.

[91] Francis X. Diebold, J.-H. Lee, and G. Weinbach. Regime Switching with Time-Varying Transition Probabilities. In *Nonstationary Time Series Analysis and Cointegration*, pages 283–302. Oxford University Press, 1994.

[92] Leena Kalliovirta, Mika Meitz, and Pentti Saikkonen. A Gaussian Mixture Autoregressive Model for Univariate Time Series. *Journal of Time Series Analysis*, 36:247–266, 2015.

[93] Kung sik Chan and Howell Tong. On Estimating Thresholds in Autoregressive Models. *Journal of Time Series Analysis*, 3:178–190, 1986.

[94] Marcelo Pereira, Sofia B. Ramos, and Jose G. Dias. The Cyclical Behaviour of Commodities. *The European Journal of Finance*, 23:1107–1128, 2017.

[95] Wai Mun Fong and Kim Hock See. A Markov Switching Model of the Conditional Volatility of Crude Oil Futures Prices. *Energy Economics*, 24:71–96, 2002.

[96] Ritika Jaiswal and Rashmi Uchil. An Analysis of Diversification Benefits of Commodity Futures using Markov Regime-Switching Approach. *Afro-Asian Journal of Finance and Accounting*, 8:20–47, 2018.

[97] David Easley, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman. Liquidity, Information, and Infrequently Traded Stocks. *The Journal of Finance*, 51:1405–1436, 1996.

[98] Jim Griffin, Jaideep Oberoi, and Samuel D. Oduro. Estimating the Probability of Informed Trading: A Bayesian approach. *Journal of Banking & Finance*, in press, 2021.

[99] Ahmet Sensoy, Erk Hacihasanoglu, and Duc Khuong Nguyen. Dynamic Convergence of Commodity Futures: Not All Types of Commodities are Alike. *Resources Policy*, 44:150—-160, 2015.

[100] Daniel P. Scheitrum, Colin A. Carter, and Cesar Revoredo-Giha. WTI and Brent Futures Pricing Structure. *Energy Economics*, 72:462—-469, 2018.

[101] Raymond P. H. Fishe and Aaron Smith. Do Speculators Drive Commodity Prices Away from Supply and Demand Fundamentals? *Journal of Commodity Markets*, 15:1–16, 2019.

[102] Peter Zangari. *RiskMetrics Technical Document*. J.P.Morgan/Reuters, 1996.

[103] Clifford S. Asness, Tobias J. Moskowitz, and Lasse Heje Pedersen. Value and Momentum Everywhere. *The Journal of Finance*, 68:929–985, 2013.

[104] Rob J. Hyndman and Yeasmin Khandakar. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27, 2008.

[105] Tilmann Gneiting and Adrian E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102:359–378, 2007.

[106] Fabio S. Dias. An Equity-Credit Hybrid Model for Asset Correlations. *International Journal of Financial Engineering and Risk Management*, 3:223–239, 2020.

[107] Alexander Jordan, Fabian Kruger, and Sebastian Lerch. Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software*, 90, 2019.

[108] Gianni Amisano and Raffaella Giacomini. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics*, 25:177–190, 2007.

[109] Eric Ghysels, Andrew Harvey, and Eric Renault. Stochastic Volatility. *Handbook of Statistics*, 14:119–191, 1996.

[110] Christopher S. Jones. The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets. *Journal of Econometrics*, 116:181–224, 2003.

[111] Oleg Bondarenko. Estimation of Risk-Neutral Densities Using Positive Convolution Approximation. *Journal of Econometrics*, 116:85–112, 2003.

[112] Reno Roberto. Nonparametric Estimation of Stochastic Volatility Models. *Economics Letters*, 90:390–395, 2006.

[113] Stefanos Drakos. Uncertain Volatility Derivative Model Based on the Polynomial Chaos. *Journal of Mathematical Finance*, 6:55–63, 2016.

[114] Y.-T. Lin, Y.-T. Shih, C.-S. Chien, and Q. Sheng. A Note on Stochastic Polynomial Chaos Expansions for Uncertain Volatility and Asian Option Pricing. *Applied Mathematics and Computation*, 393:1–14, 2021.

[115] Emiliano Torre, Stefano Marelli, Paul Embrechts, and Bruno Sudret. Data-driven Polynomial Chaos Expansion for Machine Learning Regression. *Journal of Computational Physics*, 388:601—623, 2019.

[116] Serge Provost and Min Jiang. Orthogonal Polynomial Density Estimates: Alternative Representation and Degree Selection. *International Journal of Computational and Mathematical Sciences*, 6:12—29, 2012.

[117] Dongbin Xiu and George Em Karniadakis. The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24:619—644, 2002.

[118] Fischer Black and Myron Scholes. The Pricing of Option and Corporate Liabilities. *Journal of Political Economy*, 81:637–659, 1973.

[119] Mikhail Chernov and Eric Ghysels. *What Data Should Be Used to Price Options?* https://ssrn.com/abstract=120028, 1998. Accessed on February 11, 2021.

[120] Yongxin Chen, Giovanni Conforti, and Tryphon T. Georgiou. Measure-valued Spline Curves: An Optimal Transport Viewpoint. *SIAM Journal on Mathematical Analysis*, 50:5947–5968, 2018.

[121] Francis Longstaff. Option Pricing and the Martingale Restriction. *Review of Financial Studies*, 8:1091–1124, 1995.

[122] Marcelo G. Cruz and Gareth W. Peters and Pavel V. Shevchenko. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. John Wiley & Sons, 2015.

[123] Spiridon Penev and Tenko Raykov. A Wiener Germ Approximation of the Noncentral Chi Square Distribution and of its Quantiles. *Computational Statistics*, 15:219–228, 2000.

[124] Josef Teichmann and Mario V. Wuthrich. Consistent Yield Curve Prediction. *ASTIN Bulletin*, 46:191—224, 2016.

[125] Feike C. Drost and Theo E. Nijman. Temporal Aggregation of GARCH Processes. *Econometrica*, 61:909–927, 1993.