Bond University

DOCTORAL THESIS

Revisiting the assumptions of statistical arbitrage: the role of static hedge ratios in the declining profitability of the phenomenon

Stephenson, Jeffrey

Award date:
2021

Revisiting the Assumptions of Statistical Arbitrage: The Role of Static Hedge Ratios in the Declining Profitability of the Phenomenon

Jeffrey John Stephenson

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy
March 2021
Bond Business School

Professor Bruce Vanstone and Doctor Tobias Hahn

This research was supported by an Australian Government Research Training Program Scholarship

Abstract

Statistical arbitrage refers to a suite of quantitative investment strategies employed chiefly by hedge funds and proprietary trading firms. A statistical extension of its pure arbitrage analogue, statistical arbitrage seeks to identify and exploit temporal mis-pricings between two or more securities whose dynamic evolution shares some common stochastic trend. The arbitrageur can draw on a number of different approaches to accomplish this, though the literature is broadly segmented by the distance, cointegration and time series perspectives.

Since the initial academic investigation of statistical arbitrage, its profitability has continued to diminish as the proportion of non-convergent opportunities increased, leading to the hypothesis that spread non-convergence is the cause of declining profitability. This thesis surveys the existing literature, with particular emphasis given to evidence of statistical arbitrage failure, before presenting an approach aimed at unifying the distance, cointegration and time series perspectives under a single explicit model.

The failure of statistical arbitrage opportunities is shown to be the direct consequence of implicit model assumptions that are inconsistent with the empirical literature. An alternative model, the TVHR model, is proposed with the objective of correcting spread non-convergence. A further extension of the model to consider statistical arbitrage profitability in the presence of conventional volatility and unconventional latent regimes is also investigated, offering a comparative analysis of the strengths and weaknesses of each methodology.

This thesis concludes that the declining profitability of statistical arbitrage is not attributable to spread non-convergence, but rather to the distance approach pair selection procedure. The cointegration approach presented in this thesis, by contrast, and the proposed TVHR model variant in particular, halt and even reverse the trend of declining profitability in recent history. This thesis also finds evidence to conclude that statistical arbitrage returns are at least partially dependent on the prevailing volatility regime, and that statistical learning models are better equipped than conventional models to capture and detect latent market regimes.

Declaration

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy. This thesis represents my own original work towards this research degree and contains no material that has previously been submitted for a degree or diploma at this University or any other institution, except where due acknowledgement is made.

Signature Date

Research Outputs

Publication	Statement of Contribution
J. Stephenson, B. Vanstone, and T. Hahn (2020). "A Unifying Model for Statistical Arbitrage: Model Assumptions and Empirical Failure". In: *Computational Economics.* Advance online publication. Reproduced with permission from Springer. ISSN: 09277099. DOI: `https://doi.org/10.1007/s10614-020-09980-6`	JS 90%, BV 5%, TH 5%
J. Stephenson, B. Vanstone, and T. Hahn (2017). "A Survey of Statistical Arbitrage: Commonalities and Limitations". Under review.	JS 90%, BV 5%, TH 5%

Copyright Declaration

This thesis makes careful note of all sections which have been previously published, along with relevant copyright information.

Copyright permission has been obtained from the publisher for the one manuscript accepted for publication from the publishers. Copies of the permissions are referenced in the preface of the relevant chapters within the thesis.

Acknowledgements

I would like to express my sincere gratitude to Bond University for giving me the opportunity to pursue this research. I would especially like to thank my supervisors, Bruce Vanstone and Tobias Hahn, for their constructive feedback, expert guidance, and friendly encouragement throughout the process.

I would also like to thank my family and friends for their ongoing support and patience. My deepest gratitude is reserved for my wife, Jemma, who stood by me through all my travails, my absences, my fits of pique and impatience.

The research presented in this thesis makes use of data supplied by the Center for Research in Security Prices (CRSP).

Contents

List of Tables

List of Figures

List of Acronyms

ADF Augmented Dickey-Fuller

AMEX American Stock Exchange

ANN Artificial Neural Network

APT Arbitrage Pricing Theory

ARCH Autoregressive-Conditional Heteroscedastic

AUROC Area Under Receiver Operating Characteristics

BMD Beta Mean Difference

CAC Cotation Assistée en Continu

CAPM Capital Asset Pricing Model

CRSP Center for Research in Security Prices

DAX Deutscher Aktienindex

DDF Dynamic Dickey-Fuller

DJIA Dow Jones Industrial Average

ECM Error Correction Model

ELM Extreme Learning Machine

EM Expectation-Maximisation

EMA Exponential Moving Average

EMH Efficient-Market Hypothesis

ETF Exchange-Traded Fund

FLS Flexible Least Squares

FTSE Financial Times Stock Exchange

FX Foreign Exchange

GARCH Generalised Autoregressive-Conditional Heteroscedastic

GBM Geometric Brownian Motion

GRR Granger and Ramanathan Regression

HMM Hidden Markov Model

LASSO Least Absolute Shrinkage and Selection Operator

MCMC Markov Chain Monte Carlo

MSE Mean Square Error

NASDAQ National Association of Securities Dealers Automated Quotations

NYSE New York Stock Exchange

OLS Ordinary Least Squares

OU Ornstein-Uhlenbeck

PCA Principal Component Analysis

RMSE Root Mean Square Error

ROC Receiver Operating Characteristics

SIC Standard Industrial Classification

SLFN Single-Layer Feedforward Network

SSD Sum of Euclidean Squared Distances

SVD Singular Value Decomposition

SVM Support Vector Machine

SVR Support Vector Regression

TEV Tracking Error Variance

TIM Time In Market

TVHR Time-Varying Hedge Ratio

VIX CBOE Volatility Index

1 Introduction

The pursuit of excess returns in the financial markets is inconsistent with asset pricing theory. Despite this, a number of academically scrutinised market anomalies have continued to persist, indicating that excess returns can be achieved under the right circumstances. In recent years, asset yields have waned as global economies have contracted. The attention given to alternative investments has increased as a consequence, and market anomalies have been thoroughly investigated for their ability to deliver consistent excess returns.

Statistical arbitrage is an umbrella term for the collection of trading and investment methodologies that attempt to exploit transient market inefficiencies. While the practice has gained academic recognition as a true capital market phenomenon, there remain a number of obstacles that inhibit both the academic exploration of the anomaly, and its practical exploitation. This thesis aims to address these problems by unifying the various modelling approaches into a single, coherent framework that reconciles the deficiencies of existing methodologies.

1.1 Motivation

Statistical arbitrage is concerned with the identification and exploitation of structural inefficiencies that exist between securities. There is currently no consensus as to how best to take advantage of arbitrage opportunities, though all methodologies share a simple objective: find securities that share some common relationship, and place transactions consistent with the reversion expected to follow a temporary divergence of prices. The identification of statistical arbitrage opportunities typically proceeds along quantitative lines, though some applications advocate security selection based on qualitative considerations, such as membership of a particular industry.

The identification of statistical arbitrage opportunities, and the subsequent exploitation of those opportunities, implicitly assumes that the common relationship shared by the securities constituting the arbitrage portfolio will endure for some time. This assumption is dependent on the specific methodology used to identify the opportunity, and can give a misleading representation of a causal association. Statistical tests are susceptible to misestimation of model parameters, and qualitative specifications may overlook important statistical considerations. Additionally, the possibility that a genuine relationship may disappear suddenly following its prolonged presence in the market makes the exploitation of statistical arbitrage especially prone to failure.

A statistical arbitrage model will fail if a reversion transaction is placed and the prices of securities constituting the arbitrage portfolio continue to diverge. Such an event could be caused by a temporary shock that normalises in the future, in which case the arbitrageur may still realise profit at the cost of a protracted holding period. Another reason for the continued divergence of security prices could be a structural break causing a significant shift in the data generating process, in which case the arbitrageur will assuredly make a loss on the transaction. Both scenarios are unfavourable, and are caused by the arbitrageur placing transactions consistent with the assumption of mean-reversion. If the arbitrageur were instead aware of the possibility of mean-aversion, the positions taken in the arbitrage portfolio could be reversed to take advantage of the continued divergence or, at the very least, capital could be conserved and the transaction dismissed during periods of uncertainty.

Another complication arising from the identification phase of statistical arbitrage is the difficulty associated with finding securities that meet the requisite characteristics. Accounting for arbitrageur-imposed constraints, such as appropriate levels of liquidity or mandated capital commitments, leaves relatively few candidate securities that can be assessed for their suitability in an arbitrage portfolio. There is also a growing body of evidence indicating a sharp decline in the number of convergent statistical arbitrage opportunities since academic research first investigated the phenomenon, though the literature does not attribute this decline to increasing market efficiency. Rather, it is thought to be caused by fundamental disruptions to the arbitrage relationship coupled with model failure under shifting data generating processes, making the exploitation of statistical arbitrage opportunities difficult under real-world conditions.

1.2 Research Questions and Findings

Chapter 2 surveys the current literature concerning statistical arbitrage. The reviewed journal articles and academic papers cover a broad subset of the available literature, with particular emphasis given to novel modelling methodologies, unconventional asset classes to which the methodologies are applied, theoretical and practical insights into the profitability of statistical arbitrage and evidence of its failure. The strengths of the different modelling approaches are surveyed and their weaknesses emphasised. Evidence is found suggesting the inability of the assumptions of statistical arbitrage to account for its empirical characteristics, most critically in regard to the prevalence of spread non-convergence.

Statistical arbitrage strategies, as they are conventionally implemented, are threatened by the prospect of diminishing returns. Their relevance and continued survival within the repertoire of the quantitative investor is dependent on a thorough accounting of the various sources of risk that statistical arbitrage is susceptible to. In acknowledging the increasing proportion of non-convergent trades and failure of statistical arbitrage opportunities, the following research questions arise:

1. Is the assumption of static arbitrage relationships responsible for the declining profitability of statistical arbitrage?
2. Are statistical arbitrage returns dependent on the prevailing volatility regime?
3. Are statistical learning models better equipped than conventional models to capture and detect latent market regimes?

While the first question follows from the construction of the stochastic spread and its mathematical reason for non-convergence, the second and third are the consequence of an emerging body of literature that establishes a relationship between market states and statistical arbitrage returns. The second question explores conventional volatility regimes modelled by standard statistical techniques, while the third extends the investigation to consider unconventional regimes identified by more advanced statistical learning techniques. The objective of this thesis is therefore to understand whether the failure of statistical arbitrage can be attributed to a failure of its modelling assumptions, and how best to improve the assumptions and modelling procedures if they are the source of failure.

Chapter 3 proposes a solution that reconciles theory and practical application in addressing the question of declining statistical arbitrage profitability. The development of an integrated modelling framework that operates in partnership with, and in the context of, revised statistical arbitrage assumptions is presented in the form of the proposed Time-Varying Hedge Ratio (TVHR) model, allowing both academics and arbitrageurs to consider statistical arbitrage more thoroughly as an extension of the existing market anomaly literature. The model specifically addresses spread non-convergence by allowing the estimate of the arbitrage relationship hedge ratio to vary throughout the trading period, ensuring spread convergence and addressing the first research question. The model is further refined by augmenting it with regime switching and statistical learning extensions, the former drawing on volatility information to draw conclusions about trade profitability, the latter considering exogenous data under a more universal mapping application. These extensions are intended to address the second and third research questions to determine whether statistical arbitrage returns are dependent on specific market states.

A standardised testing procedure quantifying the comparative statistical and economic performance of conventional and TVHR methodologies is replicated from landmark investigations of statistical arbitrage, with results presented in Chapter 4. The proposed TVHR model achieves greater levels of capital efficiency than conventional models and reduces the proportion of non-convergent trades to a negligible level. However, that greater efficiency comes at the expense of portfolio returns, with the TVHR model exploiting small inefficiencies that exceed transaction costs by a small margin. Nevertheless, the TVHR model delivers its objective of convergent statistical arbitrage opportunities, addressing the first research question. The regime switching and statistical learning model extensions improve per-trade performance of the proposed TVHR model at the expense of portfolio mean monthly returns, with the statistical learning extension delivering greater economic benefit, addressing the second and third research questions.

Concluding remarks are offered in Chapter 5. Sufficient evidence is found to conclude that the declining profitability of statistical arbitrage is at least partially attributable to the assumption of static arbitrage relationships, though assumed invariance is not the greatest driver of diminishing returns. Rather, the declining profitability of statistical arbitrage is largely attributable to the specific selection procedure of statistical arbitrage opportunities employed by the arbitrageur. Evidence is also found to conclude that the returns of statistical arbitrage opportunities identified under a specific selection procedure are dependent on the prevailing volatility regime, and that statistical learning models are better equipped than conventional models to capture and detect latent market regimes.

1.3 Research Contributions

This thesis addresses a number of issues lacking clarity in the literature, delivering several theoretical and practical contributions that will advance discussions about and investigations into the phenomenon of statistical arbitrage. Of particular note to the theoretician:

- A significant proportion of academic research has concerned the distance approach to the selection of statistical arbitrage opportunities. This thesis demonstrates that the distance approach is sub-optimal, considerably restricting the universe of candidate opportunities by construction of its optimisation objective.

- The three most ubiquitous approaches to the identification and exploitation of statistical arbitrage opportunities, namely the distance, cointegration and time series approaches, are shown to be equivalent.
- Proof of the failure of statistical arbitrage opportunities that are mistakenly assumed to exhibit a static arbitrage relationship is mathematically established and numerically demonstrated.
- A unifying model that allows the consideration of statistical arbitrage opportunities is proposed, allowing its extension to one which incorporates a time-varying hedge ratio in accordance with empirical observations of the phenomenon.

Practically, the chief contributions of this thesis include:

- Empirical confirmation of the continuing trend of declining profitability under the distance approach to the identification of statistical arbitrage opportunities.
- An alternative, theoretically-sound procedure for the identification of statistical arbitrage opportunities under the cointegration approach.
- Validation of the proposed TVHR model and confirmation of its favourable risk-adjusted performance relative to conventional static models.
- Validation of the modelling flexibility of the proposed TVHR model.
- Validation of the influence of latent market regimes on the economic performance of the proposed TVHR model.
- A procedure for attenuating the false positive rate of the primary TVHR model with a secondary classification model trained on trade outcomes.

The theoretical contributions of this thesis are primarily discussed in Chapter 3, while Chapter 4 addresses the empirical contributions. Both are discussed more thoroughly in Chapter 5, along with a number of considerations and recommendations for future research.

2 Literature Review

The theoretical justification for profitable speculation in the financial markets is a contentious issue among financial economists. Beginning with the contributions of Fama, Fisher, Jensen, and Roll (1969) and Fama (1970), the Efficient-Market Hypothesis (EMH) introduced the idea that financial and, in particular, stock markets rapidly adjust to new information. The hypothesis was refined by Fama (1991) to define an efficient market as one which fully reflects all available information. This implies that markets process information rationally, incorporating relevant information into asset valuations in an efficient manner to maintain price levels at the asset's fundamental intrinsic value.

EMH has several important implications for financial market speculation, most notably that asset prices fluctuate randomly through time in response to the unanticipated arrival of new information. This random walk behaviour implicitly suggests that the future price of an asset is unpredictable, and that traders are not able to outperform the market on a risk-adjusted return basis. Intuitively the implications are very logical—if asset returns were predictable, many traders would take advantage of the opportunity to generate unlimited profits. The production of unlimited wealth is not something that a stable economy is capable of, however, and EMH finds its strongest defence in this fact.

In the years since EMH was first proposed, there have been many efforts to prove or disprove the theory through empirical analysis of the markets. Though investigations by proponents of the theory such as Malkiel (2003) concluded that markets are indeed efficient and follow a random walk, there are many anomalies that have been shown to contradict EMH and its implications. These anomalies form the basis of counter-arguments to EMH.

One of the foundation principles of EMH is that current information cannot be used to predict future returns. Empirical evidence suggests that this is not always the case, with stock market anomalies providing the most significant body of evidence to reject this assertion. For example, Lakonishok, Shleifer, and Vishny (1994) found that portfolios constructed from value stocks—those with high earnings, cash flows or tangible assets relative to the current share price—routinely outperform the market over long horizons. Conversely, when considering short-horizon returns, Jegadeesh and Titman (1993) found that portfolios that generated high returns relative to the market in their recent history continued to outperform the market for the following three- to twelve-month period. This momentum can partially be accounted for by the slow adjustment of the market to surprise earnings announcements. The anomaly literature was also extended by Lo and MacKinlay (1988) and Lo and MacKinlay (2011), who found evidence of non-zero short-run serial correlation in asset returns.

The Foreign Exchange (FX) market also experiences mis-pricing anomalies. Under EMH assumptions, the forward exchange rate should be an unbiased predictor of the spot exchange rate at the settlement date of the forward contract. This amounts to the assumption that interest rate parity will hold for interest-bearing assets denominated in different currencies. Across a wide range of currencies and time periods, however, this assumption fails to materialise. Evidence suggests that the forward exchange rate is in fact a biased predictor of the spot exchange rate at the settlement date, with the current spot rate tending to move away from the future spot rate (Fama, 1984).

Anomalous behaviour can also be present in the pricing of essentially identical securities trading on different markets, leading to arbitrage opportunities. Arbitrage is defined as the simultaneous purchase and sale of separate securities with identical cash flows, whose price disparity allows risk-free profits to be realised (Sharpe, Alexander, and Bailey, 1999). The arbitrage mechanism serves to strengthen market efficiency by accelerating price discovery across various markets. Relatively few arbitrage opportunities present themselves in real-world conditions, however, and those that do are either temporary or face significant limitations to their practical consideration (Shleifer and Vishny, 1997). Though the number and magnitude of arbitrage opportunities is limited, their statistical extension finds significant support in the literature and constitutes the principal domain of research for this thesis.

The significance of these and other anomalies is to provide evidence that financial markets are not as efficient as EMH suggests, with returns being at least partially predictable. This conclusion is vital for the success of any forecasting strategy, as strong market efficiency would see asset returns following a true random walk, making them impossible to predict over any forecasting horizon.

2.1 Statistical Arbitrage

Statistical arbitrage is a highly quantitative trading and investment strategy that had its humble beginnings in the ubiquitous pairs trading methodology made famous by Morgan Stanley in the 1980s (Gatev, Goetzmann, and Rouwenhorst, 2006). The initial incarnation of pairs trading sought to identify and exploit market pricing inefficiencies between two related stocks. Though the complexity of the methods and models used in pairs trading has grown considerably, the essence of the approach remains unaltered; a pair of securities which are exposed to similar market forces or share some statistical relationship are identified, with positions being entered when the securities' prices diverge significantly from one another. The method seeks to capture profit by opening a short position in the overvalued security and a long position in the undervalued security expecting that the temporary divergence will reverse in the future, allowing prices to converge once more to their historical equilibrium.

Pairs trading proved to be a powerful asset for Morgan Stanley, earning the firm not only significant profits but also a reputation on Wall Street. Over time the pairs trading framework has become more sophisticated, accommodating multiple related securities over many asset classes and many timeframes. Models have been developed which allow the trading of a basket of securities against a target security or another basket of securities—an example of this might be the constituents of a stock index being traded against an Exchange-Traded Fund (ETF) of the index, or the constituents of another related index. It is clear that the pairs trading term cannot be applied to this class of models since they concern more than just a pair of securities. A more appropriate umbrella term for these methods is statistical arbitrage, since the approach prescribes the use of statistical models to mimic the nature of standard arbitrage. Statistical arbitrage is therefore defined as the simultaneous purchase and sale of separate securities with similar cash flows, whose price disparity potentially allows risky profits to be realised.

The definition of statistical arbitrage is further refined by Hogan, Jarrow, Teo, and Warachka (2004) to include any trading strategy which satisfies a number of constraints:

1. The strategy must be self-financing such that there is no initial cost to the arbitrageur.
2. The strategy must have positive expected discounted profits in the limit.
3. The probability of loss converges to zero in the limit.
4. The time-averaged variance of returns converges to zero if the probability of loss does not converge to zero in finite time.

The fourth condition of statistical arbitrage applies only when there is a positive probability of a loss, else a standard arbitrage opportunity exists. Using this definition, standard arbitrage can be seen as a special case of statistical arbitrage. The limiting behaviour of standard arbitrage satisfies the second and third conditions when strategy proceeds are invested at the risk-free rate, following the successful exploitation of an arbitrage opportunity, nullifying the fourth condition since there is no time at which the probability of loss is non-zero. This clearly differs from statistical arbitrage, whose probability of loss is always positive in finite time.

A comprehensive review of the statistical arbitrage literature is offered by Krauss (2017), in which the author delineates the most common and academically accessible approaches into five streams of literature; these are the distance, cointegration, time series, stochastic control, and alternative approaches. Each of the five approaches is thoroughly investigated, with landmark papers dissected and examined for their theoretical and practical merits. Additionally, the review identifies the extensions of pairs trading to quasi-multivariate and multivariate frameworks, in which one security is traded against a weighted basket of co-moving securities, and in which groups of securities are traded against other groups of securities, respectively. These terms of reference extend the nomenclature to bivariate (traditional pairs), quasi-multivariate and multivariate statistical arbitrage.

The following discussion reviews the key papers in each of the categories identified by Krauss (2017), excluding the stochastic control approach. This approach concerns the allocation of capital between a statistical arbitrage opportunity and other available assets. Consequently, the stochastic control approach is more concerned with optimising an investor's portfolio holdings relative to some utility function, and is beyond the scope of this thesis. In addition to the papers reviewed by Krauss (2017), an extensive search was conducted in Thomson Reuters' Web of Knowledge and Elsevier's Scopus literature databases. Search terms included *pairs trading*, *statistical arbitrage*, *mean reversion*, *relative value* and *cointegration*, with results refined to include only publications from 1998 to 2020, from sub-fields constituting computer science, mathematics, economics, econometrics, finance, engineering, physics, business, management, accounting, and decision sciences. Search results were further refined and publications chosen that placed a particular emphasis on practical trading applications, especially if the modelling approach or trading application was different or unique. The notation used in each of the reviewed papers is faithfully reproduced here.

Study	Objective	Data	Outcome
Gatev, Goetzmann, and Rouwenhorst (2006)	investigate economic utility of a simple statistical arbitrage methodology	US equities	positive excess returns after conservative transaction costs, bootstrap analysis indicates presence of dormant risk factor
Do and Faff (2010)	investigate the persistence of Gatev, Goetzmann, and Rouwenhorst (2006) statistical arbitrage returns	US equities	declining profitability attributable to non-convergent pairs
Perlin (2007)	extend Gatev, Goetzmann, and Rouwenhorst (2006) methodology to quasi-multivariate statistical arbitrage	BR equities	positive excess returns and low data-mining bias for most parameterisations
Huck (2015)	incorporate volatility regimes into Gatev, Goetzmann, and Rouwenhorst (2006) methodology	US, JP equities	no evidence to support volatility timing
Nath (2003)	investigate efficacy of Gatev, Goetzmann, and Rouwenhorst (2006) approach on high-frequency debt securities	US treasuries	superior returns and Sharpe ratio of statistical arbitrage relative to duration-matched benchmarks

Table 2.1: Collection of literature exploring the distance approach to statistical arbitrage.

2.1.1 Distance Approach

The distance approach to statistical arbitrage represents the first thorough academic investigation of pairs trading methodologies undertaken, beginning with the seminal work of Gatev, Goetzmann, and Rouwenhorst (2006). The paper, an update of an earlier working paper published and distributed in 1999, takes its inspiration from anecdotal evidence of strategies employed by hedge funds. The opaque and secretive nature of the institutional investment community obfuscated the strategies that were used at the time, making it difficult for academia to explore scientifically. The ansatz developed and employed by Gatev, Goetzmann, and Rouwenhorst (2006) was consequently the result of interviews with traders who disclosed the general theme of their pairs trading activities.

The authors split the dataset into two periods: a formation period, and a trading period. During the 12-month formation period, securities are normalised by calculating their cumulative return from the beginning of the period. After excluding securities that had one or more days with no trading activity, the Sum of Euclidean Squared Distances (SSD) is calculated for the $n(n-1)/2$ combinations of pairings given n initial securities. Only those 20 pairs that exhibited the smallest historic SSD are considered for the trading period, with their historical standard deviation recorded for future use. During the following six-month trading period, prices are once again normalised and positions entered in the pair once their prices diverge by more than two historical standard deviations, and closed once the prices converge again or the trading period ends. The strategy aims to capitalise on the mean-reverting behaviour of securities that are exposed to similar economic forces, consequently entering a long position in the undervalued security, and a short position in the overvalued security.

The method proposed by Gatev, Goetzmann, and Rouwenhorst (2006) seeks to maintain the essence of pairs trading without imposing any distributional assumptions. The approach, its separation into formation and trading periods, the applied pair selection criteria, and the choice of entry and exit thresholds constitute a very generalised application of the pairs trading paradigm: find securities that move together, and enter a long-short position when their prices diverge, exiting after convergence. All parameters of the trading strategy are chosen arbitrarily so as to minimise data-snooping bias, the robustness of which is verified in the true out-of-sample period between the paper's initial publication and its latter update.

Empirical simulation is conducted on daily prices of S&P 500 constituents over the period 1962–1998 in the original study, and 1962–2002 in the updated publication. Strategy returns across the top 20 pairs average a monthly excess return of 1.4%. Interestingly, the annualised excess return of the strategy over the out-of-sample period 1999–2002 averaged 10.4%, demonstrating the persistence of statistical arbitrage opportunities under the authors' simple method. These returns are found to have low correlation with common risk factors, such as momentum and mean-reversion effects, as well as the three ubiquitous Fama-French factors. Despite these encouraging results, the method advanced by Gatev, Goetzmann, and Rouwenhorst (2006) is sub-optimal in its selection of pairs during the formation period.

As discussed by Krauss (2017), the SSD selection metric works in opposition to the objectives of the rational investor, whose utility is maximised when investment returns are maximised. In the pairs trading setting, this requires finding pairs with strong and frequent divergence from and subsequent convergence to their long-term equilibria. The desirable characteristics of high spread variance and strong mean-reversion naturally minimise the risk of the equilibrium level changing following pair divergence, while generating a high number of trades whose return is proportional to the spread variance.

The SSD metric can be expressed in terms of spread variance and equilibrium drift, giving

$$SSD_{i,j} = \frac{1}{T}\sum_{t=1}^{T}\left(p_{i,t} - p_{j,t}\right)^2, \tag{2.1}$$

$$= \left[\frac{1}{T}\sum_{t=1}^{T}\left(p_{i,t} - p_{j,t}\right)^2 - \left(\frac{1}{T}\sum_{t=1}^{T} p_{i,t} - p_{j,t}\right)^2\right] + \left(\frac{1}{T}\sum_{t=1}^{T} p_{i,t} - p_{j,t}\right)^2, \tag{2.2}$$

$$= \mathbb{V}\left[p_{i,t} - p_{j,t}\right] + \left(\frac{1}{T}\sum_{t=1}^{T} p_{i,t} - p_{j,t}\right)^2, \tag{2.3}$$

where $p_{i,t}$ and $p_{j,t}$ are the normalised prices of the candidate securities. It is immediately clear that the ideal pair under the SSD metric would have zero spread variance since the normalised price series would be identical. This fact is evident in the results of Gatev, Goetzmann, and Rouwenhorst (2006), where the top pairs exhibit progressively lower spread variance. A more appropriate selection metric, as proposed by Krauss (2017), would be to minimise the second summand in Equation (2.3), selecting those pairs with the greatest spread variance as calculated by the first summand in Equation (2.3). This procedure would find pairs that exhibit low levels of equilibrium drift, but high profits per trade.

The sample data used by Gatev, Goetzmann, and Rouwenhorst (2006) is extended by Do and Faff (2010), who find continued profitability of the former authors' simple statistical arbitrage methodology in the period 2003–2009. The original approach is faithfully reproduced by Do and Faff (2010), with the authors excluding securities that had no trading activity or invalid return data on any given day. The implementation is then tested against the original dataset to verify its accuracy, with returns reflecting those of Gatev, Goetzmann, and Rouwenhorst (2006) to within a satisfactory margin of error.

The dataset is split into a number of periods which align with the sub-sampling of the original study. The results presented by Do and Faff (2010) focus on a variant of the strategy which enters positions on the day following an entry signal in an effort to minimise corruption of results due to bid-ask bounce. Results indicate excess monthly returns of 0.37% over the period 1989–2002, in agreement with the 0.38% monthly excess returns reported by Gatev, Goetzmann, and Rouwenhorst (2006) over the same period. However, these returns are markedly inferior to the monthly excess returns of 0.86% over the period 1962–1988. The declining profitability of the strategy continues through the out-of-sample period, 2003–2009, with monthly excess returns dropping to just 0.24%.

A number of potential reasons are offered for this declining profitability, and tests are conducted to identify its true source. The fundamental premise of statistical arbitrage is that mis-pricing among similar securities will arise in markets that are not fully efficient, leading arbitrageurs to exploit temporary inefficiencies. One of the potential reasons for declining pairs trading profitability is therefore the closure of price inefficiencies due to increased competition among arbitrageurs, in addition to more general increases in market efficiency. A more insidious potential reason for declining profitability is that of non-convergence, where the equilibrium relationship between two securities is shifted by fundamental disruptions, or fails completely under market stress.

A comprehensive examination of cross-sectional characteristics is conducted across four distinct groups of pairs: those that did not trade during the trading period; pairs that opened a position but did not converge before the end of the trading period; pairs that placed only one convergent trade but that may have subsequently opened a non-convergent position; pairs that placed multiple trades. The first and most striking finding is that non-convergent trades only accounted for 26% of all pairs over the period 1962–1988, jumping to 39% over 1989–2002, and 40% over 2003–2009. This increased proportion came at the expense of the multiple-trade group, whose proportions declined from 42% over 1962–1988, to 24% over both subsequent periods. The increased proportion of non-convergent trades was accompanied by declining profitability of trades within the non-convergent group, whereas profitability within the other groups remained relatively constant.

These findings indicate that the greatest source of declining profitability is the increased proportion of non-convergent pairs whose equilibrium relationships do not persist over the trading periods, as opposed to an increase in market efficiency. As noted by Krauss (2017), the assumption of price equilibrium based on spurious relationships is not sufficient indication of mean-reverting behaviour. In an attempt to improve pair selection, Do and Faff (2010) test a variant of the strategy that first requires candidate pairs be composed of securities in the same market sector—the rationale is that securities in the same sector are exposed to more closely-aligned market forces than those that span different sectors. The second requirement is that candidate pairs must have a relatively high number of zero-crossings in the formation period, in which the sign of the divergence between hedged securities constantly changes. This metric acts as a proxy for mean-reversion strength, and the combination of both constraints sees the selection process of Do and Faff (2010) produce positive, albeit marginal, excess returns after consideration of transaction costs.

Pairs trading is extended to a quasi-multivariate application by Perlin (2007), who offers a framework that combines the flexibility afforded by modelling one security in terms of multiple co-evolving securities, along with the simplicity of the approach of Gatev, Goetzmann, and Rouwenhorst (2006). The methodology begins with the normalisation of all candidate securities' prices by subtracting their mean and dividing by their standard deviation. What follows is the selection of m securities whose correlation with the target security is greatest. These m securities are then linearly combined to produce a forecast of the target security's fair price. The linear model is given by

$$p_t^* = \sum_{i=1}^{m} w_i p_{i,t} + \epsilon_t, \tag{2.4}$$

where p_t^* is the normalised price of the target security, $p_{i,t}$ is the ith paired security constituting the artificial security, w_i is the ith security's weighting coefficient, and ϵ_t is a standard normal residual.

Weights are assigned to each of the m securities by one of three schemes. The simplest and most intuitive is a simple uniform weighting, where each of the m securities is given a weight of $1/m$. The next weighting scheme assigns a weight to each of the securities according to an Ordinary Least Squares (OLS) estimation procedure, wherein the model residual, ϵ_t, is minimised. The author notes that this weighting scheme introduces a significant amount of multi-collinearity among the m securities, since their selection is made according to the greatest correlation with the target asset. Multi-collinearity can be controlled for by various dimension reduction methods, though no such considerations are made by Perlin (2007) in order to keep the method and analysis as simple as possible. Additionally, this OLS weighting scheme does not allow trading of the m securities unless constraints are placed on the assigned weights. The final weighting scheme is based on correlation, where the weight, $w_i = \rho_i / \sum_{i=1}^{k} \rho_i$, makes use of the correlation coefficients, ρ_i, that were used to select the m securities.

The author finds motivation for this quasi-multivariate statistical arbitrage strategy in the failure of classical pairs trading to determine whether price divergence between securities is attributable to a temporary mis-pricing that will be corrected in the near-term, or a more fundamental shift in the equilibrium pricing relationship between the securities. By averaging over multiple co-moving securities or data streams, the approach discussed by Perlin (2007) produces a more stable estimate of the target security's conditional mean value, offering greater certainty that accurate signals are generated when prices diverge by a significant degree.

The threshold level of divergence, d, determines when positions are entered in the securities. This parameter is chosen arbitrarily, though a number of possible values are tested in simulations on Brazilian equities data. Though the approach allows positions to be taken in the $m+1$ target and co-moving securities, only the target security is traded in simulations in order to illustrate the simplest variant of the strategy; it may be prohibitively complicated to simultaneously enter positions in the $m+1$ securities, or the associated transaction costs could erode returns by an unsatisfactory amount, so the m co-moving securities are excluded from trading. It is important to note at this point that the approach no longer satisfies the requirement that statistical arbitrage strategies must have zero initial cost. While this is a clear violation of the common definition of statistical arbitrage, a substantial proportion of literature in the field is devoted to strategies for which the arbitrageur would incur some initial cost. It would therefore be negligent to omit such papers on the violation of this criterion.

Evaluation of the strategy performance scrutinises the influence of data-mining bias, consequently comparing the strategy to two benchmark portfolios. The first is a simple naive portfolio formed by taking a long or short position in the target security in proportion to the overall position occupied by the strategy. If, for example, the strategy held a long position in the target security a percent of the time and a short position b percent of the time, then the benchmark would maintain a position of $(a-b)$ percent over the course of the trading period, with a negative number indicating a net short position. The other benchmark is the collection of N portfolios formed by randomly permuting the position history of the strategy, such that the trading positions are proportionally equivalent to those entered by the strategy, but occurring at different times. This bootstrapping method calls for the procedure to be repeated N times, with each permutation's performance metrics recorded for comparison with the original strategy. The final portfolio gives a distribution of random strategies whose time in the market is in proportion with the original strategy, so as to ascertain the extent of the trading edge that might be offered.

Simulations are presented for the quasi-multivariate strategy applied to the 57 most liquid equities on the Brazilian stock exchange over the period 2000 to 2006. The formation period is a moving window consisting of two years of the most recent data, with the conditional mean target security price re-estimated at 10-day intervals. Transaction costs of 10 basis points are incorporated, with the number of co-moving securities, m, set to five and the threshold divergence parameter, d, assuming values between 0.5 and 2 at intervals of 0.1. The returns of the strategy across all 57 equities are found to be profitable for all weighting schemes with the divergence threshold parameter assuming a value between 1.2 and 2. Lower threshold values lead to a greater number of trades being placed, causing transaction costs to have a greater influence on the profitability of the strategy.

The trading strategy beats the naive benchmark portfolio in terms of annualised return for every choice of the threshold parameter between 1.2 and 2, and the bootstrap portfolios at least 98% of the time. The OLS weighting scheme produces fewer trading signals than the uniform and correlation-based schemes, though the latter schemes produce greater annualised returns across the sample period. Another interesting feature of the OLS weighting scheme is that its profits are produced by a relatively even mix of long and short positions, while the uniform and correlation-based schemes produce the great majority of their profits from long positions. Nevertheless, all weighting schemes produce statistically significant excess returns that are further confirmed by superior returns to the benchmark portfolios, with annualised returns between 2.40% and 24.24% generated over the sample period.

While all of the literature so far reviewed concludes that statistical arbitrage under the distance approach produces significant excess returns, the magnitude of those returns was found to be in decline over recent years. Despite declining profitability, Do and Faff (2010) found that the distance approach performed well during major bear markets. One explanation for this is offered by Huck (2015), who suggests that the increased profitability during bear markets can be explained by a decline in market efficiency. The major bear markets, and market turmoil more generally, are accompanied by high volatility. Recognising this, a statistical arbitrage methodology is proposed by Huck (2015) that conditions on the prevailing volatility regime in an attempt to determine the economic utility of volatility timing.

A number of pair selection methods are presented in the paper, going beyond the distance approach to include tests of stationarity and cointegration. The two latter methods align more closely with the objective of statistical arbitrage to find and exploit securities that move together. Attempts were also made, however, to enhance the standard distance approach by accepting only those pairs whose SSD over the formation period differed by 10% or less, increasing the probability that the candidate pairs experience some measure of fundamental co-movement. Stationarity is identified in the methodology by considering the ratio of prices of a pair of securities, while cointegration considers the residual of a linear regression of one price series on the other. Stationarity and cointegration are satisfied, respectively, if the price ratio and regression residual pass a stationarity test. These concepts will be covered more thoroughly in Section 2.1.2.

The trading strategy proposed by Gatev, Goetzmann, and Rouwenhorst (2006) is replicated by Huck (2015), after suitable pairs are chosen by each of the three selection methods. An initial empirical investigation is mounted on equities data comprising the S&P 500 and Nikkei 225 indices over the period July 2003 to June 2013. Transaction costs of 30 basis points are considered, while a 1% p.a. short-selling cost is imposed for the overvalued constituent of each pairs trade. The results indicate that all selection methods find statistical arbitrage opportunities that generate significant excess returns before transaction costs are incorporated. The inclusion of transaction costs crucially erodes the profitability of the strategy for all but the cointegration selection method across both datasets, and the distance method on Nikkei 225 equities. The continued profitability of the distance selection method serves to illustrate the efficacy of the approach, though the author stresses the sensitivity of the approach to the choice of various parameters, including the length of the formation period and divergence threshold level.

Volatility regimes are then introduced to the approach, with the CBOE Volatility Index (VIX) serving as a proxy for market volatility. The regimes are identified by first subtracting a three-month moving average of the VIX from the current VIX level and then dividing by the moving average; decreasing, stable and increasing volatility regimes are defined when this statistic attains values below -10%, between -10% and 10%, and above 10%, respectively. An additional high volatility regime is defined when the volatility statistic is above 20%, or the current VIX level is above 30. The empirical analysis then follows from the initial investigation by applying the distance approach trading rule to each of the three pair selection methods, with the added caveat that entries are only transacted if they occurred during one of the four regimes. Each regime is tested separately in this manner.

The introduction of the volatility regime as an additional filter for trade entry globally decreases the performance of the strategy, across all selection methods and parameterisations, with and without transaction costs, and for both equity markets. An explanation offered by the author is that the majority of profits are realised shortly after the initiation of a statistical arbitrage trade. The regime filter serves to delay the initiation of a trade, effectively resigning trading to the largely unprofitable period toward the end of a trade in which the two securities gently revert to their historical equilibrium. The final conclusion of the paper is that volatility timing based on the VIX does not contribute to the profitability of statistical arbitrage opportunities, despite evidence that statistical arbitrage performs well during volatile markets. Another conclusion drawn by the author is that the arbitrageur should not delay trade entry after a temporary mis-pricing is identified.

High-frequency pairs trading in the U.S. bond market is investigated by Nath (2003). In a modification of the distance approach proposed by Gatev, Goetzmann, and Rouwenhorst (2006), candidate securities are first filtered by trading activity; only those with 10 or more trades per day are considered for the statistical arbitrage strategy. The SSD between each combination of securities comprising the filtered universe of treasury bills, notes and bonds is then calculated over a moving 40-day period. Selected pairs are then traded for the following 40-day period, with opening, closing and stop-loss triggers defined in terms of some percentile of the historical deviations observed over the formation period.

The author goes to significant lengths to model the effects of transaction costs and to properly account for strategy returns, given the great sensitivity of high-frequency strategies to various sources of market friction. Bid-ask spread, commission, shorting and financing costs are estimated and incorporated. The choice of opening trigger percentile threshold and its effect on various trade characteristics, including trade frequency, transaction cost, and proximity to stop-loss, is investigated. The choice of the percentile threshold presents a trade-off between profitability and trade frequency, the effect of which is illustrated by a number of thresholds in the simulation.

The financial data used in the empirical analysis consists of tick-level bid and ask quotes in addition to trade prices for all U.S. treasury securities over the period January 1994 to December 2000. This dataset comprises some 4.5 million trades and 50 million quotes across 829 securities. The opening threshold assumes values of either the 15^{th} or 20^{th} percentile, while the stop-loss threshold assumes values of either the 5^{th} or 10^{th} percentile. In the simulation, trades are placed symmetrically about the median 50^{th} percentile; when the spread diverges to the 15^{th} (85^{th}) percentile, for example, a long (short) position is taken on the spread which is subsequently closed when the spread converges to the 50^{th} percentile, or the spread widens and the stop-loss is triggered at the 5^{th} (95^{th}) percentile.

Benchmarking is accomplished by holding the securities comprising the pair for the duration of the trading period. The most successful parameterisation of the strategy, in which an opening threshold and stop-loss threshold of the 15^{th} and 5^{th} percentiles are used, respectively, produces daily returns of 2.05 basis points, while the benchmark returns a marginally lower 1.41 basis points per day. The Sharpe ratio of the strategy is 0.13, while the benchmark's is -0.01. Most other parameterisations of the strategy also produce superior returns to those of the benchmark, with and without transaction costs. While these returns are modest, they nevertheless indicate the applicability of the distance approach to statistical arbitrage opportunities that exist at high frequencies on the U.S. bond market.

Bowen and Hutchinson (2016) investigate the evidence of statistical arbitrage returns on U.K. equities, finding that returns are not significantly different from zero after accounting for risk and liquidity constraints. Returns on commodity futures, by contrast, are found by Bianchi, Drew, and Zhu (2009) to be positive and statistically significant, generating a market-neutral return for enforcing the law of one price. Returns of the distance approach on the Finnish stock market are explored by Broussard and Vaihekoski (2012), with results indicating that profit can be realised even after allowing a one-day lag between the generation and execution of a trading signal. Huck (2013) shows that the distance approach is sensitive to changes in the length of formation period. Jacobs (2015) considers the role of sentiment in the returns of long-short anomalies, finding that sentiment has predictive power for the short leg of strategies. Similarly, Jacobs and Weber (2015) find that the profitability of the distance approach is consistent across 34 international stock markets, but that news is a significant determinant of the strategy's time-varying returns. Profitable trading opportunities are also found by Mori and Ziobrowski (2011) to exist in the U.S. real estate investment trust market prior to 2000, though structural changes have eroded returns since. A dynamic variant of the distance approach is investigated by Stübinger and Bredthauer (2017) alongside correlation- and time series-based approaches using high-frequency data on the S&P 500, finding that the distance approach with time-varying estimates of spread volatility achieves a Sharpe ratio of 8.14 and an annualised return of 50.50% after transaction costs over a sample period encompassing January 1998 to December 2015.

2.1.2 Cointegration Approach

Study	Objective	Data	Outcome
Meucci (2009)	review the theoretical implications of cointegration	—	—
Caldeira and Moura (2013)	investigate economic utility of cointegration in statistical arbitrage	BR equities	cointegration-based statistical arbitrage outperforms benchmark index in terms of annualised return and Sharpe ratio
Montana, Triantafyllopoulos, and Tsagaris (2009)	state-space modelling of time-varying cointegrating relationship	US futures	positive annualised return and Sharpe ratio across all parameterisations
Alexander and Dimitriu (2005)	investigate index tracking under cointegrated portfolios	US equities	inability to outperform benchmark index tracking model in terms of tracking error, annualised return and Sharpe ratio; outperformance of benchmark model in statistical arbitrage trading
Burgess (2000)	development of statistical tests for static and time-varying cointegrating relationships	UK equities, EU indices	proposed statistical tests produce more profitable portfolios than conventional tests

Table 2.2: Collection of literature exploring the cointegration approach to statistical arbitrage.

Cointegration is a property of two or more time series that implies a long-term equilibrium relationship. Its use in statistical arbitrage builds on the anecdotal assertion that the arbitrageur will typically choose a number of securities that move together, seeking to find temporary price discrepancies that can be exploited for profit. While the distance approach of Gatev, Goetzmann, and Rouwenhorst (2006) introduced a simple heuristic that has stood the test of time, continuously delivering excess returns across varying markets and timeframes, it was ultimately shown by Krauss (2017) to be sub-optimal for the selection of suitable securities. Cointegration formalises the idea of co-movement between securities, consequently gaining considerable popularity in academia for its theoretical rigour.

Meucci (2009) offers a mathematical treatment of cointegration, its relation to the multivariate Ornstein-Uhlenbeck (OU) process, and its implications for statistical arbitrage. Given observations of N securities' prices, their joint evolution can be modelled by the multivariate OU process,

$$\mathrm{d}\boldsymbol{X}_t = -\boldsymbol{\Theta}\left(\boldsymbol{X}_t - \boldsymbol{\mu}\right)\mathrm{d}t + \boldsymbol{S}\,\mathrm{d}\boldsymbol{W}_t, \tag{2.5}$$

with $\boldsymbol{X}_t$ representing the vector of N security prices, whose unconditional expected value is $\boldsymbol{\mu}$. The deterministic component of the evolution is controlled by the transition matrix, $\boldsymbol{\Theta}$, while the stochastic component has dispersion matrix $\boldsymbol{S}$ that acts on the independent Brownian motions, $\boldsymbol{W}_t$. This process has a solution of the form

$$\boldsymbol{X}_{t+\tau} = \left(\boldsymbol{I} - e^{-\boldsymbol{\Theta}\tau}\right)\boldsymbol{\mu} + e^{-\boldsymbol{\Theta}\tau}\boldsymbol{X}_t + \boldsymbol{\epsilon}_{t,\tau}, \tag{2.6}$$

in which the invariants, $\boldsymbol{\epsilon}_{t,\tau}$, are mixed integrals of the Brownian motion, themselves being normally distributed $\mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\tau\right)$. The conditional distribution of the OU process is consequently normal, with $\boldsymbol{X}_{t+\tau} \sim \mathcal{N}\left(\boldsymbol{x}_{t+\tau}, \boldsymbol{\Sigma}_\tau\right)$.

The transition matrix, $\boldsymbol{\Theta}$, has eigenvalues that are either real or complex conjugate. Following a geometric interpretation of the dynamics of the OU process, the author notes that if the real part of any of the eigenvalues are either null or negative, then the covariance of the OU process for $\boldsymbol{X}_t$ does not converge. Those eigenvalues that do have a strictly positive real part have corresponding elements of the price vector whose covariance stabilises over time, with stationarity extending to any linear combination thereof. These combinations are said to be cointegrated, a property in which linear combinations of non-stationary series produce an artificial stationary series.

If it is supposed that a linear combination of the multivariate process, $\boldsymbol{X}_t$, can be found which leads to a stationary series, Y_t, then we have

$$Y_t^{\boldsymbol{w}} \equiv \boldsymbol{X}_t^{\intercal}\boldsymbol{w}, \tag{2.7}$$

where the cointegrating vector, $\boldsymbol{w}$, is normalised to have unit length. Since the cointegrated series, $Y_t^{\boldsymbol{w}}$, necessarily has variance that stabilises to some finite value, the best candidate for the cointegrating vector is the one which minimises the conditional variance, giving

$$\tilde{\boldsymbol{w}} \equiv \min_{\|\boldsymbol{w}\|=1} \operatorname{Var}\{Y_\infty^{\boldsymbol{w}} | \boldsymbol{x}_0\}, \tag{2.8}$$

given some initial realisation of the process, $\boldsymbol{x}_0$. The conditional covariance of the process, $\boldsymbol{\Sigma}_\infty$, can be decomposed into its eigenpairs, giving

$$\boldsymbol{\Sigma}_\infty \equiv \operatorname{Cov}\{\boldsymbol{X}_\infty | \boldsymbol{x}_0\} \equiv \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}, \tag{2.9}$$

where $\boldsymbol{E}$ is an orthogonal matrix whose N columns constitute the eigenvectors of the covariance, while $\boldsymbol{\Lambda}$ is the diagonal matrix of corresponding eigenvalues. This principal component factorisation of the covariance matrix partitions the space between those eigenvectors that lead to infinite variance, and those that lead to finite variance. A natural solution for Equation (2.8) is the eigenvector corresponding to the N^{th} eigenvalue, that is, $\tilde{\boldsymbol{w}} \equiv \boldsymbol{e}^{(N)}$. If the resulting process, $Y_t^{\boldsymbol{e}^{(N)}}$, is stationary, then $\boldsymbol{\lambda}^{(N)}$ is finite and represents the minimised unconditional variance of the process. Further cointegrating vectors can be found by sequentially working backward through the N eigenpairs until the corresponding eigenvalues are no longer finite.

The eigenpair decomposition of the covariance matrix offers a simple technique for determining cointegrating relationships within a vector autoregressive process. The approach suggests that it is not necessary to define a model for the process, nor to specify its parameters, rather only extract the eigenpairs of the covariance matrix and explore the N resulting eigenvectors for stationarity. An important caveat, however, is that the covariance matrix is known. In reality such quantities are rarely available for consideration, though the author suggests that the sample covariance is sufficient in its absence, allowing for the approximation of the true asymptotic covariance.

The paper concludes with a number of implications for statistical arbitrage. The first, derived from the univariate analogue of Equation (2.6) for which a stationary cointegrated series can be considered an example, establishes the divergence z-score of the cointegrated series as

$$z_{t,\infty} \equiv \frac{|Y_t - \mathrm{E}[Y_\infty]|}{\mathrm{std}[Y_\infty]} = \frac{|Y_t - \mu|}{\sqrt{\frac{\sigma^2}{2\theta}}}. \tag{2.10}$$

The second result follows from the fact that the series reverts to its unconditional mean, μ, at an exponential rate. The half-life of this mean-reversion can be computed as

$$\tilde{\tau} \propto \frac{\ln 2}{\theta}, \tag{2.11}$$

giving the expected time for the z-score to decay by half its value. These two results influence the practical considerations of the arbitrageur, with the z-score offering a normalised measure of the divergence of the spread from its mean that can be used to define trade opening signals, whose expected holding time will be some multiple of the half-life, depending on the chosen closing signals.

While no empirical analysis of the methodology is presented, the results offered by Meucci (2009) give a solid theoretical framework for statistical arbitrage strategies based on the cointegration approach. Though cointegration offers a more formalised framework for finding securities that move together, the author notes that the arbitrageur could erroneously select a number of securities whose cointegrated spread exhibits the greatest mean-reversion rate. While this naive decision might appear to lead to the most profitable trading system, the reversions are of a lesser magnitude than those of the preceding eigenpairs. Consequently, while trade frequency may be high, the corresponding return per trade will be prohibitively small or even negative after incorporating transaction costs. Another undesirable property of the framework is the tendency for in-sample cointegration to suggest opportunities that disappear out-of-sample. As the author notes, those securities that exhibit the strongest cointegration in-sample are the least robust out-of-sample.

An empirical investigation of the cointegration approach to statistical arbitrage is presented by Caldeira and Moura (2013), whose findings support the place of cointegration among the techniques of the arbitrageur. The methodology makes use of an alternative test of cointegration to that proposed by Meucci (2009), though one that is perhaps more commonly found in practical applications of the cointegration approach. The observation that non-stationary security prices can have common stochastic trends is offered as justification for the cointegration approach, with the methodology seeking to discover those securities that share the strongest common trends.

The approach begins with the assumption that the arbitrageur will invest an equal amount in long and short positions,

$$\alpha P_t^l = P_t^s, \tag{2.12}$$

where l and s denote long and short positions, respectively, and α is the ratio of prices of the short and long constituents, necessary to ensure capital is equally allocated to both sides of the trade—for every unit of P_t^s sold, α units of P_t^l are bought. This leads to the logarithmic investment equation,

$$0 = \ln(\alpha) + \ln(P_t^l) - \ln(P_t^s). \tag{2.13}$$

It is assumed that the logarithmic price series, $\ln(P_t^l)$ and $\ln(P_t^s)$, are both non-stationary. If there exists a value, γ, such that $\ln(P_t^l) - \gamma \ln(P_t^s)$ is stationary, then the pair of securities are said to be cointegrated. The resulting investment equation becomes

$$0 = \ln(\alpha) + \ln(P_t^l) - \gamma \ln(P_t^s), \tag{2.14}$$

where the values for α and γ are determined by a cointegrating regression.

Despite the popularity of cointegration testing in the statistical arbitrage literature, a number of limitations are identified by Caldeira and Moura (2013). Mis-estimation of the cointegrating coefficients, for example, may lead to spurious results and could even falsely confirm the presence of cointegration in series that are not. The authors suggest the use of two complementary methods for the identification of cointegration to assuage the possibility of erroneous inference. The most commonly used method for detecting cointegration is the Engle-Granger two-step test (Engle and Granger, 1987), in which the residuals of a cointegrating regression are tested for stationarity. The additional use of Johansen's test for cointegration (Johansen, 1988) is intended by the authors to introduce a comparative measure that can be used to verify the results of the Engle-Granger method.

Identification of statistical arbitrage opportunities is undertaken by testing each of the $N(N-1)/2$ combinations of N securities in the investment universe for the presence of cointegration. The investment universe chosen by the authors consists of the 50 most highly-weighted stocks comprising the Sao Paulo Stock Exchange Ibovespa index. Daily closing price data beginning in January 2005 and ending in October 2012 is used for the simulation, in which moving windows containing one year of data are used to determine the cointegrating pairs, with the following four months reserved for out-of-sample testing. On average, 90 pairs of the possible 1,225 combinations are found to be cointegrated in each sub-sample, with only 20 selected for the following out-of-sample test based on their in-sample Sharpe ratio. During the four month testing period, z-score is calculated by dividing the spread's deviation from equilibrium, $P_t^l - \gamma P_t^s$, by the standard deviation of the spread's divergence observed during the year-long formation period. Long positions are opened when the z-score drops below -2 and closed when the z-score reverts above -0.5. Similarly, short positions are opened when the z-score moves above 2 and closed when the z-score reverts below 0.75. Though no explanation is offered for this asymmetry in closing signals, it can be assumed that it is due to the positive return bias of stock markets.

Transaction costs of 0.5% are incorporated, which take brokerage, slippage and short-selling costs into consideration. The resulting pairs realise an impressive average annualised return of 16.39%, with a relatively low annualised volatility of 12.42% and an average Sharpe ratio of 1.34 during the out-of-sample testing periods. The cumulative profit over the period January 2005 to October 2012 is 189.29% with a maximum drawdown of 24.49%, while the returns are largely uncorrelated with those of the benchmark index; the maximum market-relative Beta over the entire sample is only 0.06. The statistical significance of the strategy's outperformance is verified through a bootstrapping procedure, which reveals that the returns over all years except for 2008 showed significant excess returns over those of the market index. As the authors note, 2008 was a particularly turbulent time for statistical arbitrageurs, with many fund managers experiencing deep and protracted drawdowns over the period.

Montana, Triantafyllopoulos, and Tsagaris (2009) offer an alternative formulation of the cointegration approach to statistical arbitrage. Instead of performing a cointegration test using a static dataset and fixed regression coefficient, the methodology advanced by Montana, Triantafyllopoulos, and Tsagaris (2009) models one series of prices in terms of a dynamically-regressed sequence of co-evolving data streams. By incorporating dynamic regression techniques, as realised by the Flexible Least Squares (FLS) modelling paradigm developed by Kalaba and Tesfatsion (1989) in which a penalised version of OLS accommodates time-varying regression coefficients, the authors offer a remarkably simplistic and tractable model that is capable of capturing temporal pricing inefficiencies while adapting to changing market dynamics.

The dynamic regression formulation models a target security, y_t, in terms of a vector of co-evolving data streams, $\boldsymbol{x}_t \in \mathbb{R}^p$, which can be comprised of related securities, exogenous variables, or both. The resulting regression equation,

$$y_t = \boldsymbol{x}_t^\intercal \boldsymbol{\beta}_t + \epsilon_t, \tag{2.15}$$

has time-varying regression coefficient vector, $\boldsymbol{\beta}_t$, which is assumed to evolve according to the process

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t. \tag{2.16}$$

Both $\boldsymbol{\eta}_t$ and ϵ_t are assumed to be independently and identically distributed zero-mean Gaussian noise processes, with covariances R and Q, respectively.

It is immediately clear that cointegration, in its strict mathematical sense, is incongruous with the dynamic linear formulation in Equation (2.15). The Engle-Granger two-step method, for example, estimates a fixed cointegrating vector, $\boldsymbol{\beta}$, before performing a stationarity test on the model residuals. The two-step method therefore assumes that the cointegrating relationship is constant through time. Performing a stationarity test on the residual, ϵ_t, of Equation (2.15) would give spurious results, since the dynamic regression coefficients would always force the model residual to be stationary. The assumptions of a standard cointegration test are therefore inconsistent with the dynamic regression specified in Equation (2.15), complicating the task of identifying securities that might offer statistical arbitrage opportunities.

The estimate of the target security, $\hat{y}_t = \boldsymbol{x}_t^\intercal \hat{\boldsymbol{\beta}}_t$, is interpreted as an artificial security whose price represents the true fair value of the target. Significant deviations from the fair value represent arbitrage opportunities, so it is important to accurately and dynamically model the target security with information available at each epoch. To circumvent the complication of finding securities or data streams that share some long-term statistical relationship, as would exist between cointegrated securities, the authors propose something akin to index arbitrage that ensures the target and co-evolving securities share a more fundamental, structural relationship. In the paper, S&P 500 index futures are chosen as the target security, while the constituent stocks of the S&P 500 index comprise the explanatory variables from which the true value of the index future contract is estimated.

The trading strategy used by Montana, Triantafyllopoulos, and Tsagaris (2009) enters a position only in the target security, y_t, in opposition to the sign of the deviation, $\epsilon_t = y_t - \hat{y}_t = y_t - \hat{\boldsymbol{x}}_t^\intercal \hat{\boldsymbol{\beta}}_t$. A negative value of ϵ_t leads to a long position in the S&P 500 index futures contract, while a positive value of ϵ_t leads to a short position in the futures contract. This differs somewhat from the standard statistical arbitrage framework which prescribes taking a position in both the target and co-moving securities, and only when the deviation between them reaches a significant level. Though it is not explicitly stated why only the target security is traded, it can be assumed that the prohibitive cost, both in terms of transaction cost and slippage due to difficulties with trade timing, of placing transactions for 501 securities simultaneously would dissuade the arbitrageur from such a strategy. Trading only the target security simplifies the process significantly, though it is not clear why trades are placed at every epoch rather than only when the deviation reaches a significant level.

Price data for S&P 500 index futures and constituent stocks cover the period January 1997 to October 2005 in the empirical simulation. The authors select a subset of the index consisting of 432 stocks whose continuous inclusion in the index across the sample period bypasses the task of updating the dataset to include the latest revisions to the index. Two strategies are tested, specifically one in which all constituent stocks are included in the vector of explanatory variables, and one in which the constituent stocks are dimensionally reduced to their three greatest principal components. The dimension reduction makes use of an incremental Singular Value Decomposition (SVD) technique, extracting the principal components in an online manner in keeping with the incremental nature of the FLS dynamic regression estimation procedure. The results indicate that the dimensionally-reduced explanatory variables consistently offer more profitable arbitrage opportunities, with a maximum Sharpe ratio of 0.80 and average annualised return of 13.18% realised over the sample period. Conversely, the greatest performance achieved by the full set of explanatory stocks delivered a Sharpe ratio of 0.41 and annualised return of 6.73%. Only one parameterisation of FLS gave a negative annualised return of −0.95% under the full set of variables, while the corresponding FLS parameterisation with only the principal components considered gave an annualised return of 10.82%.

Index tracking performance under the influence of cointegrated portfolios is addressed by Alexander and Dimitriu (2005), who identify the attractive property of spread stationarity as a potential source of outperformance over existing indexing methods. The objective of index tracking is the minimisation of tracking errors between some benchmark index, and a subset portfolio whose returns mimic those of the full index. Faithful reproduction of the index returns additionally requires minimal rebalancing, given the erosion of profits due to transaction costs. The existence of a long-term cointegrated relationship between securities, and the consequent stationarity of price spreads, naturally implies that rebalancing is unnecessary given the temporary nature of deviations.

The objective of any index tracking strategy is the minimisation of the Tracking Error Variance (TEV), subject to constraints that might disallow short-selling or impose a particular structure on the portfolio holdings. Construction of an index tracking portfolio begins with security selection and ends with portfolio optimisation. Existing TEV minimisation methods consider the correlation between security returns for the purposes of portfolio optimisation, which naturally introduces instability owing to the transience of the correlation statistic. The presence of cointegration, on the other hand, has a greater likelihood of persisting than relationships inferred from simple correlation. It is worth noting that the selection phase of portfolio construction can have a dramatic effect on the performance of the portfolio, with selection aided by proprietary models or simply the skill of the fund manager. No investigation of selection methods, which might otherwise obscure the performance of the cointegration-based index tracking approach, is conducted by Alexander and Dimitriu (2005).

The empirical investigation of the index tracking approach is conducted on the 30 stocks comprising the Dow Jones Industrial Average (DJIA) index over the period January 1990 to December 2003. The first simulation compares the performance between the proposed cointegration method of portfolio construction with the TEV method, across portfolios comprising 20, 25 and 30 stocks chosen according to greatest price at the time of portfolio construction. The portfolio of 30 tracking stocks is trivially cointegrated with the index, while portfolios that include fewer stocks naturally complicate the optimisation objective and may fail to produce a cointegrating relationship. For both the cointegration and TEV method of optimisation, up to five years of daily data is used for model calibration, and portfolio holdings are rebalanced either every two weeks, monthly, three-monthly or six-monthly.

Analysis of the empirical simulation indicates that the portfolios including 20 and 25 stocks consistently underperform the benchmark, while the portfolios consisting of 30 stocks outperform the benchmark. The authors found this result held true for both optimisation methods. The source of relative underperformance of the 20- and 25-stock portfolios is shown to have come from the selection method, while the outperformance of the 30-stock portfolio is attributable to both the cointegration and TEV optimisation methods. The cointegration-optimised portfolios produce slightly lower Sharpe ratios, higher transaction costs and greater tracking error than the TEV method, though with a tracking error distribution more closely resembling normality with lower kurtosis than

TEV. There is also a marginally lower probability of the cointegration approach underperforming the benchmark.

A second empirical investigation incorporates concepts familiar to statistical arbitrageurs. Synthetic indices can be constructed by adding or subtracting an annual premium to or from the index, uniformly distributed across the returns realised throughout a given year, which can then be shadowed by a tracking portfolio. Though it complicates the optimisation task since there may not exist an allocation that reproduces the returns of the synthetic index, small values of the annual premium allow the construction of reasonable tracking portfolios. Extension of the index tracking framework allows for a statistical arbitrage portfolio, in which a long position is taken in the benchmark-plus-premium index tracking portfolio, and a short position is taken in the benchmark-minus-premium tracking portfolio. Unsurprisingly, given the ubiquitous influence of cointegration in statistical arbitrage implementations, the cointegration-based statistical arbitrage portfolio outperforms the TEV portfolio across the sample period, with both portfolios using the full 30 constituent stocks to track the synthetic indices. The indices are constructed with premiums of 5%, 10% and 15%, though the best returns are observed with synthetic indices constructed from the smallest premium of 5%.

A comprehensive treatment of cointegration and its application to statistical arbitrage is offered by Burgess (2000). The author motivates the work by first noting that any random variable, which is likely the sum of deterministic and stochastic processes, has an upper limit on its predictability that is governed by the proportion of variance that can be explained by the deterministic component. It is generally accepted that the market is efficient at discounting market-wide macroeconomic risk factors, while the idea of statistical arbitrage relies on the inability of the market to fully discount the deterministic components of idiosyncratic risk. By constructing synthetic securities through cointegrating relationships, stochastic components that are common to many securities are immunised, allowing the deterministic elements to be observed more directly and modelled appropriately.

Chief among the contributions made by Burgess (2000) is a suite of diagnostic tests designed specifically to identify potential statistical arbitrage opportunities. Moving beyond standard tests of stationarity in the residual of a cointegrating regression, the author introduces the variance ratio function as a test for mean-reversion among elements of a process. The variance ratio relates average long-term variance to single-period variance, and is defined as

$$VR(\tau) = \frac{\frac{1}{\tau}\sum_t \left(\Delta^\tau y_t - \overline{\Delta^\tau y}\right)^2}{\sum_t \left(\Delta y_t - \overline{\Delta y}\right)^2}, \tag{2.17}$$

for some series, y_t, and some lag, τ. A random walk will typically have a variance ratio close to one. The author notes that suitably expressed variance ratio statistics are equivalent to weighted sums of autocorrelation coefficients, so it is unsurprising that positive and negative autocorrelation of a series over the time scale, τ, is indicated when the variance ratio is greater than or less than one, respectively. Because statistical arbitrage requires a mean-reverting spread between a target and some co-moving real or artificial security, the arbitrageur would typically select arbitrage portfolios whose variance ratio is less than one to some degree of statistical significance. The alternatives of random walk and mean-averting behaviour do not offer arbitrage opportunities under the prevailing framework.

Extension of the variance ratio statistic to a profile of the measure across different values of the lag, τ, allows a more thorough investigation of the dynamics of the series under consideration. The resulting vector of variance ratio profile, $\boldsymbol{VP}(\tau)$, is further used in the statistic $VP_{dist}(\tau)$, which measures the Mahalanobis distance of the observed variance profile from its centroid, and is known to converge to a chi-squared distribution with τ degrees of freedom. The final statistic proposed by Burgess (2000) considers the projection of the $\boldsymbol{VP}(\tau)$ vector onto the dominant eigenvectors of the covariance matrix, $\Sigma_{\boldsymbol{VP}(\tau)}$.

Extensive Monte Carlo simulation establishes the power of tests based on the proposed statistical measures over more conventional tests. In particular, the two variance profile measures are able to identify mean-reverting behaviour in series that are corrupted by non-stationary noise and short-term mean-averting dynamics, while traditional measures including Augmented Dickey-Fuller (ADF) and Box-Ljung Q-statistics fail to capture the mean-reversion that might offer profitable statistical arbitrage opportunities. The result is a testing framework that is capable of identifying profitable statistical arbitrage opportunities that are more likely to be effective in live trading circumstances than their traditional counterparts.

The author notes that the cointegration approach typically does not account for instability in the equilibrium pricing relationship between a set of securities comprising a statistical arbitrage portfolio. The major sources of instability are mis-estimation of the cointegrating vector, and potentially an equilibrium relationship whose true expression is not constant through time. The implication is that in-sample cointegration cannot guarantee future cointegration, and so the author proposes an adaptive framework incorporating a time-varying cointegrating regression, such as that defined by Montana, Triantafyllopoulos, and Tsagaris (2009). The ability of the resulting regression to track statistical mis-pricing dynamics is quantified by a modified version of the Dickey-Fuller test, given by

$$\Delta^* \epsilon_t = \epsilon^*_{t+1} - \epsilon_t = \left[y_{t+1} - \left(\boldsymbol{x}^\intercal_{t+1} \hat{\boldsymbol{\beta}}_t \right) \right] - \left[y_t - \left(\boldsymbol{x}^\intercal_t \hat{\boldsymbol{\beta}}_t \right) \right], \tag{2.18}$$

where ϵ_t is the residual of a cointegrating equation, such as Equation (2.15). In this Dynamic Dickey-Fuller (DDF) test, the regression coefficients, $\boldsymbol{\beta}_t$, are held constant over the period for which the difference operator is invoked, which is necessary to identify true mean-reverting behaviour since the time-varying regression parameters would otherwise absorb some of the non-stationarity of the price spread.

Securities are selected for inclusion in the statistical arbitrage portfolio by a test which cycles through all possible securities in the selected universe, finding those that contribute to the greatest reduction in residual variance. The first selected security is therefore the one which is most highly correlated with the target security, while sequential securities are those that are most correlated with the price spread expressed as the residual of the regression formed under the previous step. The procedure continues sequentially to add securities to the portfolio until a maximum number of securities has been added, or there is no further statistical benefit to adding securities. This selection method allows for the fully algorithmic construction of statistical arbitrage portfolios, without the need to identify common economic relationships between securities.

A simple trading rule is specified, in which a position is taken in opposition to the sign and in proportion to the magnitude of the current statistical mis-pricing. Assuming the statistical arbitrage portfolio contains one target security and N counter-securities, then the portfolio consists of the set $\{Y_t, X_{1,t}, X_{2,t}, X_{3,t}, \ldots, X_{N,t}\}$ and the corresponding portfolio holdings at any given time are expressed by $\phi_t\{1, -\beta_1, -\beta_2, -\beta_3, \ldots, -\beta_N\}$. The coefficient, ϕ_t, accounts for the sign and magnitude of the mis-pricing, and is calculated by

$$\phi_t = -\operatorname{sgn}(\epsilon_{t-1})|\epsilon_{t-1}|^k, \tag{2.19}$$

given some sensitivity parameter, $k \in \mathbb{R}^+$, which effectively inflates or shrinks the portfolio holdings in accordance with the risk appetite of the arbitrageur. An alternative formulation for ϕ_t smooths the signals over a moving average window to reduce frenetic portfolio reshuffling and limit transaction costs.

Empirical analysis of the methodology proposed by Burgess (2000) is conducted on stocks comprising London's Financial Times Stock Exchange (FTSE) 100 stock index, over the period June 1996 to October 1998. The sample spans 600 daily closing prices for all 100 stocks, with the first 400 data points reserved for the formation period, and the final 200 data points reserved for the testing period. Out-of-sample results indicate that the methodology is capable of selecting profitable statistical arbitrage portfolios, with some 450 separate portfolios evaluated with Sharpe ratios ranging from −0.93 to 3.69, with average ratios of 0.84 to 1.36, depending on the metric chosen to optimise. The time-varying cointegration methodology is tested separately on a pair comprising the French Cotation Assistée en Continu (CAC) and German Deutscher Aktienindex (DAX) stock indices, over the period August 1988 to August 1996, and found to produce a Sharpe ratio of 1.30 and a final return of 47.20%. The analysis goes on to conclude that in-sample Sharpe ratio has the greatest correlation out of all metrics with out-of-sample profitability, outperforming other performance metrics and statistical tests of mean-reversion. Though the results are without transaction costs, they indicate a powerful methodology for the construction of profitable statistical arbitrage portfolios.

Statistical arbitrage portfolio allocation is investigated by D'Aspremont (2011), with cointegration addressed by a sparse canonical correlation analysis. Dunis and Ho (2005) consider cointegration in the formation of equities portfolios for both index tracking and statistical arbitrage applications, finding that cointegration achieves good tracking performance but constructs volatile statistical arbitrage portfolios. Dunis, Laws, and Evans (2006) consider cointegration and Artificial Neural Network (ANN) models of the fair value of oil futures spreads, concluding that a simple moving average model is best able to exploit temporal inefficiencies. Identification of cointegrated series among natural gas and electricity futures is addressed by Emery and Liu (2002), with the authors finding that cointegrated series are statistically and economically significant both in-sample and out-of-sample.

Girma and Paulson (1999) investigate cointegration among crude oil, unleaded petroleum and heating oil futures, finding that historical spreads between cointegrated futures offer statistically significant opportunities for profitable trading. Huck and Afawubo (2015) find that cointegration analysis leads to more profitable statistical arbitrage opportunities than other methodologies. Cointegration among dual-listed securities is investigated by Li, Chui, and Li (2014), finding that divergent prices offer statistical arbitrage opportunities capable of generating significant annualised excess returns of 17.60%. Cointegration models are used by Liu and Chou (2003) to define the mean-reverting characteristics of precious metal price spreads that offer trading opportunities. Peters, Kannan, Lasscock, Mellen, and Godsill (2011) develop a cointegration framework for estimating and correcting prices of securities that suffer from significant overnight gaps.

A pair selection procedure inspired by the cointegration approach is investigated by Ramos-Requena, Trinidad-Segovia, and Sánchez-Granero (2017), in which the Hurst exponent of spread portfolios is used to select those that are most strongly mean-reverting. The approach achieves greater returns than distance variants on DJIA stocks over the sample period January 2000 through December 2015. Chen, Chen, Chen, and Li (2019) apply the cointegration approach to a quasi-multivariate spread portfolio formed between

CRSP stocks and their 50 nearest substitutes identified in terms of return correlation. The spread portfolio achieves significant excess returns that cannot be explained by common risk factors, though evidence is found for the returns being driven by short-term reversion and momentum factors. The strict adherence to a static cointegrating relationship is relaxed by Clegg and Krauss (2018) who propose a partial cointegration procedure that incorporates a mean-reverting component and a random walk component. Across a sample period covering 1990 through 2015 the partial cointegration procedure delivers annualised returns on S&P 500 pairs of 12% after transaction costs.

2.1.3 Time Series Approach

Study	Objective	Data	Outcome
Elliott, Van Der Hoek, and Malcolm (2005)	state-space modelling of the price spread	simulated	model estimates rapidly converge
Do, Faff, and Hamza (2006)	state-space modelling of the return spread	AU, US, UK equities	model estimates of decay rate confirm mean-reversion behaviour
Bertram (2010)	analytic derivation of optimal entry and exit thresholds	AU, NZ equities	dual-listed securities achieve Sharpe ratios between one and nine, relative to different risk-free rates
Bee and Gatti (2015)	regime switching modelling of the price spread	US futures	two-regime model outperforms one-regime model in terms of annualised return, Sharpe ratio and maximum drawdown
Chen, Chen, and Chen (2014)	regime switching modelling of the return spread	US equities	proposed model outperforms component securities in terms of return

Table 2.3: Collection of literature exploring the time series approach to statistical arbitrage.

The time series approach to statistical arbitrage differs from the distance and cointegration approaches in assuming that a number of candidate securities have already been identified by prior analysis, and that the arbitrageur is now interested in exploiting future statistical mis-pricings. Little to no effort is typically made to enumerate the possible variations of analysis capable of identifying co-moving securities. Instead, the focus of the time series approach is on the development of profitable trading signals based on models and methods arising from time series analysis.

The work of Elliott, Van Der Hoek, and Malcolm (2005) is among the most highly cited in the time series domain. The authors first assume that there is a mean-reverting hidden state process, $\{x_k | k = 0, 1, 2, \ldots\}$, for which the difference between two securities held long and short at a predetermined ratio are noisy observations. The hidden state evolves according to the difference equation,

$$x_{k+1} - x_k = (a - bx_k)\,\tau + \sigma\sqrt{\tau}\epsilon_{k+1}, \tag{2.20}$$

in which the process reverts to its long-term mean, $\mu = a/b$, at rate b. This difference equation is the discrete-time analogue of the ubiquitous OU equation, used to model stationary Gaussian and Markovian processes, and described by the stochastic differential equation,

$$\mathrm{d}X(t) = (a - bX(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t).$$

Here, as in Equation (2.20), σ specifies the variance of the noise process, while $\{\epsilon_k\}$ are realisations of an independently and identically distributed standard normal random variable, and $\{W(t)|t \geq 0\}$ is a standard Brownian motion.

The difference equation specified in Equation (2.20) can be reparameterised in terms of an AR(1) model,

$$x_{k+1} = A + Bx_k + C\epsilon_{k+1}, \tag{2.21}$$

with $A = a\tau \geq 0$, $0 < B = 1 - b\tau < 1$ and $C = \sigma\sqrt{\tau}$. While the hidden state, $\{x_k\}$, is not directly observable, measurements are obtained from an observation process,

$$y_k = x_k + D\omega_k, \tag{2.22}$$

in which $\{\omega_k\}$ are realisations of a standard normal random variable, independent of $\{\epsilon_k\}$, and with $D > 0$. The observation process in Equation (2.22), together with the state process described in Equation (2.21), jointly constitute a linear state-space model in $\{x_k\}$, with parameters (A, B, C, D).

The objective of the state-space model proposed by Elliott, Van Der Hoek, and Malcolm (2005) is to determine the best estimates of the hidden state process, $\{x_k\}$, given measurements from the observed process, $\{y_k\}$. Each estimate, $\hat{x}_{k|k}$, is computed as the expectation of the hidden state conditional on the information set, $\hat{x}_{k|k} = \mathbb{E}\left[x_k|\mathcal{Y}_k\right]$, with $\mathcal{Y}_k = \sigma\{y_0, y_1, \ldots, y_k\}$. The procedure requires the estimation of the model parameters, (A, B, C, D), in addition to the sequential estimation of the hidden state. In order to accomplish this, the authors propose the use of the Expectation-Maximisation (EM) algorithm.

A trading strategy follows from knowledge of the mean-reverting process. Once the model parameters are estimated, the divergence between the observed and hidden price spreads can be calculated and used as a trigger for entry signals. If $y_k > \hat{x}_{k|k-1}$, then the spread is regarded as too large. Counter-intuitively, the authors prescribe entering a long position in the observed spread, though the strategy implicitly assumes that the divergence will narrow and the observed spread will return to the long-term equilibrium level estimated by the hidden spread. Similarly, they prescribe entering a short position when $y_k < \hat{x}_{k|k-1}$, unwinding both long and short trades after a set length of time or once the spread corrects sufficiently.

It is also unclear why the authors do not use the full information set available at time k in the estimation of the hidden state, instead using the information set $\mathcal{Y}_{k-1}$. This arrangement is typically reserved for prediction and forecasting applications, where the full set of information is not currently available. However, since the trading rule is only concerned with the level of the observed spread relative to the hidden spread, requiring no predictions or forecasts, there is no reason why $\hat{x}_{k|k}$ could not be computed.

An alternative trading strategy advanced by Elliott, Van Der Hoek, and Malcolm (2005) sheds no further light on this oversight, but does make use of results from first passage times of the standardised OU process to effectively bypass the issue. The first passage time, T, of a stationary OU process is calculated by $T = \tau\hat{t}/(1-B)$, with discrete-time difference parameter τ, and continuous-time mean-reversion horizon, $\hat{t}$, given by

$$\hat{t} = \frac{1}{2}\ln\left[1 + \frac{1}{2}\left(\sqrt{(c^2-3)^2 + 4c^2} + c^2 - 3\right)\right]. \tag{2.23}$$

The mean-reversion horizon, $\hat{t}$, gives the time for the process to revert from some threshold level, c, back to its mean of zero. The threshold parameter is chosen by the trader, though the authors offer no guidance on appropriate values it can assume. Once all other model parameters are estimated, trades can be entered whenever

$$|y_k| \geq \left| A/(1-B) + c\left(\sigma/\sqrt{2(1-B)/\tau}\right)\right|, \tag{2.24}$$

and unwound at time T.

Numerical simulations conducted by Elliott, Van Der Hoek, and Malcolm (2005) concern the convergence rate of the EM algorithm estimation procedure. The authors find that the EM algorithm provides consistent and robust estimation of the model parameters, (A, B, C, D). Despite this finding, no empirical simulations are presented on financial data, and no indication is given of the profitability of the strategy under live trading conditions. Despite this omission, the time series method advanced by Elliott, Van Der Hoek, and Malcolm (2005) offers a significant contribution to the statistical arbitrage literature. The mathematical rigour of the mean-reverting OU process lends substantial credibility to the procedure, and its tractability ensures that it can be replicated and used in practical applications.

The stochastic spread method proposed by Elliott, Van Der Hoek, and Malcolm (2005) is extended by Do, Faff, and Hamza (2006), who made a number of additions and revisions to the state-space model that describes the dynamics of a mean-reverting price spread. Most significantly, the authors establish a link between asset pricing theory and statistical arbitrage phenomena by allowing Arbitrage Pricing Theory (APT) risk factors to influence the measurement equation in Equation (2.22). The new measurement equation,

$$y_k = x_k + \Gamma r_k^m + D\omega_k, \tag{2.25}$$

introduces the exogenous vector, $r^m \in \mathbb{R}^n$, of market risk factors, while Γ is a vector of exposure differentials between the various component securities of the spread. Together with the state transition equation, Equation (2.21), Equation (2.25) presents a linear state-space model in the spread, x_k, with parameters (A, B, C, D, Γ).

The introduction of APT factors in the model of Do, Faff, and Hamza (2006) reconciles financial theory with empirical market phenomena, though APT was chosen for inclusion in the state-space model not because of its theoretical assertions, but rather because of its flexible factor model structure. This flexibility allows the nesting of other asset pricing models and theorems, and the authors take advantage of this flexibility to incorporate market risk premia derived from the Capital Asset Pricing Model (CAPM).

Another departure from the model of Elliott, Van Der Hoek, and Malcolm (2005) is in the use of log-prices rather than level prices to define the spread between two securities, thereby modelling their relative return. The authors claim that this is necessary to ensure the mean of the spread does not change when the two securities produce the same return. However, as noted by Krauss (2017) in the author's extensive review, it is a matter of perspective and preference as to whether the price spread or the return spread is used. Under the return spread perspective, identical returns produced by the component securities will indeed leave the return spread unchanged, though their price spread will change. Conversely, under the price spread perspective, price changes in the underlying securities by an identical number of ticks will leave the price spread unchanged, but will shift the return spread.

Following estimation of the model parameters, (A, B, C, D, Γ), again by the EM algorithm, Do, Faff, and Hamza (2006) propose a trading strategy based on the accumulated residual spread. A window parameter, l, is chosen by the trader over which the sum of residuals is calculated,

$$\delta_k = \sum_{i=k-l}^{k} \mathbb{E}\left[x_i | \mathcal{Y}_i\right],$$

where δ_k is the accumulated spread at time k over the window l. It is suggested that a position might be taken once δ_k surpasses some threshold, though no guidance is given on how this threshold should be determined. It is also mentioned that the expected convergence time could be computed, implicitly suggesting that this will define the trade exit.

Simulation follows similar lines as that conducted by Elliott, Van Der Hoek, and Malcolm (2005), though the presence of real financial data allows practical consideration of the methodology. Three stock pairs across three international exchanges are chosen based on fundamental considerations: BHP and Rio Tinto, the world's largest mining companies; Target and Walmart, top retailers in the U.S.; Shell and BP, the largest energy companies in the U.K. The ASX/S&P 200, S&P 500, and FTSE All Shares indices are selected to represent the markets from which each stock pairing is selected, allowing market risk premia to be calculated for the model.

The simulation focuses on the convergence of model parameters as computed by the EM algorithm, though inference is also carried out on the estimated model parameters specifically concerning the efficacy of the trading strategy. The estimated mean-reversion rate, defined by b in Equation (2.20), is significant and strong across all three pairs, indicating that the model can capture mean-reversion in the return spread effectively and that there is minimal risk of non-convergence of the spread. However, the authors rightly infer that the speed of convergence might be too great, making it difficult to exploit for trading purposes.

The optimal trading strategy of a mean-reverting price series is addressed by Bertram (2010). Assuming that the price of an artificial security follows an OU process, analytic solutions can be found for the optimal trade entry and exit thresholds which maximise the expected return of the strategy. The author also establishes a numerical solution for the trade parameters which optimise the Sharpe ratio of the strategy, and additionally derives the expected trade length and variance of holding periods, noting that the speed of computation is vitally important for the execution of high-frequency statistical arbitrage strategies.

The price of a mean-reverting security, $p(t) = e^{X(t)}$, is assumed to be driven by an OU process in $X(t)$ which satisfies

$$\mathrm{d}X(t) = -\alpha X(t)\,\mathrm{d}t + \eta\,\mathrm{d}W(t), \tag{2.26}$$

where $\alpha\,, \eta > 0$, and $W(t)$ is a standard Brownian motion. A trading strategy designed to exploit the mean-reverting behaviour of the series prescribes entering a long position when $X(t) = a$, exiting once the series reverts to $X(t) = m$ for $a < m$. It is worth noting at this point that the analysis conducted by Bertram (2010) assumes only long positions are taken, ignoring the equally-valid short positions that could be taken when the series is overvalued.

The length of the trade cycle, $\mathcal{T}$, is decomposed into its constituent sub-intervals, $\mathcal{T}_1$ and $\mathcal{T}_2$. The time taken, $\mathcal{T}_1$, for $X(t)$ to move from a to m, and the time taken, $\mathcal{T}_2$, for $X(t)$ to return to a from m, are first passage times for the process, $X(t)$. Since the process represents the log-price of the series, the continuously-compounded return per trade is deterministic, and can be calculated by the function $r(a, m, c) = (m - a - c)$ given some transaction cost, c, while its variance can be calculated by $r(a, m, c)^2 = (m - a - c)^2$. The expected return per unit time, and variance of returns per unit time, follow from renewal theory and give

$$\mu(a, m, c) = \frac{r(a, m, c)}{\mathbb{E}[\mathcal{T}]}, \tag{2.27}$$

$$\sigma^2(a, m, c) = \frac{r(a, m, c)^2 \, \mathbb{V}[\mathcal{T}]}{\mathbb{E}[\mathcal{T}]^3}, \tag{2.28}$$

where $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$. The stochastic elements of the mean and variance of trade returns per unit time can be calculated using results from the analysis of first passage times of OU processes.

The expected trade cycle length, $\mathbb{E}[\mathcal{T}]$, is computed following the transformation by Itô's lemma of the original process, $X(t)$, to a dimensionless system, giving

$$\mathbb{E}[\mathcal{T}] = \frac{\pi}{\alpha}\left(\operatorname{erfi}\left(m\sqrt{\frac{\alpha}{\eta}}\right) - \operatorname{erfi}\left(a\sqrt{\frac{\alpha}{\eta}}\right)\right), \tag{2.29}$$

where $\operatorname{erfi}(x) = -i\operatorname{erf}(ix)$ is the imaginary error function. Similarly, the closed-form solution for the variance of trade cycle length is given by

$$\mathbb{V}[\mathcal{T}] = \frac{1}{\alpha^2}\left(w_1\left(\frac{m\sqrt{2\alpha}}{\eta}\right) - w_1\left(\frac{a\sqrt{2\alpha}}{\eta}\right) - w_2\left(\frac{m\sqrt{2\alpha}}{\eta}\right) + w_2\left(\frac{a\sqrt{2\alpha}}{\eta}\right)\right), \tag{2.30}$$

where

$$w_1(x) = \left(\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k!}\Gamma\left(\frac{k}{2}\right)\left(\sqrt{2}x\right)^k\right)^2 - \left(\frac{1}{2}\sum_{k=1}^{\inf}\frac{1}{k!}(-1)^k\,\Gamma\left(\frac{k}{2}\right)\left(\sqrt{2}x\right)^k\right)^2,$$

$$w_2(x) = \sum_{k=1}^{\infty}\frac{1}{(2k-1)!}\Gamma\left(\frac{(2k-1)}{2}\right)\Psi\left(\frac{(2k-1)}{2}\right)\left(\sqrt{2}x\right)^{(2k-1)}.$$

Substituting Equation (2.29) and Equation (2.30) into Equation (2.27) and Equation (2.28) gives the expected return and variance of returns per unit time in terms of the parameters of the process, $X(t)$, as well as the trade paramaters, (a, m, c).

The first interesting result reported by Bertram (2010) is that the optimal exit threshold is found to be $m = -a$, suggesting that the entry and exit threshold bands should be symmetric about zero. This result is in opposition to the conventional trading paradigm which proposes the use of asymmetric bands, entering a position once the divergence has reached a significant threshold level and exiting once the spread reverts to its mean. Using this result, the optimal value of a which maximises the expected return per unit time can be approximated by a third-order Taylor series expansion of Equation (2.27) around $a = 0$, giving

$$\begin{aligned} a = -\frac{1}{4}\Bigg[c + c^2\alpha\left(c^3\alpha^3 + 24c\alpha^2\eta^2 - 4\sqrt{3c^4\alpha^5\eta^2 + 36c^2\alpha^4\eta^4}\right)^{-\frac{1}{3}} \\ + \alpha^{-1}\left(c^3\alpha^3 + 24c\alpha^2\eta^2 - 4\sqrt{3c^4\alpha^5\eta^2 + 36c^2\alpha^4\eta^4}\right)^{\frac{1}{3}}\Bigg]. \end{aligned} \tag{2.31}$$

This value for the entry threshold parameter, a, maximises the return of the strategy without taking its variance into consideration. The Sharpe ratio can alternatively be maximised, though solving for a is less tractable than in the case of maximising return—a numerical optimisation routine is recommended by the author.

Simulation results establish the applicability of the method through trading of a synthetic asset, formed by a linear combination of dual-listed securities on Australian and New Zealand exchanges, in accordance with the prescribed strategy. The author investigates the range of values that the strategy return and Sharpe ratio can assume over different transaction costs and risk-free rates. An uncompetitive risk-free rate, for example, sees the Sharpe ratio of the strategy approach 10, while a competitive risk-free rate erodes the Sharpe ratio of the optimal trading strategy. Despite the limited scope of the simulation, the method described by Bertram (2010) for mean-reversion trade optimisation represents a novel contribution to the field, and one that is rooted in the dynamics of mean-reverting processes.

Evidence of different regimes is investigated by Bee and Gatti (2015), who find that the standard time series framework for statistical arbitrage benefits from modelling the mean-reverting spread conditionally on volatility-induced regimes. The authors note that mean-reverting spreads can contain structural breaks, such that the assumption of constant parameters is unreasonable. Some of the listed factors influencing structural breaks include financial crises, wars, political change and economic bubbles. The introduction of a regime switching methodology allows the trading models proposed by Bee and Gatti (2015) to absorb and even benefit from structural breaks, offering a more robust trading paradigm than conventional methods.

The authors identify three anomalies that traditional statistical arbitrage strategies might suffer from: the long-term equilibrium relationship might fail, causing the spread to no longer converge; the equilibrium relationship might be restored over a long period of time beyond allowable holding times for some investors; the spread volatility might increase, causing entry signals to be generated prematurely before the full spread divergence is realised. The objective of the methodology is to address non-constant volatility of the spread, prescribing different entry and exit thresholds for the differing regimes.

Two distinct volatility regimes are modelled according to both a two-population Gaussian mixture model, and a Hidden Markov Model (HMM). Both approaches to regime modelling seek to identify a threshold volatility level which separates the regimes, though the latter can be thought of as a generalisation of the first allowing time-dependence to be introduced into the regime switching. The approach first identifies potential pairs trading candidates based on the confidence level of a cointegration test, and then creates an artificial mean-reverting security constructed from a linear combination of security prices. The trading strategy sees a long (short) position entered once the spread level moves below (above) a multiple of the standard deviation of the spread volatility from its mean, with the position exited once the spread converges to its mean. The thresholds, along with the parameters of the regime switching models, are estimated using the EM algorithm.

The efficacy of the approach is tested on various futures contracts sampled at 10-minute intervals. In the presence of transaction costs consisting of a \$3 commission fee and one tick of slippage, the statistical arbitrage strategy proposed by Bee and Gatti (2015) realises Sharpe ratios between 2.9 and 6.3, annualised returns of between 2.8% and 41.9%, and maximum drawdowns of between only −3.6% and −0.2% over the selected sample. Additionally, the presence of two regimes offers an improvement over the same strategy executed with only one regime and the assumption of homoscedasticity in the spread.

The presence of regimes in the returns of an artificial security is investigated by Chen, Chen, and Chen (2014), where the authors propose a statistical arbitrage strategy that models the return spread of a pair of securities in terms of a three-regime autoregressive Generalised Autoregressive-Conditional Heteroscedastic (GARCH) process. The spread, $y_t = r_t^1 - r_t^2$, where r_t^i is the log-return of the ith security, is modelled as an autoregressive process specified by

$$y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + a_t, & y_{t-d} < c_1 \\ \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + a_t, & c_1 \leq y_{t-d} < c_2 \\ \phi_0^{(3)} + \phi_1^{(3)} y_{t-1} + a_t, & y_{t-d} \geq c_2 \end{cases} \tag{2.32}$$

$$a_t = \sqrt{h_t}\epsilon_t, \quad \epsilon_t \sim t_v^*, \tag{2.33}$$

$$h_t = \begin{cases} \alpha_0^{(1)} + \alpha_1^{(1)} a_{t-1}^2 + \beta_1^{(1)} h_{t-1}, & y_{t-d} < c_1 \\ \alpha_0^{(2)} + \alpha_1^{(2)} a_{t-1}^2 + \beta_1^{(2)} h_{t-1}, & c_1 \leq y_{t-d} < c_2 \\ \alpha_0^{(3)} + \alpha_1^{(3)} a_{t-1}^2 + \beta_1^{(3)} h_{t-1}, & y_{t-d} \geq c_2 \end{cases} \tag{2.34}$$

where $c_1 < c_2 \in \mathbb{R}$ are threshold values that determine the boundaries of the regimes, $d \in \mathbb{N}$ is the threshold lag, and t_v^* is a standardised Student-t error distribution with zero mean and unit variance.

The model parameters, $\left(\boldsymbol{\phi}^{(j)}, \boldsymbol{\alpha}^{(j)}, \beta_1^{(j)}, d, c_1, c_2\right)$ for $j = 1, 2, 3$, are estimated by Markov Chain Monte Carlo (MCMC) simulation over a training partition, for each pair selected according to the minimal SSD of their normalised price series. A very simple trading strategy follows, where a long position is taken during the testing partition when the realised spread, y_t, falls below c_1, and exited once it crosses back above c_1. A short position is taken when the realised spread crosses above the upper threshold, c_2, and exited once it crosses back below c_2.

Though the methodology proposed by Chen, Chen, and Chen (2014) models the effects of different regimes, those regimes do not alter the trade logic in an explicit way as they do in the model of Bee and Gatti (2015). The regimes simply influence the estimation of the upper and lower entry threshold parameters, c_1 and c_2, with no consideration given to how the strategy should be managed in the presence of different regimes. Despite this apparent model rigidity, the simulation results offer some evidence that the approach has economic potential.

The training partition of the dataset consists of daily prices of DJIA index constituents over the period January 2, 2006 to February 28, 2013. Five pairs are selected which exhibit the smallest SSD of their normalised prices, and their model parameters are estimated over the partition. Daily price data over the period March 1, 2013 to May 28, 2013 is withheld for the testing partition, with trading rules applied relative to each pair's estimated threshold values. The strategy returns over the testing period are 15.8%, 10.4%, 8.1%, 3.2% and 1.6% for the five pairs, respectively.

The methods proposed by Bee and Gatti (2015) and Chen, Chen, and Chen (2014) both investigate the presence of different regimes in statistical arbitrage equilibrium relationships, though they approach the matter in vastly different ways. While the former seeks to identify unique regimes and the trading signals that best apply to those regimes, the latter simply models the spread as an autoregressive process with its parameters determined by the prevailing regime. Despite the distinct contrast in methodologies, both give evidence of profitable trading potential and lend credibility to the idea that statistical arbitrage can benefit from inference regarding different market states.

Japanese technical indicators are considered in a framework proposed by Bogomolov (2013), with theoretical profitability of the framework estimated by an OU process. Tests indicate the methodology produces monthly excess returns of between 1.40% and 3.60% on Australian and U.S. equities data. Cummins and Bucca (2012) propose a time series approach that aggregates estimates of mean-reversion strength across 861 oil futures spreads, generating Sharpe ratios greater than two in many instances. Triantafyllopoulos and Montana (2011) follow a similar framework to Elliott, Van Der Hoek, and Malcolm (2005), though allowing for time-variation in the parameters which are subsequently estimated in a computationally-efficient state-space model. Zeng and Lee (2014) verify the continued profitability of statistical arbitrage under the time series approach in contrast to the distance approach.

The emergence of cryptocurrencies in general and Bitcoin in particular has offered arbitrageurs the opportunity to engage in statistical arbitrage on less mature, more volatile markets. Lintilhac and Tourin (2017) develop a vectorised OU process model for a portfolio of cointegrating cryptocurrencies, empirically verifying the model's ability to deliver positive returns both in- and out-of-sample. Commodities, though generally not as volatile as cryptocurrency markets, nonetheless offer arbitrageurs ample opportunity to exploit statistical arbitrage. Liu, Chang, and Geman (2017) investigate the use of the OU process in modelling spreads in oil company stocks, delivering an annualised Sharpe ratio and return of 3.90 and 188%, respectively, across the sample period June 2013 to April 2015. Applying their model only to market data from 2008, the authors report an annualised Sharpe ratio and return of 7.20 and 1,788%, respectively, verifying their hypothesis of market turmoil driving greater statistical arbitrage performance. Statistical arbitrage among oil company stocks from the S&P 500 is also investigated by Stübinger and Endres (2018), incorporating a jump-diffusion model in the authors' formulation of the mean-reverting OU process. Across a sample period covering 1998 to 2015, the authors report annualised returns of 60.61% and annualised Sharpe ratios of 5.30 after transaction costs.

2.1.4 Alternative Approaches

Statistical arbitrage opportunities can be identified and exploited by any number of strategies and modelling approaches, though the ones previously reviewed share strong conceptual commonalities that allow them to be categorised into either the distance, cointegration or time series approach. The distance approach was the first to bring academic scrutiny to the ill-defined concept of statistical arbitrage which had, up until that point, been the closely guarded proprietary knowledge of hedge funds. The cointegration approach built on the foundation laid by the distance approach, extending the concept of discovering relative mis-pricings between securities that move together to that of a rigorous mathematical framework. The time series approach, in turn, extended the formalised framework of cointegration by modelling the stationary spread between related securities in terms of an OU process, for which a number of analytic results are known.

Study	Objective	Data	Outcome
Huck (2010)	combination of machine learning and multi-criteria decision methods	US equities	proposed methodology produces positive daily returns for most parameterisations
Montana and Parrella (2009)	extension of Montana, Triantafyllopoulos, and Tsagaris (2009) framework to include machine learning forecasts	US equities	positive Sharpe ratio for most parameterisations
Avellaneda and Lee (2010)	modelling of bivariate spreads against hidden market factors	US equities	as few as ten hidden factors are required to explain 50% of return variance; proposed methodology produces positive Sharpe ratios for most time periods
Krauss and Stübinger (2017)	copula-based modelling of bivariate pairs in mean-reversion and mean-aversion settings	US equities	mean-reversion and mean-aversion pairs achieve comparable returns and Sharpe ratios
Hogan, Jarrow, Teo, and Warachka (2004)	development of statistical tests to identify statistical arbitrage	US equities	statistical tests indicate presence of statistical arbitrage in value and momentum strategies

Table 2.4: Collection of literature exploring alternative approaches to statistical arbitrage.

Alternative approaches to statistical arbitrage constitute a much broader subset of the literature, unified by the notion of identifying relatively overvalued and undervalued securities and forming a dollar-neutral and potentially market-neutral portfolio from their numbers. This relaxed criterion of the statistical arbitrage phenomenon allows many more modelling paradigms to be considered in the construction of the problem.

ANN model forecasts and their combination is addressed by Huck (2010) in one of the most cited papers of the alternative approach literature. Multi-step forecasts of security returns are first calculated; one-, two-, three- and four-period cumulative returns for securities are forecast using ANNs with input data consisting of the five most recent single-period returns of each security. The non-linear, non-parametric ANN model estimates the mapping function, f, in

$$r_i^{t+h|t} = f\left(r_i^t, r_i^{t-1}, r_i^{t-2}, r_i^{t-3}, r_i^{t-4}\right), \tag{2.35}$$

where $r_i^{t+h|t}$ is the cumulative return for security i over the horizon h beginning at time t, and r_i^t is the single-period return of security i from time $t-1$ to time t. The author notes that any forecasting methodology would be equally viable for this step of the procedure, though ANNs have the benefit of not requiring a rigid theoretical model in order to perform the task. This generality allows the models to forecast returns without requiring exogenous independent variables, simplifying the task considerably.

The next step of the procedure combines the ANN forecasts by calculating the spread between different securities' forecast returns for a given forecast horizon. A spread-based decision matrix, S, is constructed, in which the $(i, j)^{th}$ element is given by

$$s_{i,j} = \hat{r}_i^{t+h|t} - \hat{r}_j^{t+h|t}, \tag{2.36}$$

for $i, j = 1, 2, \ldots, N$ given N securities under consideration. The resulting anti-symmetric decision matrix, representing all information about forecast future return differentials, is then used to rank the securities in terms of their relative under- or over-valuation. The method by which the securities are ranked, ELECTRE III, is an outranking method based on the concept of fuzzy logic. Though the workings of the ELECTRE III method are not discussed in the paper, the parameter values used for the empirical investigation are published for reference.

Once the forecast security returns have been ranked, the proposed trading strategy selects the p most undervalued and overvalued securities for a given horizon for construction of statistical arbitrage portfolios. The resulting portfolios are dollar-neutral, with an equal allocation of capital afforded to the short overvalued securities and the long undervalued securities. The number of securities in a portfolio at the beginning of the trading period is $2p$, while the number of overlapping portfolios at any given time is h, given the possibility of running multiple instances of the strategy across more granular time intervals than the forecast horizon, h. The securities are either held for the duration of the forecast horizon, or liquidated earlier if cumulative returns are greater in magnitude than some threshold, c. This threshold applies equally to positive and negative cumulative returns, working concurrently as a stop-loss and profit cap.

The dataset considered under the empirical investigation consists of stocks from the S&P 100 equity index, with only those companies that were listed on the market before 1992 considered in the analysis. The resulting 90 companies' weekly returns are then calculated, based on the closing price from one Friday to the next. The sample spans the period January 1993 to December 2006, with separate simulations run on statistical arbitrage portfolios formed from one, two, three and four week-ahead forecasts. Possible values for the number of long and short positions, p, are 5, 10 and 15, while the possible closing threshold values, c, are 10% and 20%.

The results indicate the relative performance of statistical arbitrage portfolios formed from short forecast horizons, with the one-step weekly forecasts producing the greatest mean return per pair of 0.73%. This result was achieved with parameter value $p = 5$ and no closing threshold, indicating that the top and bottom five securities form the arbitrage portfolios which are held for the entire duration of the forecast horizon. The number of pairs that realised a positive return at the end of trading was found to be 52.49%, a figure which decreased for different parameterisations of the strategy. The mean daily return of the portfolio was found to be 0.15%, with the maximum positive daily return of 35.39% greatly surpassing the maximum negative return of −7.79%. By contrast, the worst performance was realised by the parameterisation $h = 4$, $p = 5$, with no closing threshold imposed. Of all 36 parameterisations, this was the only one to realise a negative mean return per pair, indicating the robustness and strength of the proposed methodology.

Statistical learning is again applied in a framework proposed by Montana and Parrella (2009), sharing similarities with the cointegration approach of Montana, Triantafyllopoulos, and Tsagaris (2009). The quasi-multivariate implementation of the cointegration approach with only the target considered for trading purposes is preserved, though with a different target security from that of the former work. The methodology is intended as a more generalised application of the cointegration-based statistical arbitrage framework, leveraging the approximation and mapping capabilities of statistical and machine learning in a trading setting.

The proposed methodology begins with streaming observations of $n + 1$ securities' prices beginning at time t; n independent securities, and one co-evolving target security, y_t, that is fundamentally tied to the other n. The dimensionality of the n securities is then reduced to $k \leq n$ principal components that are incrementally extracted from the streaming data. In this work only the first principal component, x_t, is extracted. The methodology proceeds with the approximation

$$y_t = f\left(x_t; \boldsymbol{\phi}\right), \tag{2.37}$$

in which f is an incremental version of Support Vector Regression (SVR), whose hyperparameters are specified by the vector, $\boldsymbol{\phi}$. The rationale for the use of incremental algorithms for both dimension reduction and function approximation lies in the necessity for trading algorithms to efficiently parse data to extract useful information, accounting for both relevant historical relationships and evolving data generation processes. The authors argue that the algorithms should ideally meet these requirements without requiring access to the full dataset that has been observed up until the present time.

The generalised statistical arbitrage approach proposed by Montana and Parrella (2009) assumes the realised price of the target security, y_t, can be decomposed into its true fair value, z_t, and a potential mis-pricing, m_t. The approximation performed by the SVR model is thought to model the true fair value of the security giving the approximation $\hat{y}_t \approx z_t$, since the true value can never be directly observed. An estimate of the potential mis-pricing is then constructed by subtracting the approximation of the theoretical fair price from the observed price, giving $m_t = y_t - \hat{y}_t$. A simple trading strategy would see a long position taken in the security when it is undervalued, and a short position when the security is overvalued. The decision rule, $d_t(m_t) = -\operatorname{sgn}(m_t)$, therefore suggests a position be taken at every epoch in opposition to the sign of the potential mis-pricing.

The final requirement of the proposed approach is the specification of the SVR hyperparameters, $\boldsymbol{\phi}$. Ordinarily this vector would be chosen to optimise some objective function over a training partition of the data, or incrementally tuned to improve the performance of the approximation. The authors instead propose the creation of a phase space signifying the different possible parameter values of the modelling approach, all of which are considered jointly. This ensembling method sees many different SVR models perform the approximation task with slight variations in parameter values, the results of which are combined through weighted majority voting—those models that falsely suggest a particular transaction have their decision rule, $d_t(m_t)$, multiplied by a fixed scalar coefficient, $0 \leq \beta < 1$, every epoch that a false signal is given. The effect is to gradually reduce the influence of inaccurate models so that the sum of signals more often reflects those of the remaining accurate models.

Data for the empirical analysis consists of daily closing prices of S&P 500 stocks, comprising the list of n independent variables, and the target iShare S&P 500 Index Fund ETF, both over the period May 2000 to June 2007. The phase space of SVR parameterisations leads to 2,560 separate models of the potential mis-pricing. The trading activity derived from these models' decision rules gives an average Sharpe ratio of 0.51 after transaction costs, realising an average annualised return of 8.00%. The best model achieved a Sharpe ratio of 1.10, while the worst model achieved a Sharpe ratio of -0.39. Applying the weighted majority voting scheme to these 2,560 models, the greatest average Sharpe ratio of 1.45 was achieved with a β value of 0.7, though all tested values of β achieved a greater Sharpe ratio than the sample average before including weighted majority voting. This simple ensembling technique, combined with mis-pricing estimates generated by SVR models, offers a very attractive and robust methodology for exploiting departures from fair value equilibrium.

Principal Component Analysis (PCA) and its application to the identification of statistical arbitrage opportunities is investigated by Avellaneda and Lee (2010). The proposed methodology assumes that security returns are the sum of systematic and idiosyncratic risk premia that can be decomposed into their component parts. Though neither systematic nor idiosyncratic risk factors can be observed directly, their influence can be approximated by extracting significant orthogonal hidden factors and modelling security returns in terms of their projection onto these factors. The resulting methodology allows the arbitrageur to specify a model that accounts for market-wide influences and achieves some degree of market neutrality.

It is assumed that stock returns follow the model

$$R_i = \beta_i F + \tilde{R}_i, \tag{2.38}$$

where R_i is the return of stock i over some period, F is the return of the market portfolio, β_i is the loading of stock i on the market portfolio, and $\tilde{R}_i$ is the idiosyncratic return of stock i. This specification is consistent with CAPM, and allows for the conceptual extension of the market portfolio to a collection of uncorrelated risk factors onto which the returns of stock i are projected. This leads to the multi-factor model

$$R_i = \sum_{j=1}^{m} \beta_{ij} F_j + \tilde{R}_i, \tag{2.39}$$

where F_j represents the return of the jth hidden factor. This representation allows the returns to be modelled with arbitrary accuracy, depending on the number of factors, m, chosen to represent the aggregated market portfolio. An alternative formulation is also offered by the authors, in which industry-specific ETFs are assumed to proxy the market portfolio of which stock i is a constituent. In this case, $m = 1$ and the hidden factor, F_1, is simply the ETF return series.

The hidden market factors are empirically estimated and extracted from the data by performing PCA on the stock returns. This allows the arbitrageur to specify up to N principal components to represent the aggregated market portfolio, given N securities in the stock universe. The authors suggest that the first 15 principal components are sufficient to explain the majority of variance in stock returns, though they also propose varying the number of principal components in order to account for a fixed percentage of the model returns; 45%, 55% and 65% were chosen as target levels of variance explained by the principal components. One notable result discovered using this approach was the inverse relationship between the number of principal components and the level of market volatility—times of high volatility required few principal components to explain a given percentage of model variance, while times of low volatility required a greater number of principal components.

Following the approximation of m market risk factors, idiosyncratic stock returns are further decomposed into their trend and dispersion elements, and modelled by the OU process,

$$\mathrm{d}\tilde{R}_i(t) = \alpha_i \,\mathrm{d}t + \kappa_i \left(\mu_i - \tilde{R}_i(t)\right) \mathrm{d}t + \sigma_i \,\mathrm{d}W_i(t), \tag{2.40}$$

given some mean-reversion rate, $\kappa_i > 0$, long-term mean, μ_i, and dispersion, σ_i. The additional drift term, α_i, is included to absorb the excess return of the stock relative to the market or sector ETF that it is modelled in relation to. In the empirical analysis, all model parameters and principal components are estimated and extracted over a rolling window of 60 days, in which the parameters are assumed to be constant. A trading strategy is then proposed which makes use of the dimensionless s-score metric,

$$s_i = \frac{\tilde{R}_i(t) - \mu_i}{\sigma_i \sqrt{\frac{1}{2\kappa_i}}}. \tag{2.41}$$

The strategy prescribes entering a long position in stock i when $s_i < -1.25$ and exiting when $s_i > -0.50$, while a short position should be entered when $s_i > 1.25$ and exited when $s_i < 0.75$. The asymmetric exit criteria reflect the tendency for stock markets to have an upward bias, with the parameter values chosen and validated in-sample in preparation for out-of-sample testing.

The empirical analysis considers daily closing prices of U.S. companies with a market capitalisation greater than or equal to USD $1 billion at the time of trading, over the period 1996 to 2007. The results indicate a Sharpe ratio of 1.10 over the sample period for the strategy in which sector ETFs were used as proxy for the market portfolio. The strategy which used 15 principal component factors achieved a Sharpe ratio of 1.44 over the period, while results for the varying number of principal component factors were only published over the sub-sample period 2002 to 2007; Sharpe ratios of 0.60, 0.70 and 0.40 were realised for the factors constituting 45%, 55% and 65% of explained variance, respectively.

Statistical arbitrage and its theoretical assertions are rooted in the concept of mean-reversion—if related securities exhibit some divergence from their historical equilibrium, it is expected that the equilibrium will be restored in the future by market forces that direct the reversion. It is unsurprising, therefore, that little effort has been made to consider the antithetical phenomenon of mean-aversion within the context of statistical arbitrage. In a paper that helps to bridge the divide between the two phenomena, both mean-reversion and mean-aversion are considered in a statistical arbitrage framework proposed by Krauss and Stübinger (2017) that inverts trading signals depending on the type of phenomenon under consideration.

Given a universe of N securities, $N(N-1)/2$ combinations of paired securities are formed and considered as potential statistical arbitrage candidates. The methodology proceeds by calculating the empirical marginal distributions, F_{R_1}, F_{R_2}, for the returns of both securities, R_1, R_2, comprising the candidate pair. Scaled ranks are then calculated for the returns given the transformation $U_i = F_i(R_i)$, resulting in a sequence of uniformly distributed variables for security i. These scaled ranks are then passed to the elliptical bivariate Student's t-copula, $C_{\rho,\nu}(u_1, u_2)$, to model the joint distribution, $F_{R_1,R_2}(u_1, u_2)$.

The t-copula is given by

$$C_{\rho,\nu}(u_1,u_2) = t_{\rho,\nu}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\right), \tag{2.42}$$
$$= \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^2(1-\rho)^2}} \left(1 + \frac{s^2+t^2-2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} \mathrm{d}t\,\mathrm{d}s,$$

where the parameters, ρ and ν, are chosen so as to maximise the likelihood. The partial derivatives of the t-copula allow the conditional distributions of U_1 and U_2 to be extracted, and their consequent $(1-\alpha)\%$ confidence bands to be calculated, for $\alpha > 0$. The two regions beyond these confidence bands represent potential mis-pricings between the two securities due to a move beyond the empirical equilibrium of returns—one region suggests the first security is undervalued and the second security is overvalued, while the obverse applies to the other region.

During a formation period, the cumulative return of each candidate pair is calculated, given trading signals generated when the transformed returns fall outside the copula confidence bounds. The cumulative returns serve to rank the pairs in terms of their profitability given adherence to a mean-reversion trading strategy, where the spread is bought when it is undervalued and sold when it is overvalued. The authors recognise that a number of pairs exhibit mean-averting behaviour during the formation period, characterised by a negative cumulative return given a particular parameterisation of the strategy. This sustained divergence of equilibrium following a trading signal leads to the concept of momentum pairs, in which the trading signals are reversed in order to buy the spread when it is overvalued, and sell the spread when it is undervalued.

For both mean-reversion and momentum pairs, only candidates whose returns have a correlation coefficient greater than or equal to 0.6 are considered, and then only the top k pairs of each type are traded in the testing period. The result is a set of two portfolios traded concurrently, in which the k pairs that offered the greatest cumulative return in the formation period are traded in a mean-reversion setting, and the k pairs that offered the poorest cumulative return in the formation period are traded in a mean-aversion setting. This winners-minus-losers portfolio methodology is a common feature of momentum research and bears striking similarity to the concept of multivariate statistical arbitrage, suggesting that the two phenomena are intrinsically connected.

The data used in the empirical analysis consists of daily observations of S&P 100 stock prices over the period January 1990 to December 2014. A 60-month formation period is applied to each pair, in which 48 sets of 12-month estimation and subsequent one-month pseudo-trading periods are used to determine the ranking of the pairs. The k top-ranked pairs and k bottom-ranked pairs are then traded over the following 12-month period in mean-reversion and momentum capacities, respectively. In the analysis, $k \in \{5, 10, 20\}$ are the values chosen for comparison.

Monthly returns for the mean-reversion pairs of 0.65%, 0.63% and 0.38% for $k = 5, 10, 20$ respectively are observed, while the momentum pairs generated monthly returns of 0.59%, 0.54% and 0.49%. These returns correspond with Sharpe ratios of 1.54, 1.25 and 0.56 for the mean-reversion pairs, and ratios of 1.33, 1.19 and 1.01 for the momentum pairs. Immediately it is clear that fewer pairs are preferable to a greater number considered for trading, with returns diminishing with increased portfolio size. There additionally appears to be greater consistency in returns and Sharpe ratios across momentum pairs than mean-reversion pairs, displaying a more marginal reduction in profitability over the different values of k. Nevertheless, the mean-reversion pairs achieved the greatest profitability across all pairs for the parameterisation $k = 5$. These results compare favourably to a market buy-and-hold strategy which, while offering a greater monthly return of 0.76%, displays significantly greater volatility leading to a Sharpe ratio of only 0.35.

The consideration of momentum pairs represents a significant departure from traditional statistical arbitrage methodologies which are motivated by the concept of mean-reversion. Statistical arbitrage is, however, a generalisation of standard arbitrage, and one that imposes no caveats on the form of the statistical mis-pricing, provided such an opportunity exists. In a paper that establishes the theoretical and mathematical requirements of statistical arbitrage, Hogan, Jarrow, Teo, and Warachka (2004) extend the discourse to include traditional value and momentum strategies—the cross-sectional analogues of mean-reversion and mean-aversion—finding that roughly half of the tested strategies exhibit statistical arbitrage. This result establishes the presence of statistical arbitrage in unconventional methodologies, and makes a strong case for the consideration of mean-averting behaviour.

An arbitrageur is assumed to follow a zero initial cost, self-financing trading strategy, $\{x_t, y_t : t \geq 0\}$, comprising x_t units of a security, S_t, and y_t units of a money market account, B_t. The security may itself be representative of a portfolio of long and short positions in multiple securities, provided the strategy satisfies $x_0 S_0 + y_0 B_0 = 0$, with the money market account initialised at one dollar, $B_0 = 1$. If the cumulative profits at time t are denoted V_t, and their discounted present value are denoted $v_t = V_t / B_t$, the properties of v_t must satisfy four conditions in order for the strategy to generate statistical arbitrage:

1. $v_0 = 0$

2. $\lim_{t \to \infty} \mathbb{E}[v_t] > 0$

3. $\lim_{t\to\infty} Pr\left(v_t < 0\right) = 0$

4. $Pr\left(v_t < 0\right) > 0 \Rightarrow \lim_{t\to\infty} \frac{\mathbb{V}[v_t]}{t} = 0$

The first condition imposes the requirement of zero initial cost, while the second condition requires the strategy to have positive expected discounted returns. The third condition requires the probability of loss to converge to zero in the limit, and the fourth condition requires the time-averaged variance of discounted profits to converge to zero in the limit if there is a non-zero probability of loss at any given time. Standard arbitrage can be considered a special case of statistical arbitrage in which the probability of loss is zero across all time, nullifying the conditional statement of the fourth condition. Standard arbitrage can be transformed into an infinite horizon self-financing strategy by investing the arbitrage profits into the money market at some finite horizon, satisfying the first three conditions.

The convergence of statistical arbitrage to standard arbitrage is established in the paper, with the incremental trading profits satisfying

$$\Delta v_i = v_i - v_{i-1} = \mu i^\theta + \sigma i^\lambda z_i, \tag{2.43}$$

where $i = 1, 2, \ldots, T < \infty$, and $z_i \sim \mathcal{N}(0, 1)$ with $z_0 = 0$. Given this formulation, the discounted cumulative trading profits follow

$$v_t = \sum_{i=1}^{t} \Delta v_i \sim \mathcal{N}\left(\mu \sum_{i=1}^{t} i^\theta, \sigma^2 \sum_{i=1}^{t} i^{2\lambda}\right), \tag{2.44}$$

for $t \leq T$, while the log-likelihood of the incremental discounted profits,

$$\log L\left(\mu, \sigma^2, \lambda, \theta | \Delta v\right) = -\frac{1}{2} \sum_{i=1}^{t} \log\left(\sigma^2 i^{2\lambda}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{t} \frac{1}{i^{2\lambda}} \left(\Delta v_i - \mu i^\theta\right)^2, \tag{2.45}$$

allows parameter estimates to be obtained by maximising the likelihood. A trading strategy is said to generate statistical arbitrage with $(1 - \alpha)\%$ confidence if the following conditions are satisfied:

H1: $\hat{\mu} > 0$

H2: $\hat{\lambda} < 0$

H3: $\hat{\theta} > \max\left\{\hat{\lambda} - \frac{1}{2}, -1\right\}$

The individual hypotheses' p-values form an upper bound for the test's Type I error, requiring the sum of the p-values to be below α to conclude that the trading strategy generates statistical arbitrage. An alternative formulation of the definition assumes constant mean incremental profits, μ, modifying Equation (2.43) to give

$$\Delta v_i = \mu + \sigma i^{\lambda} z_i. \tag{2.46}$$

This simpler formulation has simpler requirements for a strategy to generate statistical arbitrage, with statistical arbitrage generated with $(1-\alpha)\%$ confidence if the first two hypotheses, H1 and H2, are satisfied. This constrained-mean formulation was preferred by the authors for the empirical analysis.

Data for the empirical analysis consists of monthly equity returns over the period January 1965 to December 2000, and across most of the stocks traded on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automated Quotations (NASDAQ) markets. The trading strategies employed form the basis of momentum and value research, respectively, with the momentum strategies modelled directly after those of Jegadeesh and Titman (1993), and the value strategies modelled after those of Lakonishok, Shleifer, and Vishny (1994). Different parameterisations of the momentum strategies result in 16 individual strategies that are tested, of which nine are found to produce statistical arbitrage at the 90% confidence level, and six at the 95% confidence level. Additionally, of the 12 parameterisations of value strategies, six are found to produce statistical arbitrage at the 90% level, and five at the 95% level. Of the momentum strategies, only one produced negative mean returns, while all of the value strategies produced positive mean returns.

This result serves to illustrate the strength of the statistical test proposed by Hogan, Jarrow, Teo, and Warachka (2004), which is able to determine the presence of statistical arbitrage even among strategies that consistently generate positive mean returns. Additionally, the result lends credibility to the notion of statistical arbitrage being generated by both mean-reversion and mean-aversion phenomena. The statistical test is also shown to have a slight bias for acceptance of the null hypothesis, further illustrating the continued profitability of the tested strategies, along with their ability to consistently generate statistical arbitrage.

ANN and cointegration models of oil future spreads are investigated by Dunis, Laws, and Evans (2008), finding that ANN models achieve the greatest in-sample and out-of-sample performance. Genetic algorithm and ANN models are also used by Dunis, Laws, Middleton, and Karathanasopoulos (2015) to model the corn/ethanol crush spread, finding that the genetic algorithm produces the greatest statistical arbitrage returns. Huck (2009) aggregates multiple return forecasts across bivariate combinations of securities to generate a global arbitrage portfolio based on return rankings.

Alternative approaches to statistical arbitrage have continued to benefit from the enthusiasm for big data made possible by the advancement of computing resources. At the time of writing, alternative approaches constitute the most active field of statistical arbitrage research. Stübinger, Mangold, and Krauss (2018) apply vine copulas to the modelling of multivariate arbitrage relationships among S&P 500 stocks, reporting annualised returns of 9.25% and an annualised Sharpe ratio of 1.12 after transaction costs across the sample period 1992 to 2015. The model also limits downside risk, realising a maximum drawdown of only 6.57% during the in-sample period. Causal relationships are investigated by Stübinger (2019), with a framework proposed for exploiting statistical mis-pricings arising from a temporally-dislocated arbitrage relationship. Among pairs selected from the S&P 500 spanning 1998 through 2015, the approach delivers annualised returns of 54.98% and an annualised Sharpe ratio of 3.57. Huck (2019) embraces the potential of big data, applying machine learning models to the prediction of future stock returns with the aid of some 600 predictive variables. Despite delivering positive excess returns between 1993 and 2008, they are not significant after considering transaction costs and accounting for common risk factors.

2.1.5 Failure of Statistical Arbitrage

The objective of any statistical arbitrage trading model is to identify a suitable opportunity for exploiting temporary mis-pricings between two or more related securities. The suitability of these opportunities is contingent on the stability of the relationship between the securities' prices. Consequently, the modelling assumptions of statistical arbitrage are crucial to the sustained profitability of the various identification and exploitation approaches. The declining profitability of pairs trading, as discovered by Gatev, Goetzmann, and Rouwenhorst (2006) and subsequently validated by Do and Faff (2010), represents a threat to conventional statistical arbitrage and its precepts. In order to fully understand the nature of the problem, the sources and determinants of statistical arbitrage profitability, and more critically its failure, must be fully understood.

The increasing proportion of non-convergent trades was identified as the principal cause of declining statistical arbitrage profitability by Do and Faff (2010). Conversely, the relative increase in profitability during times of market volatility and stress during the authors' sample period was attributable to a marked reduction in the proportion of non-convergent trades. The tendency for a pair to fail to converge following the opening of a position illustrates a form of arbitrage risk assumed by the arbitrageur. Do and Faff (2010) identify fundamental risk, noise-trader risk and synchronisation risk as three sources of arbitrage risk. Fundamental risk refers to the possibility of an unexpected disruption to the relative pricing relationship, noise-trader risk refers to further divergence of the price spread due to increased activity from irrational market participants, while synchronisation risk arises when there is uncertainty about when arbitrageurs will exploit a temporal mispricing. Any of these three risks can manifest as pair non-convergence in an affirmation of arbitrage risk, though the specific source may be obscured from the arbitrageur.

Andrade, Di Pietro, and Seasholes (2005) replicate the methodology of Gatev, Goetzmann, and Rouwenhorst (2006) and apply it to the 647 eligible constituents of the Taiwanese Stock Exchange over the period January 1994 to August 2002. The authors' interest in the Taiwanese market is due to its institutional features, allowing large pools of uninformed trades to be distinguished from informed transactions. The first notable contribution of the paper is the validation of the distance approach on Taiwanese equities, generating annualised excess returns of 10.18% over the sample period. The second contribution addresses the authors' hypothesis that pairs trading profitability is driven by uninformed trading shocks, confirming that uninformed net buying is highly correlated with the initial divergence of the arbitrage opportunity.

Pairs trading is conceptualised by Andrade, Di Pietro, and Seasholes (2005) as a trade between two securities with similar loadings on the latent risk factors that drive asset returns. Assuming those factor loadings remain relatively constant through time, the paired securities' returns should be highly correlated and their normalised prices minimise the SSD metric. Andrade, Di Pietro, and Seasholes (2005) find that there are two mechanisms by which a pairs trade in the identified securities can be opened: uninformed demand shocks and idiosyncratic shocks. The former arises when uninformed trading takes place in one of the constituent securities, accommodated by informed traders on a market with limited risk-bearing capacity. The price change is sufficient to trigger the opening of a pairs trade, though the divergence from fundamental value is corrected by the market since the security's factor loadings remain unchanged. Idiosyncratic shocks, on the other hand, can potentially change those factor loadings.

An example is given of two mining companies whose returns are highly correlated over a sufficiently long observation period. If one of those companies were to discover a significant mineral deposit, for example, the market would value that company's stocks more highly than those of its counter-party. Though the returns may continue to move together in the future, the price change would persist and potentially lead to losses for the arbitrageur. In this scenario the factor loadings remain unchanged but the idiosyncratic shock is sufficient to trigger an unprofitable pairs trade. Another possible idiosyncratic shock would arise if one of the companies were to diversify its capital by investing in another market sector, such as information technology through an internet venture. This scenario would fundamentally change the data generating process, de-coupling the factor loadings of the company from its counter-party and destroying the arbitrage relationship.

Both of these idiosyncratic shocks are examples of fundamental risk as described by Do and Faff (2010).

The divergence of the spread necessary to trigger a statistical arbitrage position is the subject of several event studies in the literature. Investigating the determinants of pairs trading profitability, Jacobs and Weber (2015) apply the distance approach of Gatev, Goetzmann, and Rouwenhorst (2006) to constituents of 34 global equity markets over the period January 2000 to December 2013. Large markets are found to produce more profitable pairs than smaller markets, a finding that is attributed to the greater likelihood of matching stocks with economically-significant substitutes in a pair. Additionally, the limits to arbitrage imposed on emerging markets, either by chance or design, lead to greater pairs trading returns than their developed market counterparts. The analysis moves on to a thorough event study of U.S. equities, following formulation of the hypothesis that limited investor attention on international pairs boosts profitability by delaying the inclusion of information to one of the stock's prices; those pairs whose constituents are less visible to investors are typically more profitable.

The characteristics of the average pair, identified on U.S. equity markets in accordance with the methodology of Gatev, Goetzmann, and Rouwenhorst (2006), indicate that successful statistical arbitrage opportunities exist among large and liquid securities. They are generally well-diversified among market sectors and industries, with more than 40% of the divergence necessary to initiate a pairs trade coming on the day of divergence. Moreover, Andrade, Di Pietro, and Seasholes (2005) find that 71.08% of pair divergence is attributable to the rising stock, which subsequently assumes the role of the short leg of the pair once a trade is initiated. Pairs therefore do not slowly drift apart, but rather experience a bullish shock to one of the stocks on the day of divergence that is usually sufficient to open a mean-reverting position. Like Do and Faff (2010), Jacobs and Weber (2015) find that the profitability of pairs trading of U.S. equities is largely attributable to the 36.2% of pairs that converge within a month of trade initiation, compared to the average of 28.6% and 19.2% of large and small global equities markets, respectively.

Idiosyncratic shocks in the form of firm-specific earnings, dividend news or general coverage are found to degrade the performance of pairs trades by a considerable degree. By contrast, pairs that are initiated following divergence on days when broad macroeconomic news are delivered experience increased profitability relative to the average. This increased profitability is especially pronounced for pairs comprised of less visible stocks, such as those that do not attract much investor attention. Returns are found to be greatest when a pair converges within several days of trade initiation, and similarly the probability of convergence is highest immediately following trade initiation, slowly decreasing thereafter. These results indicate that the slow diffusion of information across both constituents of the pair following a shock to one of the constituents is what drives pairs trading profitability. Furthermore, low analyst coverage, high idiosyncratic volatility, and disunity among analyst forecasts are associated with greater abnormal returns for pairs that diverge following common macroeconomic news. There is also evidence to suggest that some of the abnormal returns generated by pairs that are initiated during times of high investor distraction are related to limits to arbitrage.

The returns of pairs trading strategies are explored by Engelberg, Gao, and Jagannathan (2009), who consider both the standard six-month holding period used by Gatev, Goetzmann, and Rouwenhorst (2006), and a shorter ten-day holding period designed to exploit the convergence that immediately follows trade initiation. The results for the sample period spanning 1993 to 2006 indicate that the shorter holding period generates excess returns of 1.75% per month, while the standard six-month holding period generates excess returns of 0.70% per month, trading pairs formed from U.S. equities within the same industry. This result illustrates the magnitude of profit that can be earned during the beginning period of spread convergence following trade initiation. The authors also find evidence that pairs whose constituent stocks have relatively few common analysts outperform those that share a greater number of analysts, and that common institutional holdings can also have a detrimental effect on returns.

In their own event study of pair divergence and subsequent trade initiation, Engelberg, Gao, and Jagannathan (2009) conduct an extensive investigation of the determinants of trade profitability and various sources of risk to the arbitrageur. Idiosyncratic news is found to harm trading profits despite increasing the probability of trade initiation. Liquidity and short-term changes in liquidity are found to increase profitability, the probability of trade initiation and convergence speed while decreasing noise-trader risk and synchronisation risk at the expense of greater arbitrage risk—risks associated with holding the security or executing a trade are ameliorated only by taking on more noise-trader and synchronisation risk. The interplay of these various sources of risk demonstrates that statistical arbitrage is not a riskless strategy, but rather compensates the arbitrageur for assuming varying levels of these risks.

All of the findings discussed here consider the profitability and failure of the distance approach of Gatev, Goetzmann, and Rouwenhorst (2006), though Rad, Low, and Faff (2016) extend the analysis to the cointegration approach and a copula-based alternative approach. Over the sample period spanning 1962 to 2014, the authors report average monthly excess returns of 0.91%, 0.85% and 0.43% for the distance, cointegration and copula approaches, respectively, before transaction costs. Though the inclusion of transaction costs degrades the performance of the strategy, all approaches continue to deliver statistically significant excess returns. Investigating the profitability of the strategy across different sub-periods, the authors note that the risk-adjusted returns of statistical arbitrage increase until 1985, declining thereafter. The copula approach, in particular, delivers negative risk-adjusted returns from 2000 until 2014, with the distance and cointegration approaches delivering short intervals of negative risk-adjusted returns during this period.

For converged trades—that is, positions that are opened and subsequently closed during the trading period—Rad, Low, and Faff (2016) report that the distance approach achieves an average return of 4.26% and a Sharpe ratio of 1.79, while the cointegration approach achieves an average 4.37% return with a Sharpe ratio of 1.62, and the copula approach achieves an average return of 3.95% with a Sharpe ratio of 1.05. Of all converged trades, the proportion that are profitable is 98.42%, 98.64% and 94.41% for the distance, cointegration and copula approaches, respectively. These high proportions suggest that if a pairs trade is able to converge during the trading period, it will most likely be profitable.

Non-convergent trades produce returns of −3.99%, −4.36% and −2.15% for the distance, cointegration and copula approaches, respectively. Despite the high proportion of profitable trades among convergent pairs, the three methods generate positive returns for only 71%, 69% and 59% of all trades, respectively. Non-convergence drives the degradation of pairs trading performance that has been observed since its initial academic exploration, and the findings of Rad, Low, and Faff (2016) help to strengthen this assertion. Only 62.53% of distance approach pairs converge during the trading period, with the cointegration approach producing even fewer convergent trades at only 61.35%. Nearly 40% of all trades initiated under the distance and cointegration approaches fail to converge and have a high probability of accruing losses. The copula approach is significantly worse than the other two approaches in this regard, with only 39.98% of trades converging during the trading period. More than half of all copula approach pairs fail to converge, explaining the relatively poor performance of the approach in comparison with the distance and cointegration approaches.

Non-convergence is the principal cause of failure for statistical arbitrage strategies. The mechanisms for non-convergence include idiosyncratic shocks, which can either cause persisting price divergence or a complete disruption and breakdown of the arbitrage relationship. Despite the enhanced statistical rigour employed by the more advanced cointegration and alternative approaches, idiosyncratic shocks still remain a salient feature of the domain. The failure of statistical arbitrage strategies to accommodate idiosyncratic shocks in its assumptions and modelling methodologies is a significant source of academic tension, offering the opportunity to expand the literature by addressing this deficiency.

2.2 Regime Switching and Statistical Learning

The presence and subsequent investigation of regimes within the statistical arbitrage literature is catalogued in Section 2.1. Do and Faff (2010) find that periods of high volatility and market turmoil are conducive to statistical arbitrage profitability. Caldeira and Moura (2013) echo these findings in their own investigation, though Huck (2015) is unable to find evidence of volatility regimes influencing statistical arbitrage returns. Chen, Chen, and Chen (2014) incorporate a regime switching framework into their statistical arbitrage model, and both Krauss and Stübinger (2017) and Hogan, Jarrow, Teo, and Warachka (2004) consider mean-aversion in statistical arbitrage as alternative regimes under which the arbitrage relationship can operate. More broadly, the failure of statistical arbitrage due to structural changes in the arbitrage relationship between paired securities, and the consequent emergence of time-varying estimation procedures as discussed in Section 2.1, implicitly strengthen arguments in favour of considering statistical arbitrage under a regime switching framework.

Financial markets are subject to shifts in fundamental economic conditions that manifest in distinctive behaviour observed in the evolution of security prices. Such shifts can result from technological advancement, changing consumption trends, legislative and central bank intervention, geopolitical stressors, and evolving regulatory requirements to name but a few sources. The breadth of potential triggers for regime shifts is too great to count or consider, and too complex to model or forecast comprehensively. These shifts may manifest as persistent periods with certain distributional properties, or transitory jumps that are unlikely to recur.

The importance of narrative in empirical finance has served to advance the use of regime switching models. Different phases of the business cycle, for example, generate security returns with different distributions that can be modelled and interpreted, though the identification of economic regimes is typically only possible ex-post. Nevertheless, regime switching models allow contemporaneous forecasting and prediction to be conducted within the context of an economically intuitive modelling paradigm.

The number of regimes, K, chosen to characterise a system can be inferred from the specific modelling application. If the practitioner were concerned with modelling bull and bear markets, for example, or high and low volatility regimes, a suitable value for K is two. Alternatively, if the application required the modelling of regimes instigated by technological change, regimes that are unlikely to recur in the future, then K might assume a much greater value. In this change point process, the number of unique regimes would grow over time as previous regimes were no longer able to capture the salient features of evolutionary market forces.

According to the survey of regime switching models conducted by Ang and Timmermann (2012), the economic rationale for the model should be stressed, leading to a choice of K that reflects some fundamental consideration of the process. The authors argue that it is difficult to base the decision of number of regimes on econometric tests due to the difficulty of implementing such tests, thanks largely to the non-standard distributions of the test statistics. The extensive survey of Markov regime switching literature conducted by Guidolin (2011) finds that roughly half of the applications use Markov switching models as a non-linear alternative to simpler linear models, in which the model is permitted to take the form that attains the best fit of the data. In contrast to the economically-inferred number of regimes, this approach is typically used in applications where the accuracy of point estimates is of greater concern than regime interpretability.

The strength of specific modelling and forecasting applications is a function of data quality and modelling paradigm. The rigidity of the canonical regime switching model constrains its application considerably, making it difficult to consider complex non-linear or non-parametric interactions that could otherwise offer some predictive power. If the principal concern of a practitioner is obtaining reasonable estimates of the prevailing regime, the modelling of regimes and their switching behaviour can alternatively be undertaken by more generic statistical learning models with universal approximation capabilities. Such models offer the practitioner a greater degree of flexibility in terms of data format, modelling application, and mapping capability. The breadth of statistical and machine learning models augmenting statistical arbitrage frameworks presented in Section 2.1.4 is testament to their broad applicability, flexibility and popularity.

ANNs and their variants represent the state-of-the-art in statistical learning and universal approximation. The resulting models optimise the mapping function relating a vector of independent input variables to a potentially transformed vector of dependent output variables—an application of supervised learning, in which the desired outcome is supplied *a priori*. The computational burden of training ANN models is significant, though recent advances have simplified the training procedure of Single-Layer Feedforward Network (SLFN) architectures considerably, allowing the number and breadth of applications to grow substantially.

Following Hastie, Tibshirani, and Friedman (2001), the archetypal statistical learning model, such as that to which ANNs adhere, optimises a functional mapping, $f : \mathbb{R}^n \to \mathbb{R}^m$, between a vector of inputs, $\boldsymbol{x} \in \mathbb{R}^n$, and a vector of desired outputs, $\boldsymbol{y} \in \mathbb{R}^m$. The modelling procedure typically requires data to be partitioned into training and testing sets, with the former dataset used for optimising the mapping function, and the latter reserved for model validation and calibration. Multiple observations of paired input and output vectors are presented to the learning algorithm, and subsequently fed through a lattice network of computational nodes resembling the ANN's biological analogue.

There is significant intersection between regime switching models and statistical learning models. While regime switching models can be considered an unsupervised classification task with $m = K$ output variables, the supervised classification tasks to which ANNs are applied may be considered a form of non-linear, non-parametric regime modelling—provided a number of suitably defined classes are selected as output variables, the learning algorithm is capable of approximating any functional mapping that may truly exist between input and output variables, allowing the practitioner to define their own regimes. Section 2.2 therefore jointly considers the applications of regime switching and statistical learning models in finance, with particular emphasis placed on investigations of statistical arbitrage.

2.2.1 Regime Switching Applications

The majority of regime switching literature is devoted to equity return modelling though, as noted by Guidolin (2011), regime switching models have also found application in foreign exchange, corporate bond, real estate and fixed income markets. The initial wave of research sought to use regime switching models as a non-linear alternative to simpler Autoregressive-Conditional Heteroscedastic (ARCH) and GARCH models, for example, often providing a better fit for the data and improved forecasting ability. According to Ang and Timmermann (2012), the stylised behaviour of financial returns series is captured in regime switching models, which are able to account for fat-tailed distributions, volatility persistence, skewness and time-varying correlations. Later contributions to the literature, beginning in the late 1990s, instead saw attempts to reconcile the presence of multiple regimes with asset pricing theory by testing its assumptions and hypotheses.

Schaller and Van Norden (1997) investigate the presence of regimes in CRSP value-weighted monthly equity returns over the period January 1929 to December 1989. The sample covers a number of abnormally volatile periods, and represents a convenient dataset for assessing the ability of regime switching models to identify distinct regimes. The models considered allow for switching mean with constant variance, switching variance with constant mean, and switching mean and variance. Each model provides progressively better likelihood ratio statistics compared to a model in which no regime switching is considered, leading the authors to conclude that regime switching models are better able to capture equity returns than models that only consider a single regime.

Regime switching models began to infiltrate the statistical arbitrage domain in the late 2000s, with a number of contributions proposing different trading rules to be used under different regimes. Cui and Cui (2012) investigate regime switching models for cointegration, allowing the model to identify regimes under which a cointegration relationship breaks down temporarily. In a departure from standard Markov regime switching model estimation, the authors specify a Bayesian framework for the model that is estimated by MCMC methods. The objective is to identify periods in which the assumption of cointegration does not hold, allowing the arbitrageur to modify their trade entry logic. The model allows the autoregressive parameter in an Error Correction Model (ECM) to switch between some value within the unit circle under the stationary cointegrating regime, and a value of zero under the non-stationary non-cointegrating regime.

Using simulated data, an empirical analysis of the regime switching cointegration model estimated the parameters of the data with a high degree of accuracy, though it should be noted that the model parameters were estimated *a posteriori*—that is, smoothed regime probabilities were calculated. It is clear that the entire dataset would not be available to a practitioner for inference, and as such this model suffers from forward-looking bias. The model is subsequently applied to pairs trading, with differing trade entry thresholds estimated under the regime switching model and a model that assumes persistent cointegration. Under the regime switching model, the entry thresholds are significantly different from and substantially narrower than those estimated under the assumption of persistent cointegration.

A fundamental implication of cointegration is the mean-reverting behaviour of the price spread formed between two securities. Rather than modelling breaks in the cointegrating relationship, Yang, Tsai, Shyu, and Chang (2016) consider the evolution of the price spread in terms of an OU process with shifting mean-reversion, mean and variance terms. Given some price series, p_t, and related security, q_t, the evolution of the spread, $\delta_t = p_t - q_t$, is assumed to follow

$$\mathrm{d}\delta_t = \kappa_{S_t} \left(\mu_{S_t} - \delta_t \right) \mathrm{d}t + \sigma_{S_t} \, \mathrm{d}W_t, \tag{2.47}$$

with $S_t \in \{1, 2\}$ indicating the different regimes. The model is estimated by maximum likelihood, revealing a regime in which the mean-reversion rate and series mean are relatively high while the variance is low, and a second regime in which the mean-reversion rate and series mean are relatively low while the variance is high. The former regime corresponds with typical market dynamics, while the latter corresponds with periods of market turmoil.

The empirical analysis considers daily observations of constituent stocks comprising the S&P 500 index, over the period January 2006 to September 2012. Three trading rules are applied to test the relative performance of the regime switching strategy: the first trading rule enters positions in the spread when the regime switching stochastic spread model indicates a significant level of divergence; the second trading rule enters positions in the spread when a stochastic spread model with fixed mean-reversion rate, mean and variance indicate a significant level of divergence; and the third trading rule follows the original distance approach proposed by Gatev, Goetzmann, and Rouwenhorst (2006). Trading periods covering one, two, three, six and 12 months are considered, with each trading strategy used to generate entry and exit signals for the duration of the trading period before re-estimating the models. The regime switching model produces greater Sharpe ratios than the two competing models, and greater average excess return across all trading periods. The authors also note the declining profitability of the distance approach, supporting the conclusions discussed by Do and Faff (2010).

If a pair of securities are cointegrated, their price spread will be stationary and mean-reverting. Similarly, the ratio of prices of cointegrated securities will be stationary with a long-term mean representing the hedge ratio between securities. Making use of this knowledge, Bock and Mestel (2009) investigate a departure from the distance approach by considering the mean-reverting ratio of prices in a regime switching framework. Given cointegrated price series, p_t and q_t, their ratio, $z_t = p_t/q_t$, is assumed to follow

$$z_t = \mu_{S_t} + \sigma_{S_t}\varepsilon_t, \tag{2.48}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$. This simple model with switching mean and variance is continually re-estimated within a moving window of 250 observations, though it is not mentioned precisely how the estimates are obtained.

The proposed trading strategy mimics the distance approach by placing transactions once the observed price ratio exceeds the regime-dependent mean by 1.645 standard deviations, corresponding with a 90% confidence interval about the mean. The regime switching model estimates two regimes: one in which the mean ratio is relatively high and volatility is low, and one in which the mean ratio is low and volatility is high. To ensure the greatest probability of trading success under circumstances where the price ratio lies between the estimated high and low mean values, the trade logic of the regime with the greatest smoothed or filtered probability is followed to generate a trading signal. The investigation focuses on the returns of regime switching statistical arbitrage applied to DJ STOXX 600 stocks, with little mention of the regime switching model and its performance. The proposed methodology nevertheless generates positive excess returns over the period June 2006 to November 2007.

While the Markov regime switching methodology offers a tractable framework for modelling non-linearities, such as those arising from the mixture of multiple distributions, its rigid assumptions and parametric form ultimately limit its scope of application. Moving beyond the standard Markov switching framework, Liu and Zhang (2010) investigate the relative improvement in modelling performance offered by ANNs over the Markov regime switching model in modelling the regimes of Chinese equity markets. The authors reason that the non-linear, non-parametric nature of ANNs are preferred because of their universal approximation capabilities, without needing to satisfy any requirements or assumptions about the modelling application. This broadens the range of applications considerably, though the specific task to which ANNs are applied by Liu and Zhang (2010) is the same as the Markov switching model, allowing a direct comparison between the two competing methodologies.

Adapting the canonical regime switching model to allow for factor modelling, the base model chosen to investigate regime switching in returns, r_t, of the Shanghai A Share Composite Index is given by

$$r_t = \beta_{1,S_t} x_{1,t} + \beta_{2,S_t} x_{2,t} + \beta_{3,S_t} x_{3,t} + \epsilon_t, \tag{2.49}$$

where $S_t \in \{1, 2\}$ indicates the prevailing regime. The covariates, $x_{1,t}, x_{2,t}, x_{3,t}$, are three macroeconomic variables selected by the authors which are believed to have predictive power for equity index returns. Specifically, consumer and corporate goods pricing indices are chosen, in addition to a variable that indicates the relative change in money supply between periods. The same three variables are parsed by a radial basis function ANN, whose output is the sign of the index return at time t. The alternative model is given by

$$\operatorname{sgn}(r_t) = f\left(x_{1,t}, x_{2,t}, x_{3,t}\right) + \epsilon_t, \tag{2.50}$$

where f is the radial basis ANN mapping function. The objective of the paper is to compare the estimated regime probabilities and forecasts, emphasising the relative performance of the two methodologies.

Daily observations of the Shanghai A Share Composite Index are taken over the period October 1999 to August 2009, allowing the logarithmic returns to be calculated for 2,368 periods. The estimated Markov regime switching model reveals a state in which mean daily returns are high and volatility is high, and another state in which daily returns are slightly negative but volatility is low. This finding suggests structural dissimilarities exist between Chinese and Western markets, which conversely tend to produce evidence of a high return, low volatility regime, and a low return, high volatility regime. The switching model estimates also indicate that the negative return regime tends to persist for twice the duration of the high return regime. For a more general comparison, the observed and predicted regime for each time period is compared between the two competing models. The total observed duration of the high return regime is 54 months, while the Markov model estimated returns that indicated the high return regime for only 24 months, and the ANN model estimated the sign of the return that indicated the high return regime for 48 months. The observed total duration of the negative return regime is 66 months, with the Markov switching and ANN models predicting durations of 57 months and 72 months, respectively. These results indicate the potential of ANN and other statistical learning models to be deployed in regime switching applications.

International portfolio diversification in the presence of regime switching is investigated by Ang and Bekaert (2002a). A domestic investor is found to achieve diversification benefit by allocating capital to international investments, even under regimes in which domestic and international return correlations and volatilities increase. Ang and Bekaert (2002b) find that there is strong evidence to suggest the presence of regimes in interest rates, and that Markov regime switching models are better able to forecast regimes and estimate sample moments than univariate models. Similarly, Dai, Singleton, and Yang (2007) develop a dynamic term-structure model incorporating regime-dependent risk factors, finding that the state-dependence captures asymmetry in the cyclical behaviour of interest rates. Rare disasters, including sharp contractions associated with war and economic crises, are modelled as regimes by Barro (2006), finding that the methodology is capable of explaining a number of economic puzzles such as high equity premiums and volatile stock returns.

Lettau and Van Nieuwerburgh (2008) investigate stock return predictability in the presence of regime switching. Under the proposed model, financial accounting ratios are adjusted to reflect the current regime, with strong empirical evidence supporting regime switching behaviour. Welch and Goyal (2008) and Pástor and Stambaugh (2001) estimate the equity premium in a regime switching framework, with regimes characterised by structural breaks. U.S. monetary policy is considered by Sims and Zha (2006), who find that the regimes inferred from a regime switching model closely align with commonly accepted opinions of shifts in monetary policy held by observers. Optimal portfolio allocation under changing regimes is investigated by Tu (2010), finding that the failure to account for regimes can lead to annualised losses of between 2% and 10%.

2.2.2 Statistical Learning Applications

Krauss, Do, and Huck (2017) investigate the relative performance of several statistical learning models, including deep learning ANNs, gradient-boosted trees, and random forests, the latter two constituting ensembling applications of decision trees. The investigation considers a classification task in which a given stock's future return is expected to outperform its industry median return. Model parameters are chosen according to heuristics and guidelines discussed in the literature. The specific architecture of the ANN, for example, has three hidden layers; the first hidden layer has the same number of nodes as the input layer, while the second and third hidden layers have successively fewer nodes to force a bottleneck that reduces the dimensionality of the model.

The binary classification task allows the models to calculate the probability that a stock's return in the next period will be greater than the industry median return. A relatively high probability indicates an expectation of outperformance, while a relatively low probability indicates an expectation of underperformance. This allows a statistical arbitrage portfolio to be constructed, with a long position taken in the k stocks with the highest probability, and a short position in the k stocks with the lowest probability, for $k \in \{1, 2, \ldots, N/2\}$ given an investable universe of N stocks. This long-short portfolio construction mirrors the winners-minus-losers methodology of the canonical momentum framework, pioneered by Jegadeesh and Titman (1993). Positions are held in the k upper and lower stocks for as long as their membership in their respective upper and lower portfolios continues, and liquidated thereafter. In addition to the three base models, ensembles are formed from the predictions generated by the ANN, gradient-boosted tree and random forest models. Results are only reported for one ensemble, in which the forecast probability of the stock exceeding the industry median return is averaged across all three models.

Empirical analysis of the statistical learning models is conducted on daily returns of constituents of the S&P 500 index over the period January 2000 to October 2015. The input variables for each model consist of cumulative returns over the previous $\{1, 2, \ldots, 19\}$ days, after which point the cumulative returns are calculated over the past $\{20, 40, \ldots, 240\}$ days to account for lower resolution monthly returns, generating a feature space of 31 variables. The results indicate that the portfolios with low values of k generate the greatest risk-adjusted returns, effectively excluding the relatively uncertain return predictions comprising the middle of the probability distribution. For $k = 10$, in which a long position is taken in the 10 stocks with the greatest positive return probability and a short position in the 10 stocks with the smallest positive return probability, the average daily return of the ANN model is 0.33% per day with a Sharpe ratio of 2.44—substantially greater than the market average daily return of 0.04% and Sharpe ratio of 0.35.

Despite the strong market outperformance offered by the ANN model, its performance is weakest among the various models and ensembles, with the random forest model achieving returns of 0.43% per day with a Sharpe ratio of 5.12. The greatest average daily return was achieved by the ensemble model, with daily returns of 0.45% and a Sharpe ratio of 4.71. The directional accuracy of all statistical learning models was greater than 50%, indicating that the models were correctly able to forecast the sign of the next period's return more than half the time. While all models are economically and statistically viable, even after incorporating transaction costs, the relatively weak performance of the

ANN model was attributed by the authors to the heuristic selection of model parameters, concluding that careful consideration of network architecture and input features would significantly improve its performance.

In a similar investigation, Qiu, Song, and Akagi (2016) apply a highly optimised ANN model to the prediction of monthly returns of the Nikkei 225 index. Using econometric variables, including T-bill rates, Yen spot rates, indices of industrial production and banknotes in circulation, the authors apply a fuzzy curve analysis dimension reduction technique in order to extract the most relevant variables for the prediction of Nikkei 225 monthly returns. The authors partition the dataset into training and testing sets, consisting of 170 months of observations over the period November 1993 to December 2007, and 67 months of observations over the period January 2008 to July 2013, respectively. The dimension reduction procedure is performed on the in-sample training set and limits the 71 potential variables to 18 significant ones, which are then pre-processed by standardising them within the range $[0, 1]$.

The ANN model is optimised relative to the training partition by repeating the training procedure 900 times, each time with a different parameterisation of the network. The authors selected 10 possible values for the size of the hidden layer, with the number of hidden nodes lying between 10 and 100 nodes. Additionally, the learning rate and number of training iterations were optimised, with the final network having 10 hidden nodes, 3,000 training iterations and a learning rate of 0.1. The authors additionally improved the ANN model by incorporating simulated annealing and genetic algorithms to further optimise the transfer weights and bias node activation values. The results indicate that the training procedure proposed by Qiu, Song, and Akagi (2016) offers greater prediction accuracy of the Nikkei 225 index than a standard backpropagation ANN in terms of Mean Square Error (MSE), though the economic performance of the procedure is not discussed.

A comparison between ANN and Support Vector Machine (SVM) models applied to stock index directional forecasting is offered by Kara, Boyacioglu, and Baykan (2011). The application concerns the directional accuracy of the two models when used to forecast daily returns of the Istanbul Stock Exchange National 100 Index. The authors note that the majority of literature in the statistical learning domain is devoted to applications in established equity markets, while emerging markets, and the Turkish equity market in particular, are often overlooked. Building on the established market literature, the authors investigate directional forecasting over the period January 1997 to December 2007 by first selecting 10 technical market indicators, commonly applied by fund managers, as input variables for the two models. The procedure then requires sub-setting of the data into training and testing partitions, with the number of positive and negative daily returns in the training set selected so as to be proportional to the number of positive and negative daily returns in the testing set.

The partitioning of training and testing sets in proportion to the observed occurrence of directional returns introduces a significant source of forward-looking bias, and one that is not addressed in the paper. Furthermore, observations from the entire sample are used to optimise parameters of the two statistical learning models. Nevertheless, the authors find that the proposed combination of technical indicator input variables and directional forecasting using statistical learning models offers encouraging results that might be useful in funds management and investment. The SVM model achieved directional accuracies of 71.52% on average over the period, with the greatest directional accuracy of 80.16% achieved in 2005. Similarly, the ANN model achieved an average directional accuracy of 75.74% over the period, with its greatest directional accuracy of 79.37% occurring in 1997. The two models are both able to extract meaningful relationships from the data, though the authors conclude that the ANN model offers the greatest modelling capability.

The burgeoning literature devoted to Extreme Learning Machine (ELM) models—the computationally-efficient generalisation of SLFNs pioneered by Huang, Zhu, and Siew (2004)—contains a large number of financial applications, though very few investigations of statistical arbitrage. One exception, authored by Nóbrega and Oliveira (2013), explores the comparative performance of ELM and SVR models, along with four ensembling techniques that combine the forecasts of the two individual models. The application considers five equity pairs chosen from the Brazilian iBovespa index, all of which are confirmed to exhibit stationarity in their price spreads, which are subsequently modelled by the OU process in order to estimate the mean-reversion half-life of the spread. This half-life, along with the first 10 lagged values of the price spread, historical volatility and spread mean are selected as input variables for both the ELM and SVR models. The dataset consists of all asset quotes observed between 2 January 2013 and 15 March 2013, with the first 25 days of data devoted to the training set, the following 12 days devoted to a validation set that is used to tune the ELM and SVR model parameters, and the final 13 days reserved for the out-of-sample testing set.

Ensembling techniques include Bayesian model averaging, Least Absolute Shrinkage and Selection Operator (LASSO), Granger and Ramanathan Regression (GRR) and Kalman filtering. While the former three techniques are static, the Kalman filter updates its forecasts sequentially in order to minimise the dynamic variance of the model error. The task of forecasting the next period's price spread is addressed, with the performance of each model and ensemble assessed in terms of Root Mean Square Error (RMSE) and the Theil-U statistic, with lower values for each measure indicating a better model forecast. While the ELM consistently outperformed the SVR model, both in-sample and out-of-sample, the Kalman filter achieved the greatest forecast accuracy by a substantial margin.

In terms of economic performance, a simple statistical arbitrage rule is used to generate entry and exit signals for each leg of the pair depending on whether the current price spread exceeds upper or lower confidence bounds around the forecast price spread. The performance measures are annualised return, annualised volatility and Sharpe ratio. The results indicate that the Kalman filter achieved the greatest in-sample performance with an annualised return of 89.27% for one pair, and a corresponding out-of-sample annualised return of 112.81%. The authors conclude that the iterative nature of the Kalman filter gives it a clear advantage over the static models. Additionally, the ELM model outperforms the SVR model in terms of in-sample return and both in-sample and out-of-sample Sharpe ratio, while the SVR model achieves the greatest out-of-sample annualised return for one pair.

In other financial applications, Abdou, Pointon, and El-Masry (2008) assess the performance of ANNs in credit scoring relative to conventional discriminant analysis, probit analysis and logistic regression models, finding that ANNs achieved the greatest classification rate. Credit scoring and bankruptcy prediction are also addressed by West, Dellana, and Qian (2005) and Tsai and Wu (2008), where both investigations consider ensembles of ANN models. The former finds that ensembles outperform single models, while the latter finds this to be true only when the ensembles are trained on the same data as the single model. Panda and Narasimhan (2007) use regression ANN models to forecast the Indian rupee relative to the U.S. dollar, finding that ANNs outperform autoregressive and random walk models both in-sample and out-of-sample. The determinants of capital structure are investigated by Pao (2008), who introduce ANN models as an alternative to standard linear regression models, finding that ANNs offer a greater model fit and forecasting accuracy.

2.3 Literature Review Summary and Research Questions

The distance approach to the identification and exploitation of statistical arbitrage opportunities laid the foundations of the framework upon which rigorous academic investigation could be mounted. Its most cited papers by Gatev, Goetzmann, and Rouwenhorst (2006) and Do and Faff (2010) propose a simple economic model-free algorithm which neither suffers from model mis-specification nor mis-estimation. The simplicity of the approach is its greatest strength, requiring only that pairs of normalised stock prices have a relatively low SSD over a 12-month formation period. Subsequent divergence of the pairs' prices by two or more historical standard deviations are viewed as temporal mis-pricings that can be exploited for profit.

This pairs trading approach to statistical arbitrage is quite intuitive, however it suffers from a number of inconsistencies in its objective of maximising a rational investor's excess returns. The ideal pair would exhibit frequent and strong divergence from and subsequent convergence to its equilibrium price. The spread between the ideal pair would consequently have a high variance and be strongly mean-reverting, neither of which are characteristics that the approach explicitly seeks to find. On the contrary, by selecting pairs whose SSD is relatively low, the distance approach of Gatev, Goetzmann, and Rouwenhorst (2006) favours those pairs that exhibit low spread variances and are not guaranteed to be mean-reverting.

In a cross-sectional investigation of strategy returns across different time periods, Do and Faff (2010) discovered declining profitability of the simple distance approach proposed by Gatev, Goetzmann, and Rouwenhorst (2006). The cause of the diminishing returns was found to be due to a greater proportion of non-convergent trades, as opposed to an increase in market efficiency. These pairs' equilibrium relationships did not persist over the trading period, so stronger conditions for pair selection were proposed by Do and Faff (2010) leading to moderate improvements in profitability.

Quasi-multivariate statistical arbitrage is addressed by Perlin (2007), though only the target security is traded in the empirical analysis. Monte Carlo data-mining bias tests confirm the robustness of the methodology, with various parameterisations of the strategy outperforming bootstrap portfolios constructed to approximate the distribution of returns under the assumption of random trader luck. Different pair selection methods are investigated by Huck (2015), with cointegration tests delivering the only pairs that are able to produce significant excess returns in the presence of transaction costs, though the standard distance approach is found to generate profit before transaction costs. Volatility regimes are introduced in an effort to capture the apparent market inefficiency that accompanies financial crises, though no evidence is found in support of volatility driving statistical arbitrage profitability. High-frequency trading in U.S. treasury securities is addressed by Nath (2003), with the distance approach generating positive excess returns even after full consideration of the restrictive transaction costs imposed by high-frequency trading. This subset of the literature verifies the results of Gatev, Goetzmann, and Rouwenhorst (2006), across multiple asset classes and timeframes, establishing statistical arbitrage as a true capital market anomaly.

The cointegration approach to statistical arbitrage further improves on the distance approach by requiring candidate pairs to exhibit a long-term equilibrium relationship, thereby improving the likelihood that any divergence between prices will be followed by convergence to the equilibrium level. Once a pair has satisfied a given cointegration test, the exploitation typically follows that of Gatev, Goetzmann, and Rouwenhorst (2006) where trades are entered once the pair has diverged by a multiple of its historical standard deviation, and closed following convergence.

The mathematical treatment of cointegration offered by Meucci (2009) represents a theoretical approach to security selection for potential exploitation of statistical arbitrage opportunities. Instead of specifying a complicated model of inter-security dynamics based on fundamental considerations, the arbitrageur need only consider the covariance of returns between securities. Positions can be taken in the securities in proportion to their cointegrating vector coefficient once the artificial security diverges significantly from zero, and closed when it returns to zero in the future. This approach to cointegration-based portfolio identification suffers from the requirement of a known covariance matrix of security returns. While the author notes that the sample covariance matrix approximates the true asymptotic covariance, it is also noted that in-sample cointegration is no guarantee of out-of-sample cointegration. In fact, those relationships that exhibit the greatest level of cointegration in-sample are typically the least robust out-of-sample.

This undesirable property is not unique to the covariance-based approach to estimation of cointegrating vectors. While noting that cointegration-based pairs have a higher rate of convergence than distance-based pairs, it was found by Huck (2015) that a substantial number of pairs identified by cointegration can either fail to converge or fail to do so within a reasonable timeframe. Though the cointegration approach advances a more rigorous framework for pairs trading and statistical arbitrage, and its empirical returns indicate an improvement over the distance approach, cointegration tests do not guarantee convergence of price spreads to their historical equilibrium level.

A time-varying cointegrating vector is considered by Montana, Triantafyllopoulos, and Tsagaris (2009), for which an estimate is derived by a given parameterisation of the FLS time-varying regression technique. This approach is shown to be more robust than traditional invariant cointegrating vectors with respect to model specification and evolving data generation processes. If a given model is a poor fit, the regression technique will update the cointegrating vector to reflect the changing dynamics of the model and its relationships. Consequently, the approach is able to generate significant excess returns in an application that relies solely on fundamental considerations to select the statistical arbitrage portfolio. The time-varying cointegrating regression is further developed by Burgess (2000), in which a statistical test capable of discovering mean-reversion in the presence of dynamic cointegrating relationships is proposed.

In a departure from the distance and cointegration approaches, the time series approach to statistical arbitrage is solely concerned with the exploitation of an equilibrium relationship, assuming that a number of co-moving securities or data streams has already been identified by some suitable method. The model advanced by Elliott, Van Der Hoek, and Malcolm (2005) describes the spread between securities in terms of a mean-reverting Gaussian Markov Chain observed in Gaussian noise. The continuous-time version of the mean-reverting spread can be represented as an OU process, while its discrete-time analogue is parameterised in the form of a state-space model. The strength of the time series approach is in its ability to derive closed-form solutions for a number of crucial quantities. The expected trade holding period, trade cycle duration, and optimal entry and exit thresholds were derived by Bertram (2010). Using these insights, an empirical investigation was able to produce pairs with Sharpe ratios approaching 10—a figure far surpassing that of any other methodology in the literature.

The presence of regimes is investigated by Bee and Gatti (2015), in which volatility is modelled in terms of a two-state HMM. In contrast to the results of Huck (2015), substantial evidence in support of switching volatility regimes is found, with Sharpe ratios between 2.9 and 6.3 realised in a simulation concerning futures contracts. Trading regimes are also investigated by Chen, Chen, and Chen (2014), though the proposed methodology cannot identify specific market states that correspond with periods of high or low profitability of the trading framework.

Proponents of the time series approach point to its comparatively high excess returns and its closed-form solutions of crucial trading quantities as evidence of its successful application to statistical arbitrage. It is important to note, however, that the approach nevertheless relies on existing methods for quantitatively identifying co-moving securities and data streams. The distance and cointegration approaches to security selection are typically assumed to have been employed, both of which assume invariance in the model parameters. Engle-Granger and Johansen tests for cointegration assume a fixed regression coefficient, while the distance-based identification ansatz relies on a fixed historical relationship estimated from the data. The time series approach consequently suffers from the same complications that plague the distance and cointegration approaches, namely sensitivity to model specification, and the potential dynamic evolution of the true equilibrium relationship.

The rigid structure of the distance, cointegration and time series approaches to statistical arbitrage is relaxed in the literature devoted to alternative approaches. The unifying concept is that of identifying some statistical mis-pricing, however it might present itself. The use of ANNs and multi-critera decision methods is explored by Huck (2010), generating a matrix of bivariate spreads between the forecasts of securities' future returns. The proposed methodology dispassionately considers all bivariate pairs, and allocates capital to those securities that have the greatest likelihood of having been mis-priced by the market. A similar approach is advanced by Montana and Parrella (2009), where all combinations of model parameters are considered jointly to produce a stronger picture of the mis-pricing between a target security and a fundamentally-related portfolio. While the former aggregates signals across securities, the latter aggregates across model parameterisations, with both approaches producing significant excess returns.

Hidden market factors constituting systematic risk drivers are considered by Avellaneda and Lee (2010), with individual statistical arbitrage opportunities identified in reference to those factors. One of the more interesting developments proposed in the framework is the extension of the OU process to include an idiosyncratic drift component. This model consideration accommodates equilibrium relationships that establish new levels over the course of the formation period, though the static estimation procedure implicitly assumes that the equilibrium drift occurs at a constant rate. Similarly, the consideration of mean-aversion in addition to traditional mean-reversion in the methodology of Krauss and Stübinger (2017) establishes the profitability of trading pairs that are expected to diverge further after some initial divergence. Though the presented copula-based approach to pair selection and trading differs considerably from the formalised model specified by Avellaneda and Lee (2010), both approaches seek to profit from pairs that exhibit both a fixed and variable equilibrium level.

Statistical arbitrage is formalised by Hogan, Jarrow, Teo, and Warachka (2004), with a set of four conditions determining whether a particular trading strategy generates statistical arbitrage or not. These conditions require the strategy to have zero initial cost, the discounted incremental profits to have positive expected value, the probability of loss to converge to zero, and the time-averaged variance of profits to converge to zero if there exists a non-zero probability of loss. Following extensive investigation of the proposed statistical tests for robustness, traditional value and momentum strategies are considered and found to generate statistical arbitrage for a number of parameterisations. This result serves to illustrate the presence of statistical arbitrage in unconventional places, and establishes the efficacy of the proposed statistical tests.

A number of conclusions can be drawn from the current state of the statistical arbitrage literature:

- The entry signals generated by the distance approach, in which the spread is bought or sold once it reaches some significant z-score, is replicated in most applications.
- Cointegration analysis, while capable of identifying arbitrage portfolios with greater profitability than the distance approach, still suffers from an inability to identify

a causal relationship between securities. The consequence is a framework that exaggerates the importance of potentially transient relationships.

- Time-varying regression methods offer greater flexibility in the estimation of a cointegrating relationship, adapting to absorb noise due to model mis-specification and evolving market conditions.
- The time series approach, with its theoretical considerations of the OU process, offers a number of analytical results that can be used to optimise the trading parameters of a statistical arbitrage opportunity.
- There is significant evidence of regimes in statistical arbitrage frameworks, though there is considerable variability in their modelling and application.

In light of these findings, it is clear that there are a number of commonalities that exist between all approaches. There are a number of inconsistencies, however, that need to be addressed in order to ensure compatibility between the various elements.

As noted in Section 2.1.5, the failure of statistical arbitrage can be attributed to model mis-specification or the evolution of the underlying system dynamics. If the equilibrium level shifts over time, the standard statistical arbitrage framework, which relies on static arbitrage relationships, will be unable to capture mean-reversion in the stochastic spread. The emergence of time-varying approaches for modelling the arbitrage relationship, such as those explored by Burgess (2000), Montana, Triantafyllopoulos, and Tsagaris (2009), Triantafyllopoulos and Montana (2011), Dunis, Laws, and Evans (2006), Nóbrega and Oliveira (2013) and Stübinger and Bredthauer (2017), illustrate a broad range of attempts to integrate a dynamic estimation procedure into the conventional approaches, though none have offered a theoretical justification nor investigated such procedures in reference to the phenomenon of declining profitability. Time-varying arbitrage models therefore offer an attractive avenue for research due to their ability to absorb and accommodate the stochastic variability of the arbitrage relationship, allowing the spread to adapt to evolving market dynamics irrespective of their underlying cause.

The first and most critical research question posited in this thesis concerns the declining profitability of pairs trading and how the rigid model assumptions of the conventional approaches fail to consider the empirical dynamics of the anomaly.

> *Is the assumption of static arbitrage relationships responsible for the declining profitability of statistical arbitrage?*

Nested within this question is the implicit acceptance of a specific modelling approach, be it a distance, cointegration, time series or alternative specification. There is, however, sufficient commonality among the distance, cointegration and time series approaches to motivate the development of a unified model for statistical arbitrage. Such a model would ideally capture the simplicity and tractability of the distance approach, the theoretical rigour of the cointegration approach, and the flexibility of the time series approach. The unified model proposed in Chapter 3 delivers these objectives by reformulating the spread variable to one that can be expressed directly as an OU process. The flexibility of the proposed unified model allows it to consider time-varying hedge ratios between cointegrated securities, allowing the research question to be explored through an extensive empirical simulation in Chapter 4.

Regime switching models as discussed in Section 2.2 offer a way to rationalise and reconcile security returns with asset pricing theory, accounting for fat-tailed distributions that result from temporal regimes driving the data generating process. Capturing the switching dynamics additionally allows the practitioner to infer the expected duration of a given regime, allowing for the careful planning and implementation of policy decisions. Though regime switching models are generally unable to forecast the kind of regime shift that might arise from technological advancement or a significant change in market microstructure, for example, they offer a powerful framework for ex post analysis of changing economic conditions. The different ways regime switching models can be used makes them attractive for fund managers, central bankers and financial policy regulators in understanding the underlying market conditions within which they interact.

The literature devoted to Markov regime switching models in empirical finance is surveyed and reviewed by Guidolin (2011), leading the author to a number of conclusions. The proportion of papers that use Markov switching models because the data requires it versus papers that propose the use of Markov switching based on some economic consideration stands at roughly 50:50, though there is increasing interest in using the models in concert with asset pricing theory, leading to greater interest in formulating the model in terms of some economic rationale. Another finding is the proportion of papers that allow for more than two regimes against those that only allow for two, which again stands at roughly 50:50. This is related to the first finding, given the tendency for an economic rationale to consider two competing regimes only, while a data-driven exploration of regimes may warrant many more than two.

Within the statistical arbitrage literature, regime switching models take many and varied forms, with significant deviations from the canonical Markov switching model observed in the literature. Regime switching models were used by Cui and Cui (2012) to characterise structural breaks in cointegrating relationships, which is the fundamental driver of mean-averting momentum behaviour in statistical arbitrage portfolios. Other investigations, such as those by Do and Faff (2010), Caldeira and Moura (2013), Huck (2015) and Bee and Gatti (2015), have considered the volatility of the equilibrium relationship to vary over time according to a regime switching model, and others such as Yang, Tsai, Shyu, and Chang (2016), Bock and Mestel (2009) and Chen, Chen, and Chen (2014) have considered a fully parametric regime switching process that drives the specific parameterisation of the arbitrage relationship. The specific role of a regime switching model depends on its intended application, though a number of conclusions can be drawn about the current state of the literature with reference to statistical arbitrage.

The number of regimes is typically fixed *a priori* and motivated by economic considerations. There is scope to extend the application to that in which the number of regimes is unbounded, particularly in the case of structural breaks in a cointegrating relationship, though evidence in the literature suggests that only two regimes are ever considered. Irrespective of the specific form that the regime switching model takes, all investigations discussed in Section 2.2.1 consider volatility as the most critical element of each implementation, and the most likely driving force behind regime switching behaviour.

Statistical learning models generated by ANN and ELM methodologies, as discussed in Section 2.2.2, offer a number of advantages over conventional statistical modelling and regime switching techniques, including their non-linear, non-parametric universal function approximation capability. They can be used in both regression and classification models, with any number of objective optimisation functions that can be specifically catered to the intended application. They can approximate functions with arbitrary precision, given the specification of a suitable number of hidden nodes, and ELMs have the additional advantage of good generalisation thanks to the minimum norm estimate obtained by OLS regression.

The use of statistical learning models in finance is well documented in the literature. A substantial number of ANN modelling applications in finance concern classification tasks, which can be considered a form of regime modelling under the right context. The regime switching literature surveyed in Section 2.2.1 typically requires the analyst to specify the number of suspected regimes present in a time series, subsequently fitting the best model to the data. This modelling paradigm does not allow the specification of the regimes' characteristics, and the analyst may consequently derive a model that is inconsistent with knowledge of the data generating process. By specifying the desired output in statistical learning models and allowing the model to optimise the mapping function, ANNs and ELMs offer the advantage of modelling flexibility at the expense of model transparency.

Given the evidence of regimes in statistical arbitrage, and the somewhat contradictory evidence of the impact of volatility regimes in particular, the second research question posited in this thesis considers the performance of the proposed unified model in the presence of regimes.

Are statistical arbitrage returns dependent on the prevailing volatility regime?

To simplify the analysis and provide a foundation on which to mount a solid investigation of the effects of volatility, the regime switching model proposed in Chapter 3 considers a two-regime process estimated by a logistic regression model. In particular, and in accordance with the guidance offered by Ang and Timmermann (2012) to select economically-relevant regimes, the two regimes modelled by the logistic regression framework are those of profitability and unprofitability, mirroring the methodologies presented by Krauss, Do, and Huck (2017) and Kara, Boyacioglu, and Baykan (2011). Though this simple approach does not employ the Markov switching framework favoured by many practitioners, it allows the specification of regimes that are ultimately of most interest to arbitrageurs, and allows the impact of volatility to be directly assessed.

Following the review of statistical learning applications in Section 2.2.2, and in view of the second research question, the third and final research questions posited in this thesis considers the additional economic benefit offered by statistical learning models.

Are statistical learning models better equipped than conventional models to capture and detect latent market regimes?

The universal approximation capability and computational efficiency of ELMs makes them the ideal candidate with which to test the comparative economic and statistical advantage offered by statistical learning models. Chapter 3 presents the ELM model as an alternative to the logistic regression model used to assess the impact of volatility on profitability proposed under the second research question, albeit with the inclusion of various exogenous variables that are hypothesised to contribute to model performance. In this way, the third research question and its consequent investigation seeks to determine whether exogenous data in concert with state-of-the-art statistical learning models offer additional benefit to the arbitrageur in pursuit of statistically significant excess returns.

3 Methodology

3.1 Introduction

The literature surveyed in Chapter 2 spans the evolution of statistical arbitrage, from its beginnings with the distance approach, to its cointegration and time series approach extensions, and finally to its abstraction and consideration as a data mining exercise under alternative approaches. The different approaches, and more specifically the different contributions that constitute the body of literature, have a simple unifying theme: find two or more securities that share some common relationship, estimate their equilibrium price, and execute a trading strategy when the observed deviation from equilibrium is significant. Following the research questions proposed in Section 2.3, and the first research question in particular which assumes a common approach to the exploitation of arbitrage relationships, Chapter 3 first seeks to establish a unifying model that links the distance, cointegration and time series approaches under a single formulation.

The simplicity of the concept of statistical arbitrage belies the complexity of its practical implementation, thanks largely to the myriad free parameters that must first be considered by the arbitrageur. What constitutes a common relationship can be defined quantitatively or qualitatively, with various statistical testing procedures available for ascertaining the strength of the relationship. This design consideration is perhaps the first point of difference between the various approaches; the simpler distance approach prescribes minimisation of the SSD statistic, while the cointegration approach requires the more rigorous condition of spread stationarity according to, for example, an ADF test. Neither the time series nor alternative approaches have a prescribed method for identifying statistical arbitrage opportunities, so a unifying framework must first be able to identify candidate pairs, with spread-stationarity a necessity for profitable trading.

Modelling of the spread under the time series approach is accomplished in terms of the mean-reverting OU process. The process is stationary, Gaussian and Markovian, consequently offering a number of attractive properties that have been studied extensively. Mean-reversion time can be determined by calculating the half-life of the process decay, and the expected holding period of a trade can be estimated by the first passage time of the process. Additionally, estimation of the first passage time allows analytic solutions for optimal trade entry and exit thresholds to be calculated. Another contribution of time series analysis is the possibility of a time-varying equilibrium relationship—the financial markets are variable and volatile, so any unified statistical arbitrage framework should be similarly adaptive to evolving market conditions.

Alternative approaches to statistical arbitrage include those contributions that suggest the possibility of mean-aversion among some select pairing of securities, in opposition to the traditional statistical arbitrage paradigm of mean-reversion trading. Additional evidence of mean-aversion is offered in the small but growing subset of literature incorporating regime switching behaviour. The presence of different regimes necessitates modification of the traditional trading rules, in which a position is taken in opposition to the sign of a significant deviation from the spread equilibrium; under some regime switching applications, trade might be suspended during different regimes, while the trading rule might be inverted under others. The flexibility of trading both mean-reverting and mean-averting spreads would be attractive for a unifying framework.

The first and most critical element that must be addressed by a unifying framework is that of spread non-convergence, identified by the literature as the greatest challenge to the continued profitability of statistical arbitrage. The distance and cointegration approaches consider securities that hold the relation $Y(t) \approx \beta X(t)$, with $\beta > 0$ and $\{X(t), Y(t) > 0 | t \geq 0\}$. Since these securities are considered adequate substitutes for each other, a trading opportunity is triggered when their hedged prices diverge significantly. This gives rise to the tradable spread,

$$U(t) = Y(t) - \beta X(t), \tag{3.1}$$

where the spread variable, $U(t)$, accounts for non-zero differences in the approximation, $Y(t) \approx \beta X(t)$. This spread variable is simple and intuitive, allowing the arbitrageur to identify trading opportunities, and buy or sell the related securities in proportion to the hedge ratio, β; that is, for every unit of $Y(t)$ bought, β units of $X(t)$ must be sold, and vice versa. While this spread variable is ubiquitous in statistical arbitrage literature for the construction of pairs trading portfolios, it suffers from a reliance on the hedge ratio remaining constant. If the arbitrage relationship between $X(t)$ and $Y(t)$ were to change or break down completely, β would no longer represent the appropriate hedge ratio between the two securities. The result would see the spread variable, $U(t)$, move away from its natural and assumed mean of zero, facilitating spread non-convergence.

A time-varying hedge ratio, $\beta(t)$, would more accurately model the evolution of the arbitrage relationship through time. Unfortunately, estimation of $\beta(t)$ is complicated. The Kalman filter is the simplest and most obvious choice for the estimation of $\beta(t)$, but it suffers from a kind of time-varying spurious regression—when the arbitrage relationship breaks down, the Kalman filter's estimates of $\beta(t)$ fluctuate wildly, making it difficult if not impossible for the arbitrageur to buy or sell the correct number of securities to preserve the hedge ratio between $X(t)$ and $Y(t)$.

An alternative formulation of the spread variable, proposed in this thesis as one of its chief contributions to the literature, offers a more intuitive representation of the arbitrage relationship. If $Y(t) \approx \beta X(t)$, then

$$\begin{aligned}\frac{Y(t)}{X(t)} &\approx \beta, \\ \ln\left(\frac{Y(t)}{X(t)}\right) &\approx \ln\beta, \\ V(t) &= \ln\left(\frac{Y(t)}{X(t)}\right) - \ln\beta, \end{aligned} \tag{3.2}$$

where the new spread variable, $V(t)$, again accounts for non-zero differences in the approximation. By framing the spread variable in this way, the logarithm of the hedge ratio, β, is considered the mean level of the observable log-ratio, $\ln(Y(t)/X(t))$, rather than the coefficient of the paired security, $X(t)$. Estimation of β is much simpler in $V(t)$ than in $U(t)$, allowing the consideration of its time-varying equivalent, $\beta(t)$. This simple reformulation of the tradable spread, $V(t)$, is what allows the distance, cointegration and time series approaches to be unified under a single model.

The principal contributions of this thesis detailed in Chapter 3 are summarised as follows: Section 3.2.1 proves the sub-optimality of the distance approach and its exclusion of potentially viable arbitrage opportunities; Section 3.2.2 presents the alternative spread formulation, $V(t)$; Section 3.2.3 proves the near-equivalence of the normalised z-scores of $U(t)$ and $V(t)$ for practical trading purposes, establishing $V(t)$ as a unifying spread formulation; Section 3.2.4 proposes the TVHR model to ensure model convergence before illustrating the expected time until failure when using a model with a constant hedge ratio, addressing the first research question posited in this thesis; Section 3.3 presents a simple discrete-time modelling and estimation procedure for the proposed TVHR model; Section 3.4 discusses the practical implementation of the proposed TVHR model along with other considerations relevant to the arbitrageur; Section 3.5 presents regime switching and statistical learning extensions of the proposed TVHR model, addressing the second and third research questions posited in this thesis; Section 3.6 presents a standardised procedure for the empirical evaluation of the distance, cointegration and TVHR models, and proposes a new cointegration approach under the unifying spread formulation that uses zero-crossing rate as a proxy for ADF testing.

3.2 Time-Varying Hedge Ratio Model

3.2.1 Distance and Cointegration Approach Bivariate Spread

The three primary approaches to statistical arbitrage, specifically the distance, cointegration and time series approaches, are intrinsically linked by their implicit model. The model requires a stochastic spread that is stationary and mean-reverting, for which the simplest and most tractable model is the OU process described by

$$\mathrm{d}S(t) = -\theta_S \left(S(t) - \mu_S\right) \mathrm{d}t + \sigma_S \, \mathrm{d}B_S(t), \tag{3.3}$$

for some security, $S(t)$. While the distance and cointegration approach implicitly assume a linear combination of candidate securities can be found such that the result is an OU process, the time series approach requires only that a security, artificial or otherwise, can be described by an OU process. The only distinguishing feature of the three approaches is the estimation of their respective model parameters.

Under the distance and cointegration approaches, the bivariate spread portfolio is represented by the stochastic variable $U(t)$, constructed by

$$U(t) = Y(t) - \beta X(t), \tag{3.4}$$

where $\beta = Y(0)/X(0)$ under the distance approach, and $\beta = \hat{\beta}$ under the cointegration approach. The distance approach value of β does not necessarily produce a zero-mean OU process in $U(t)$. Conversely, the cointegration approach uses OLS regression to estimate β in Equation (3.4) so that $U(t)$ is a zero-mean OU process when appropriate cointegrating securities, $X(t)$ and $Y(t)$, are identified. Both the distance and cointegration approaches seek to produce a stationary, mean-reverting spread, but while that spread may be shifted away from zero under the distance approach, the out-of-sample failure of cointegrating relationships will also shift $U(t)$ away from zero under the cointegration approach. With knowledge of this potential failure in mind, the choice of β is therefore arbitrary with respect to the out-of-sample evolution of $U(t)$.

The distance approach selects securities that minimise the SSD metric, $SSD\left(U_\tau\right) = \sum_\tau U_\tau^2$, where U_τ is the discrete-time analogue of $U(t)$, while the cointegration approach imposes no such requirement, requiring only that its securities are cointegrated. The limiting selection criteria imposed under the distance approach not only inhibits profitability, as noted by Krauss (2017), it also significantly restricts the universe of candidate securities. This can be seen by decomposing the SSD metric into its component parts; it is equivalent to the sum of the squared mean and variance of the spread, as shown by Krauss (2017). The continuous-time analogue of these components requires the mean and variance of the OU process, whose own evolution is described by the stochastic differential equation

$$\mathrm{d}U(t) = -\theta_U \left(U(t) - \mu_U\right) \mathrm{d}t + \sigma_U \, \mathrm{d}W_U(t), \tag{3.5}$$

$$= -\theta_U \left(Y(t) - \beta X(t) - \mu_U\right) \mathrm{d}t + \sigma_U \, \mathrm{d}W_U(t), \tag{3.6}$$

where $\theta_U \in (0, 2)$, $\sigma_U > 0$, and $\{W_U(t) | t \geq 0\}$ is a standard Brownian motion. The mean and variance of Equation (3.5) are given by

$$\mathbb{E}\left[U(t)\right] = U_0 e^{-\theta_U t} + \mu_U \left(1 - e^{-\theta_U t}\right),$$

$$\mathbb{V}\left[U(t)\right] = \frac{\sigma_U^2}{2\theta_U} \left(1 - e^{-2\theta_U t}\right).$$

As $t \to \infty$, the mean and variance of $U(t)$ tend toward $\mathbb{E}\left[U(t)\right] = \mu_U$ and $\mathbb{V}\left[U(t)\right] = \sigma_U^2 / 2\theta_U$. The continuous-time SSD metric, $SSD\left(U(t)\right)$, is therefore equivalent to the sum

$$SSD\left(U(t)\right) = \mathbb{E}\left[U(t)\right]^2 + \mathbb{V}\left[U(t)\right],$$

$$= \mu_U^2 + \frac{\sigma_U^2}{2\theta_U}. \tag{3.7}$$

If, during the formation period, the value of β produces an OU process whose mean, μ_U, is non-zero, $SSD\left(U(t)\right)$ will also have a non-zero bias, inflating the selection metric and likely leading the arbitrageur to discard the candidate securities. The estimation of β in-sample should therefore be made such that $\mathbb{E}\left[U(t)\right] = 0$ to allow consideration of the full universe of co-evolving securities, a feature of the cointegration approach which contributes to its superior performance over the distance approach. Assuming the arbitrageur were to select cointegrated series, $X(t)$ and $Y(t)$, and estimate β using OLS regression, the in-sample process mean would be zero, simplifying Equation (3.5) and Equation (3.6) to give

$$\mathrm{d}U(t) = -\theta_U U(t) \, \mathrm{d}t + \sigma_U \, \mathrm{d}W_U(t), \tag{3.8}$$

$$= -\theta_U \left(Y(t) - \beta X(t)\right) \mathrm{d}t + \sigma_U \, \mathrm{d}W_U(t), \tag{3.9}$$

negating the first term in Equation (3.7) and minimising $SSD\left(U(t)\right)$. However, the cointegration approach does not require the minimisation of the SSD metric, a constraint of the distance approach which simultaneously minimises the dispersion term, σ_U. The ideal pair of securities under the distance approach would therefore have null spread, and would consequently preclude all trading opportunities.

The form of OU process presented in Equation (3.9) has some interesting implications for the paired securities. Specifically, if $X(t)$ is assumed to be a Geometric Brownian Motion (GBM), it demonstrates that the evolution of $Y(t)$ is dependent on $X(t)$. If such a model were a true representation of the relationship, $Y(t)$ would therefore be a mean-reverting process relative to its non-stationary mean of $\beta X(t)$. Conceptually, $Y(t) = \beta X(t) + U(t)$ might be thought of as the superposition of two independent stochastic processes: a GBM, $X(t)$, and an OU process, $U(t)$. Though this simplifies the relationship between the paired securities, it also offers insights into the mechanics of the arbitrage relationship.

3.2.2 Time Series Approach Stochastic Spread

Both the distance and cointegration approach conform to the representation of the spread, $U(t)$, as the difference between the dependent security, $Y(t)$, and its independent non-stationary mean, $\beta X(t)$. Conventionally, the OU process used by the time series approach models the evolution of a stochastic variable about its long-term stationary mean. The stochastic differential equation in Equation (3.9), however, describes a non-standard OU process. An alternative representation of the spread variable, $U(t)$, is required in order to facilitate use of the standard OU process. If $Y(t) \approx \beta X(t)$, then

$$V(t) = \ln\left(\frac{Y(t)}{X(t)}\right) - \ln\beta, \tag{3.10}$$

is an alternative representation of the spread to $U(t)$ that offers the flexibility of being modelled as an OU process. This new stochastic spread, $\{V(t)|t \geq 0\}$, follows a similar evolution to $U(t)$, but represents its spread as the difference between a stochastic variable, $\ln(Y(t)/X(t))$, and its long-term mean, $\ln\beta$. Alternatively, the observed stochastic variable, $\ln(Y(t)/X(t))$, might be thought of as the superposition of a constant and an independent stochastic process, $\ln(Y(t)/X(t)) = \ln\beta + V(t)$. Its evolution follows the conventional OU process

$$\begin{aligned} \mathrm{d}V(t) &= -\theta_V\left(V(t) - \mu_V\right)\mathrm{d}t + \sigma_V\,\mathrm{d}W_V(t), \\ &= -\theta_V\left(\ln\left(\frac{Y(t)}{X(t)}\right) - \ln\beta - \mu_V\right)\mathrm{d}t + \sigma_V\,\mathrm{d}W_V(t), \end{aligned} \tag{3.11}$$

where $\theta_V \in (0,2)$, $\sigma_V > 0$, and $\{W_V(t)|t \geq 0\}$ is a standard Brownian motion. If $\mathbb{E}[V(t)] = \mu_V = 0$, then Equation (3.11) simplifies to

$$\mathrm{d}V(t) = -\theta_V\left(\ln\left(\frac{Y(t)}{X(t)}\right) - \ln\beta\right)\mathrm{d}t + \sigma_V\,\mathrm{d}W_V(t). \tag{3.12}$$

In this form, the relationship between the candidate securities is simplified—the arbitrageur need no longer estimate the cointegrating coefficient in Equation (3.4), which is susceptible to estimation difficulties, but rather the long-term mean of the observed log-ratio, $\ln\beta$. Once this mean level is estimated, it can be subtracted from the observed log-ratio to yield the zero-mean stationary OU process, $V(t)$. This reformulation of the spread process forms the basis of the unifying model proposed in this thesis to draw the distance, cointegration and time series approaches together under a single framework.

3.2.3 Equivalence of Spread Formulations

In order for $V(t)$ to be a viable alternative representation to $U(t)$ of the arbitrage relationship, their normalised z-scores and, consequently, their trade dynamics need to be equivalent. The arbitrageur ordinarily enters a position when $U(t)$ reaches some significant positive or negative value, signifying a statistical divergence of the security prices under consideration. If $V(t)$ were to offer the same executable opportunities as $U(t)$, their normalised values must be equal. That is,

$$\begin{aligned} z_U(t) &= z_V(t), \\ \frac{U(t) - \mathbb{E}[U]}{\sqrt{\mathbb{V}[U]}} &= \frac{V(t) - \mathbb{E}[V]}{\sqrt{\mathbb{V}[V]}}. \end{aligned}$$

The Euclidean distance between the normalised vectors, z_U, z_V, is given by $d(z_U, z_V) = \sqrt{2(1 - \rho_{z_U,z_V})}$, where ρ_{z_U,z_V} is the Pearson correlation coefficient between z_U and z_V. In order for the Euclidean distance to equal zero, and consequently for the statistical arbitrage z-scores to be equal to each other, Pearson correlation must equal one. It is therefore sufficient to show that $\rho_{U,V} = 1$, since

$$\begin{aligned} \rho_{z_U,z_V} &= \frac{\mathbb{E}\left[\left(z_U - \mathbb{E}[z_U]\right)\left(z_V - \mathbb{E}[z_V]\right)\right]}{\sqrt{\mathbb{V}[z_U]\mathbb{V}[z_V]}}, \\ &= \mathbb{E}[z_U z_V], \\ &= \mathbb{E}\left[\left(\frac{U - \mathbb{E}[U]}{\sqrt{\mathbb{V}[U]}}\right)\left(\frac{V - \mathbb{E}[V]}{\sqrt{\mathbb{V}[V]}}\right)\right], \\ &= \frac{\mathbb{E}\left[\left(U - \mathbb{E}[U]\right)\left(V - \mathbb{E}[V]\right)\right]}{\sqrt{\mathbb{V}[U]\mathbb{V}[V]}}, \\ &= \rho_{U,V}. \end{aligned}$$

If $U(t)$ and $V(t)$ are perfectly positively correlated, their z-scores are identical and both series generate the same trading signals for the arbitrageur's consideration.

Two series with equal rates of change are perfectly positively correlated. Consider, for example, $A(t)$ and $B(t)$, whose rates of relative change are identical. That is,

$$\begin{aligned}\frac{A(t+1)}{A(t)} &= \frac{B(t+1)}{B(t)},\\ A(t) &= A(0)\frac{A(t)}{A(0)},\\ &= A(0)\frac{B(t)}{B(0)},\\ &= B(t)\frac{A(0)}{B(0)},\\ &= cB(t),\end{aligned}$$

where c is a constant. The correlation between $A(t)$ and $B(t)$ is therefore

$$\begin{aligned}\rho_{A,B} &= \frac{\mathbb{E}\left[(A-\mathbb{E}[A])\,(B-\mathbb{E}[B])\right]}{\sqrt{\mathbb{V}[A]\mathbb{V}[B]}},\\ &= \frac{\mathbb{E}\left[(cB-\mathbb{E}[cB])\,(B-\mathbb{E}[B])\right]}{\sqrt{\mathbb{V}[cB]\mathbb{V}[B]}},\\ &= \frac{c\mathbb{E}\left[(B-\mathbb{E}[B])\,(B-\mathbb{E}[B])\right]}{\sqrt{c^2\mathbb{V}[B]\mathbb{V}[B]}},\\ &= \frac{c\mathbb{V}[B]}{\sqrt{c^2\mathbb{V}[B]^2}},\\ &= 1.\end{aligned}$$

The complex relationship between $U(t)$ and $V(t)$, due largely to the effect of the natural logarithms in $V(t)$ on the scale of the process, means that $U(t)$ and $V(t)$ are not perfectly positively correlated. Their almost-perfect positive correlation, however, can be witnessed in the approximate equivalence between the relative change of $U(t)$ and that of $V(t)$. That is,

$$\begin{aligned}\frac{U(t+1)}{U(t)} &\approx \frac{V(t+1)}{V(t)},\\ \frac{Y(t+1)-\beta X(t+1)}{Y(t)-\beta X(t)} &\approx \frac{\ln\left(\frac{Y(t+1)}{\beta X(t+1)}\right)}{\ln\left(\frac{Y(t)}{\beta X(t)}\right)}.\end{aligned} \tag{3.13}$$

Equation (3.13) is analogous to the approximate equivalence of percentage price changes and log-price changes discussed by Fama (1965), in which the two are considered appropriate substitutes for each other given the limited variability observed in security returns; the closer to zero, the greater the approximation. Taking a first-order Taylor series approximation of the natural logarithms on the right-hand side of Equation (3.13) gives

$$
\begin{aligned}
\frac{\ln\left(\frac{Y(t+1)}{\beta X(t+1)}\right)}{\ln\left(\frac{Y(t)}{\beta X(t)}\right)} &\approx \frac{\left(\frac{Y(t+1)}{\beta X(t+1)}-1\right)}{\left(\frac{Y(t)}{\beta X(t)}-1\right)}, \\
&= \frac{\left(\frac{Y(t+1)-\beta X(t+1)}{\beta X(t+1)}\right)}{\left(\frac{Y(t)-\beta X(t)}{\beta X(t)}\right)}, \\
&= \frac{X(t)}{X(t+1)}\frac{Y(t+1)-\beta X(t+1)}{Y(t)-\beta X(t)}, \\
&= \frac{X(t)}{X(t+1)}\frac{U(t+1)}{U(t)}. \qquad (3.14)
\end{aligned}
$$

As $X(t+1)-X(t) \to 0$, the rate of change of $V(t)$ approaches the rate of change of $U(t)$, and the two processes are highly correlated.

The conditional dependence of high positive correlation on small changes in $X(t)$ is reflected in the requirements of perfect correlation. In order for $U(t)$ and $V(t)$ to be perfectly positively correlated, there must exist an affine transformation of $V(t)$, such that

$$
\begin{aligned}
V(t) &= \ln\left(\frac{\beta X(t)+U(t)}{\beta X(t)}\right), \\
&= a + bU(t),
\end{aligned}
$$

where $a \in \mathbb{R}$ and $b > 0$. Solving for $X(t)$ gives

$$
X(t) = \frac{U(t)}{e^{(a+bU(t))}-1},
$$

which is inconsistent with its specification as a GBM. However, assuming only small changes, $X(t+1)-X(t) \to 0$, such that $X(t)$ is approximately constant at a value of μ_X with $X(t) \approx \mu_X \gg U(t)$, then

$$
\begin{aligned}
V(t) &= \ln\left(\frac{\beta X(t)+U(t)}{\beta X(t)}\right), \\
&\approx \frac{U(t)}{\beta X(t)}, \\
&\approx \frac{1}{\beta\mu_X}U(t),
\end{aligned}
$$

which is affine with $a = 0$ and $b = 1/\beta\mu_X$, where the first approximation is a first-order Taylor series expansion of $V(t)$. Here again, small changes in $X(t)$ lead to almost-perfect positive correlation of $U(t)$ and $V(t)$.

A bootstrapping simulation of $U(t)$ and $V(t)$ further illustrates the correlation between the two processes. Each simulation generates 1,000 samples of two GBMs, $X(t)$ and $Y(t)$,

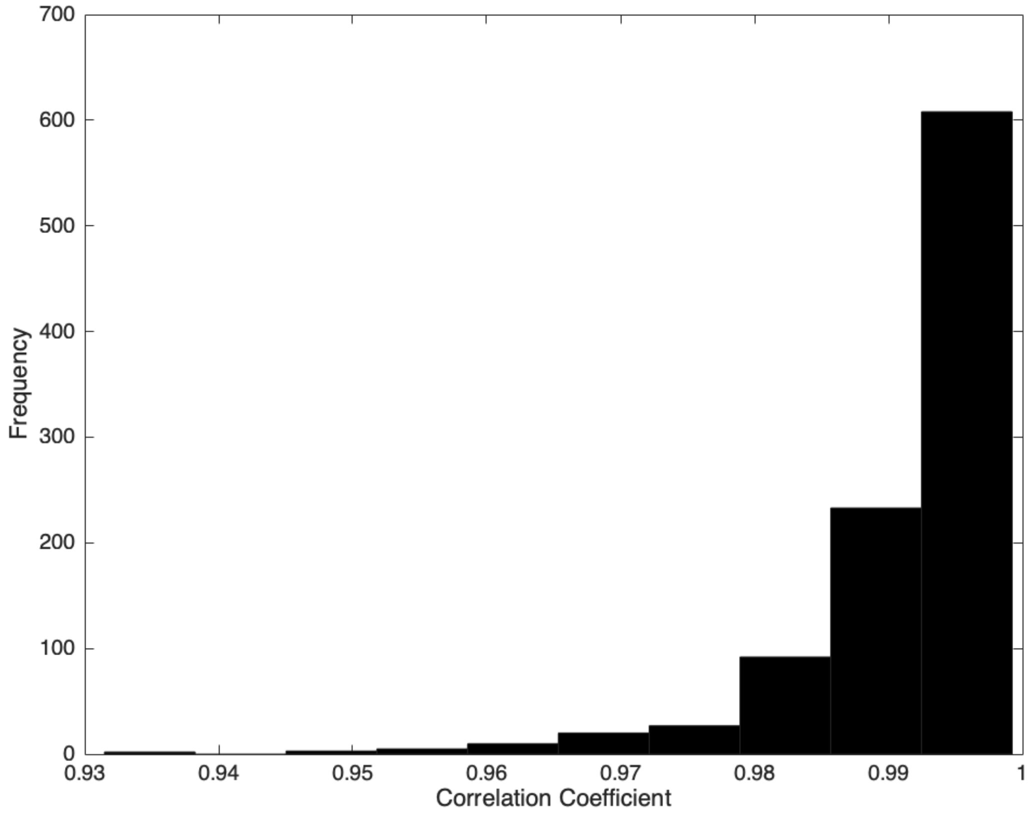

Figure 3.1: Histogram of correlation coefficients between simulated stochastic spread series, $U(t)$ and $V(t)$.

and estimates their regression coefficients, β, using OLS regression to minimise the sum of squared values of $U(t)$ in Equation (3.4). $U(t)$ and $V(t)$ are then constructed and their correlation calculated. Figure 3.2 offers a visualisation of $X(t)$ and $Y(t)$ for a single simulation in the top panel, and the constructed normalised z-scores of $U(t)$ and $V(t)$ in the bottom panel. Each simulation is repeated 1,000 times to generate a distribution of correlation coefficients, the histogram of which is displayed in Figure 3.1. Its mean and median are 0.9922 and 0.9943, respectively.

Given the near-perfect positive correlation of $U(t)$ and $V(t)$, it can be concluded that $V(t)$ is an adequate substitute for $U(t)$, and that the two series offer the same arbitrage opportunities. The distance, cointegration and time series approaches to statistical arbitrage are therefore identical in terms of their implicit model—that is, an OU process describing the evolution of a stochastic variable about its long-term stationary mean. The arbitrage portfolio would still be constructed in the exact same way under $V(t)$, with β units of $X(t)$ short-sold for every unit of $Y(t)$ bought and vice versa, so there is no difference in how the opportunity is exploited. The consideration of the OU process governing $V(t)$ presented in Equation (3.12), however, allows for closer scrutiny of the failure of statistical

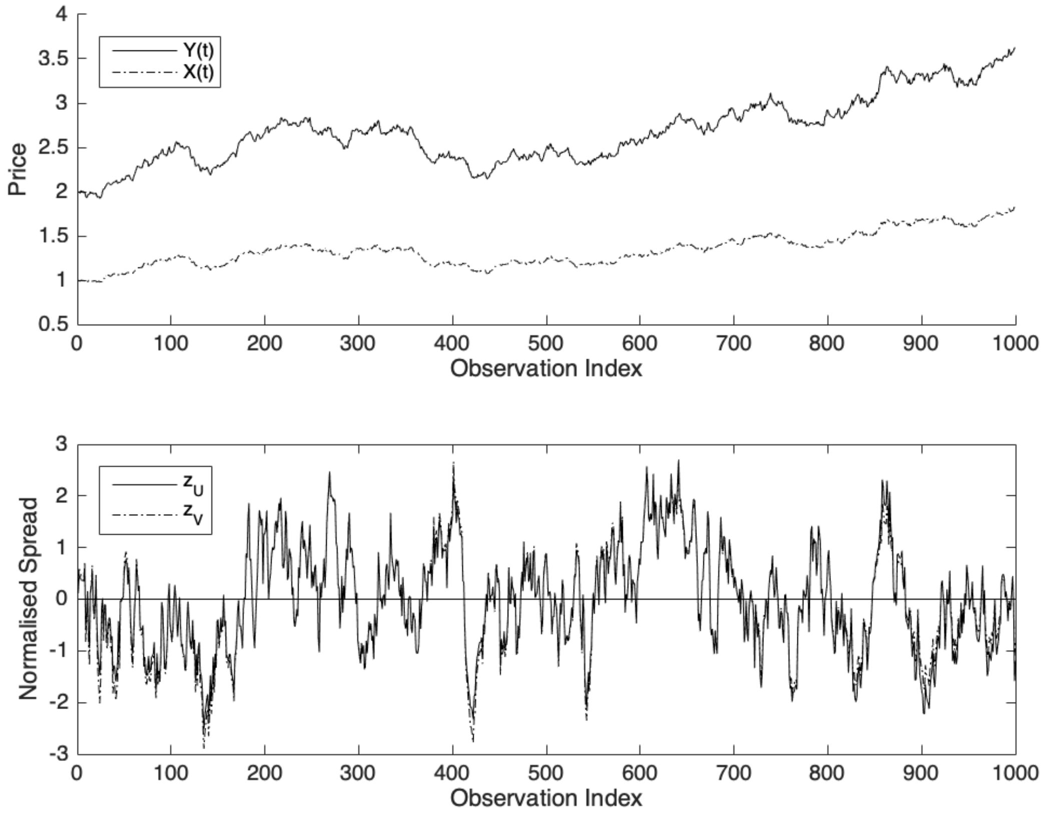

Figure 3.2: A single simulation of Geometric Brownian Motions in the top panel, $X(t)$ and $Y(t)$, and the normalised z-scores of $U(t)$ and $V(t)$ constructed from $X(t)$ and $Y(t)$ in the bottom panel.

arbitrage opportunities and the rising prevalence of non-convergence documented in the literature, addressing the first research question posited in this thesis.

3.2.4 Proposed Model and Model Assumptions

The implicit statistical arbitrage model of a conventional OU process is one which is inconsistent with the empirical literature. The model, as expressed by Equation (3.2) and shown in Section 3.2.3 to be equivalent to the conventional distance and cointegration models expressed by Equation (3.1), establishes a mean-reverting spread from the difference between a stochastic variable and its long-term mean, assuming that mean is time-invariant. The distance approach normalises the candidate securities' prices at the beginning of formation and trading periods, such that the normalisation is applied for the duration of the period. Similarly, the cointegration approach requires the estimation of a cointegrating coefficient that remains fixed for the duration of the formation and trading periods. The time series approach, despite its theoretical purity relative to the distance and cointegration approaches, also assumes a constant mean that is estimated from the data.

Statistical arbitrage non-convergence presents the greatest challenge to the continued profitability of the strategy, as discussed in Section 2.1.5. The literature indicates that the cause of non-convergence is idiosyncratic shocks experienced by one or both of the candidate securities, causing either persisting divergence or a complete disruption and breakdown of the arbitrage relationship. In the case of persisting divergence, the mean of the OU process describing the spread will shift significantly from its historically-observed level. The stationarity and mean-reversion of the spread will be preserved under such circumstances, albeit with respect to a different mean; the process will have jumped to a new level away from $\ln\beta$ in Equation (3.12). In the event of complete failure of the arbitrage relationship, the mean of the OU process will assume the characteristics of a random walk. In both cases, the assumption of a static mean is erroneous.

The evolution of the observed ratio, $S(t) = \ln(Y(t)/X(t))$, can be conceptualised as the superposition of two independent stochastic processes; a time-varying mean, $M(t) = \ln\beta + \sigma_M B_M(t)$, and a stationary OU process, $V(t)$, such that $S(t) = M(t) + V(t)$. If their evolutions follow

$$\begin{aligned} \mathrm{d}M(t) &= \sigma_M \,\mathrm{d}B_M(t), \\ \mathrm{d}V(t) &= -\theta_V V(t)\,\mathrm{d}t + \sigma_V \,\mathrm{d}B_V(t), \end{aligned}$$

then the evolution of the sum, $S(t)$, follows

$$\mathrm{d}S(t) = -\theta_V \left(S(t) - M(t)\right)\mathrm{d}t + \sigma_M \,\mathrm{d}B_M(t) + \sigma_V \,\mathrm{d}B_V(t), \tag{3.15}$$

where $\{B_M, B_V | t \geq 0\}$ are independent standard Brownian motions. When $\sigma_M = 0$, $S(t)$ is an OU process that reverts to a time-invariant mean. The securities, $Y(t)$ and $X(t)$, that constitute the ratio, $S(t)$, would be perfect candidates for statistical arbitrage thanks to their static cointegrating relationship. When $\sigma_M > 0$, however, the cointegrating relationship is not constant, and the mean of the process will eventually shift away from $\ln\beta$. If the arbitrageur fails to estimate the time-varying mean, the stationary spread, $V(t)$, will shift away from zero facilitating non-convergence of the spread. The assumption of a constant mean is therefore inconsistent with the salient features of the empirical data.

The proposed Time-Varying Hedge Ratio (TVHR) model comprises the time-varying relationship,

$$\begin{aligned} V(t) &= S(t) - M(t), \\ &= \ln\left(\frac{Y(t)}{X(t)}\right) - \left(\ln\beta + \sigma_M B_M(t)\right), \end{aligned} \tag{3.16}$$

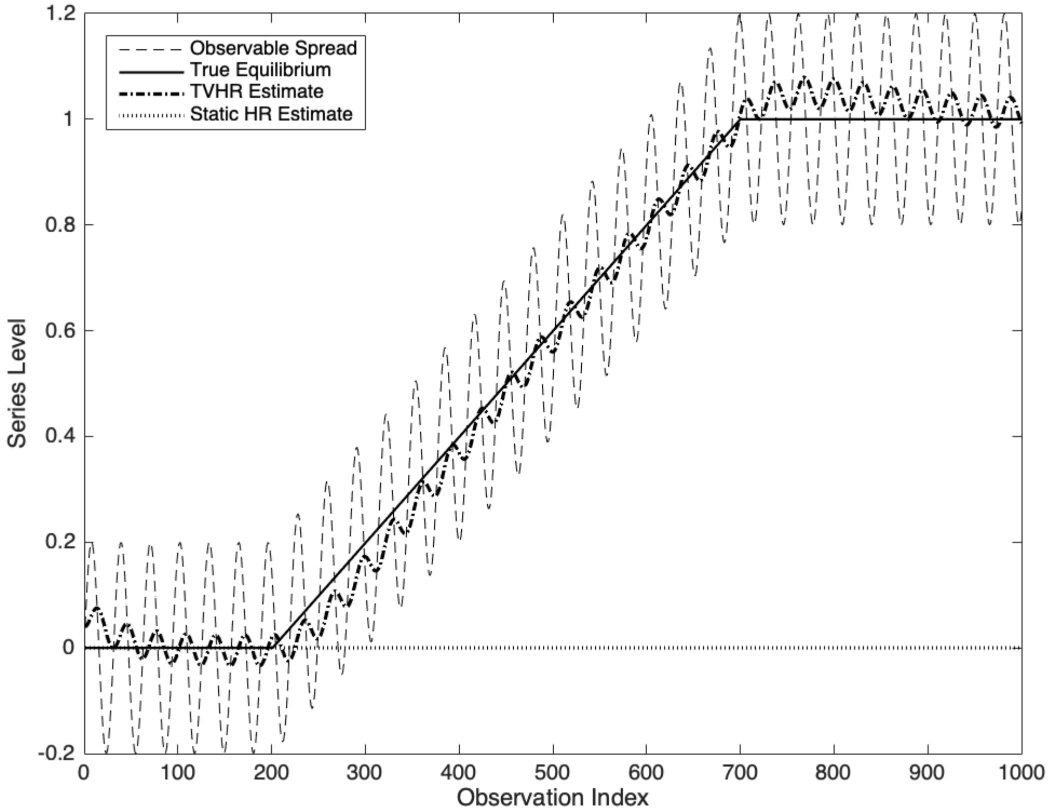

Figure 3.3: Illustration of the TVHR model relative to a static hedge ratio in the estimation of the unobservable process equilibrium.

where the second summand represents the time-varying hedge ratio. The second summand can be reparameterised as $\beta(t)$, allowing the final form for the TVHR model to be expressed as

$$V(t) = \ln\left(\frac{Y(t)}{X(t)}\right) - \ln\beta(t), \tag{3.17}$$

whose evolution follows the standard OU process,

$$\mathrm{d}V(t) = -\theta_V V(t)\,\mathrm{d}t + \sigma_V\,\mathrm{d}B_V(t). \tag{3.18}$$

In this form, $\ln\beta(t)$ can be considered the time-varying mean of the observable log-ratio, $S(t) = \ln(Y(t)/X(t))$. Unlike conventional approaches that assume a fixed hedge ratio throughout the life of the arbitrage opportunity, the TVHR model allows the hedge ratio to evolve with the market. An illustration of the TVHR model's adaptivity is shown in Figure 3.3, relative to an observable spread, $S(t)$, that shifts to a new level after the first 200 observations. The true equilibrium level is an unobservable variable which must be estimated by the arbitrageur. The TVHR model estimate contrasts the static hedge ratio estimate that does not change throughout the observation period.

The variance of the superposed stochastic processes, $S(t)$, is unbounded when $\sigma_M > 0$, facilitating non-convergence of the statistical arbitrage opportunity, and the failure of the strategy. The variance can be calculated as

$$\begin{aligned} \mathbb{V}\left[S(t)\right] &= \mathbb{V}\left[M(t) + V(t)\right], \\ &= \mathbb{V}\left[M(t)\right] + \mathbb{V}\left[V(t)\right], \\ &= \sigma_M^2 t + \frac{\sigma_V^2}{2\theta_V}\left(1 - e^{-2\theta_V t}\right), \end{aligned} \tag{3.19}$$

whose limit as $t \to \infty$ is infinite. An arbitrageur may mistakenly believe that candidate securities constituting a statistical arbitrage opportunity have produced an OU process with null dispersion in the time-varying mean, $\sigma_M = 0$, when in fact there is some non-zero dispersion, $\sigma_M > 0$. A sufficiently small value of σ_M will give the appearance of stationarity in $S(t)$, but for the arbitrageur who assumes a time-invariant mean, the process, $S(t)$, is no better than a random walk—it will eventually shift away from its long-term mean, and the arbitrageur will be susceptible to losses resulting from non-convergence of the spread. Spread non-convergence is visualised in Figure 3.4, comparing the spread of the TVHR model in the top panel to the spread of a conventional approach with a static hedge ratio. While the TVHR model initially experiences a slight disruption to the spread upon a change in the data generating process, the conventional static hedge ratio spread experiences a prolonged divergence from which it is unable to recover.

To understand the risk posed by spread non-convergence, we must find the time at which $\mathbb{V}\left[M(t)\right] \geq \mathbb{V}\left[V(t)\right]$, when the mean process will have potentially shifted the stationary process sufficiently away from zero such that the variance of the stationary process is not likely to facilitate convergence. We therefore have

$$\begin{aligned} \mathbb{V}\left[M(t)\right] &\geq \mathbb{V}\left[V(t)\right], \\ \sigma_M^2 t &\geq \frac{\sigma_V^2}{2\theta_V}\left(1 - e^{-2\theta_V t}\right), \\ 0 &\leq \sigma_M^2 t - \frac{\sigma_V^2}{2\theta_V}\left(1 - e^{-2\theta_V t}\right), \\ 0 &\leq 2\theta_V \frac{\sigma_M^2}{\sigma_V^2} t - \left(1 - e^{-2\theta_V t}\right), \\ 0 &\leq \alpha\gamma t - \left(1 - e^{-\alpha t}\right), \end{aligned}$$

with $\alpha = 2\theta_V$ and $\gamma = \sigma_M^2/\sigma_V^2$, which is a non-linear function in t that must be solved numerically to find the smallest value of t for which the inequality holds. Since $\theta_V \in (0, 2)$, then $\alpha \in (0, 4)$, however $\theta_V \in [1, 2)$ corresponds with OU processes whose reversion is so strong as to be indistinguishable from white noise processes. It is unlikely that an arbitrageur would be able to capitalise on opportunities presented by such a process, so more realistic mean-reversion rates of $\theta_V \in (0, 1)$ and, consequently, $\alpha \in (0, 2)$ are

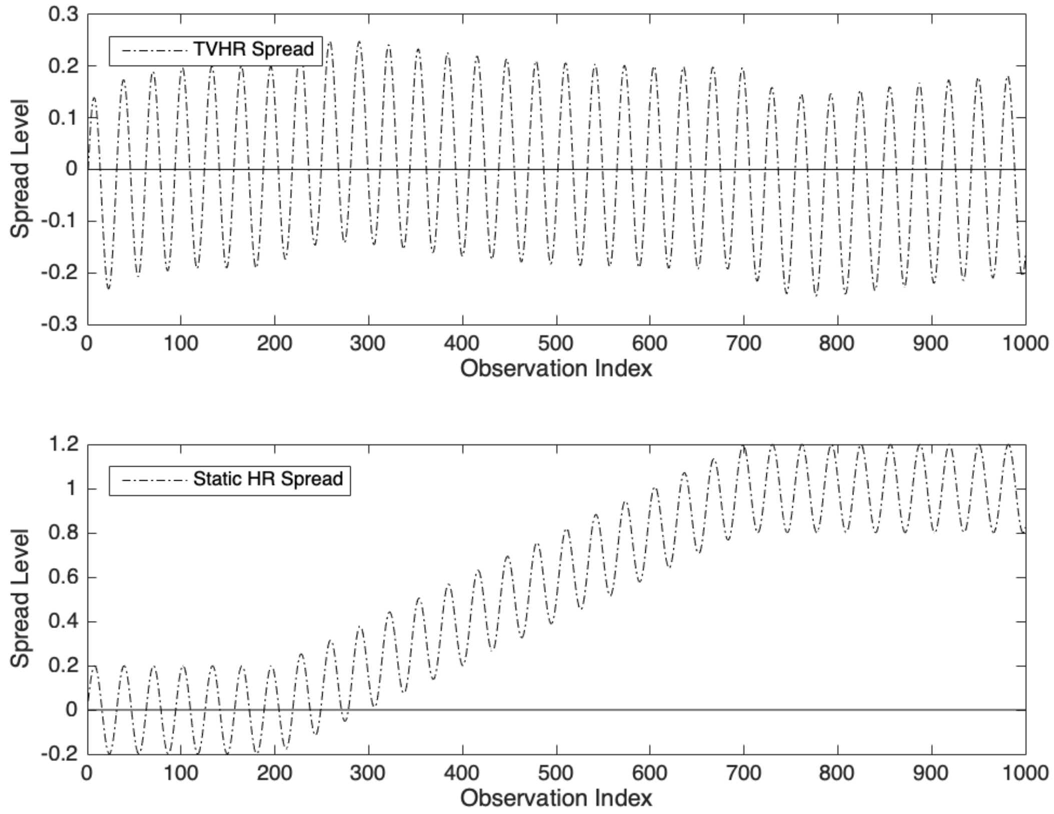

Figure 3.4: Comparison between the spread calculated using the TVHR estimate of the process equilibrium and that calculated using a static hedge ratio estimate, illustrating non-convergence of the latter.

assumed. If $\gamma \geq 1$, there is no time, $t > 0$, for which the inequality is untrue; if the noise term for the mean, $M(t)$, is greater than the noise term for the stationary process, $V(t)$, then the variance of the mean will immediately be greater than the variance of the stationary process. Any statistical arbitrage opportunities with $\gamma \geq 1$ would likely fail within a short time from the commencement of trading. Estimated values of t for a range of α and γ values are shown in Table 3.1. A surface plot of times is shown in Figure 3.5, with values of $t < 1$, for which the arbitrageur would not be able to act on the divergence, reassigned a value of $t = -100$ for visual clarity.

The time, t, at which $\mathbb{V}[M(t)] \geq \mathbb{V}[V(t)]$ is the time that a statistical arbitrage opportunity could be expected to fail due to non-convergence. The failure of the opportunity is due to the estimation and use of a fixed mean, $\ln \beta$, which fails to capture the non-stationary nature of the true process mean, $M(t)$. The observed variable, $S(t)$, moves away from $\ln \beta$ and, consequently, $V(t)$ moves away from its desired mean of zero. Unable to determine the true magnitude of divergence or convergence, the arbitrageur must abandon the opportunity at the end of the trading period, likely accruing losses due

		α				
		0.1000	0.5500	1.0000	1.4500	1.9000
	0.1000	99.9955	18.1810	9.9995	6.8962	5.2629
	0.3000	31.9706	5.8128	3.1971	2.2049	1.6827
γ	0.5000	15.9362	2.8975	1.5936	1.0991	0.8387
	0.7000	7.6143	1.3844	0.7614	0.5251	0.4008
	0.9000	2.1456	0.3901	0.2146	0.1480	0.1129

Table 3.1: Estimated time, t, at which $\mathbb{V}[M(t)] \geq \mathbb{V}[V(t)]$ for different combinations of α and γ.

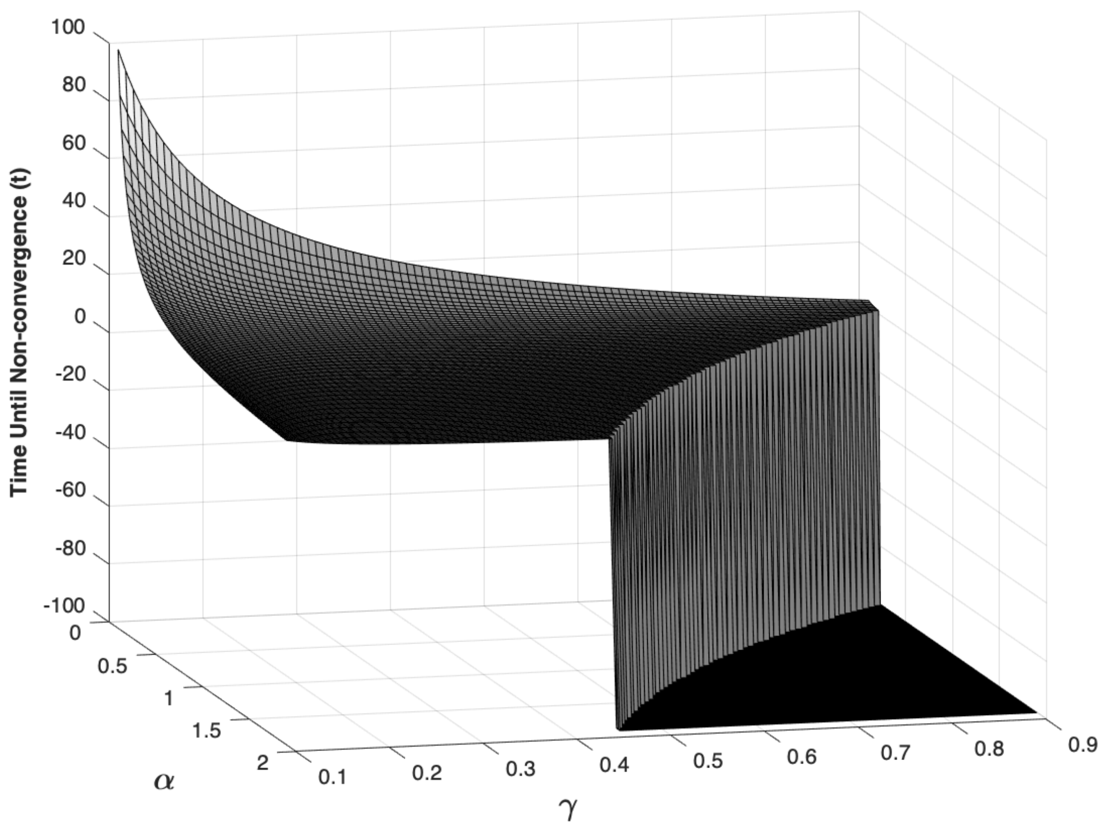

Figure 3.5: Estimated time, t, at which $\mathbb{V}[M(t)] \geq \mathbb{V}[V(t)]$ for different combinations of α and γ.

to non-convergence.

An OU process with time-varying mean, such as that expressed by Equation (3.17), is more appropriate for statistical arbitrage than the implicit model used by the distance, cointegration and time series approaches expressed by Equation (3.13). Such a model accounts for perturbations in the process mean, $M(t)$, maintaining a zero-mean stationary process, $V(t)$, for consideration of trading opportunities by the arbitrageur. It is this TVHR model that is proposed as a solution for the problem of spread non-convergence and, consequently, the declining profitability of statistical arbitrage, addressing the first research question posited in this thesis.

3.3 Discrete Model Derivation

The dynamic equilibrium level, $M(t)$, of the artificial security's price, $S(t)$, can be estimated in discrete-time by the Kalman filter. The state-space representation of the estimation procedure is given by

$$\begin{aligned} S_\tau &= M_\tau + V_\tau, \\ M_\tau &= M_{\tau-1} + \sigma_M \omega_\tau, \end{aligned}$$

where $\omega_\tau \sim \mathcal{N}(0,1)$ is Gaussian noise. Under this random walk-plus-noise model, the Kalman filter estimate of M_τ is algebraically equivalent to an Exponential Moving Average (EMA) estimate of S_τ, whose smoothing parameter is uniquely determined by the Kalman filter's noise covariance statistics (Bruder, Dao, Richard, and Roncalli, 2011). The smoothing parameter, $\alpha_0 \in (0,1)$, determines the amount of history used in the estimation of the current hidden state, M_τ; a value of α_0 close to one considers mostly recent history, while a value close to zero smoothes the estimate over a greater length of time.

The EMA estimate of M_τ is given by

$$\hat{M}_\tau = \alpha_0 S_\tau + (1-\alpha_0)\,\hat{M}_{\tau-1}. \tag{3.20}$$

Let $\delta_\tau = S_\tau - \hat{M}_\tau$ be the difference between the artificial security and the EMA estimate of its mean. Substituting in Equation (3.20) gives

$$\begin{aligned} S_\tau &= \hat{M}_\tau + \delta_\tau, \\ &= \alpha_0 S_\tau + (1-\alpha_0)\,\hat{M}_{\tau-1} + \delta_\tau, \end{aligned} \tag{3.21}$$

where the EMA can be initialised as $\hat{M}_0 = S_0$. The objective is now to rearrange Equation (3.21) to show that the EMA estimation procedure yields a stationary, mean-reverting process, V_τ.

Rearranging Equation (3.21) gives

$$\begin{aligned}\delta_\tau &= (1-\alpha_0)\left(S_\tau - \hat{M}_{\tau-1}\right),\\ &= (1-\alpha_0)\left(S_\tau - S_{\tau-1} + \delta_{\tau-1}\right),\\ &= (1-\alpha_0)\,\delta_{\tau-1} + (1-\alpha_0)\left(S_\tau - S_{\tau-1}\right). \end{aligned} \tag{3.22}$$

Let $S_\tau - S_{\tau-1} = R_{S,\tau} \sim \mathcal{N}\left(\mu_{S,\tau}, \sigma^2_{S,\tau}\right)$, be the one-period return of the artificial security, S_τ. This formulation allows the mean and variance of the artificial security's returns to vary over time. Substituting into Equation (3.22) gives

$$\begin{aligned}\delta_\tau &= (1-\alpha_0)\,\delta_{\tau-1} + (1-\alpha_0)\,R_{S,\tau},\\ &= c_\tau + (1-\alpha_0)\,\delta_{\tau-1}, \end{aligned} \tag{3.23}$$

where $c_\tau = (1-\alpha_0)\,R_{S,\tau}$. Let $\mathbb{E}\left[\delta_\tau\right] = \mu_{\delta,\tau}$, and take expectations of Equation (3.23), giving

$$\begin{aligned}\mathbb{E}\left[\delta_\tau\right] &= \mathbb{E}\left[c_\tau + (1-\alpha_0)\,\delta_{\tau-1}\right],\\ \mu_{\delta,\tau} &= c_\tau + (1-\alpha_0)\,\mu_{\delta,\tau-1}. \end{aligned}$$

Rearranging for c_τ, we have $c_\tau = \mu_{\delta,\tau} - (1-\alpha_0)\,\mu_{\delta,\tau-1}$. Substituting into Equation (3.23) gives

$$\begin{aligned}\delta_\tau &= \mu_{\delta,\tau} - (1-\alpha_0)\,\mu_{\delta,\tau-1} + (1-\alpha_0)\,\delta_{\tau-1},\\ \delta_\tau - \mu_{\delta,\tau} &= (1-\alpha_0)\left(\delta_{\tau-1} - \mu_{\delta,\tau-1}\right). \end{aligned} \tag{3.24}$$

The process mean, $\mu_{\delta,\tau}$, can be conceptualised as a bias that shifts δ_τ away from its natural zero-mean level. The bias can be estimated by an additional EMA with smoothing parameter, $\alpha_1 \in (0,1)$, giving

$$\hat{\mu}_{\delta,\tau} = \alpha_1\delta_\tau + (1-\alpha_1)\,\hat{\mu}_{\delta,\tau-1}. \tag{3.25}$$

Let $\mu_{\delta,\tau} = \hat{\mu}_{\delta,\tau} + e_\tau$, and substitute into Equation (3.25), giving

$$\begin{aligned}\mu_{\delta,\tau} - e_\tau &= \alpha_0\delta_\tau + (1-\alpha_0)\left(\mu_{\delta,\tau-1} - e_{\tau-1}\right),\\ \mu_{\delta,\tau} - \alpha_0\delta_\tau - (1-\alpha_0)\,\mu_{\delta,\tau-1} &= e_\tau - (1-\alpha_0)\,e_{\tau-1}, \end{aligned} \tag{3.26}$$

where the left-hand side is the difference between the true process bias and its EMA estimate. The left-hand side therefore constitutes a zero-mean white noise process, giving

$$\begin{aligned}\mu_{\delta,\tau} - \alpha_0\delta_\tau - (1-\alpha_0)\,\mu_{\delta,\tau-1} &= e_\tau - (1-\alpha_0)\,e_{\tau-1},\\ &= \sigma_{\mu,\tau}\nu_\tau, \end{aligned} \tag{3.27}$$

where $\sigma_{\mu,\tau} > 0$ and $\nu_\tau \sim \mathcal{N}(0,1)$ is Gaussian noise. Let $V_\tau = \delta_\tau - \mu_{\delta,\tau}$. Substituting $\mu_{\delta,\tau} = \hat{\mu}_{\delta,\tau} + e_\tau$ into Equation (3.24), we have

$$\begin{aligned}
\delta_\tau - \hat{\mu}_{\delta,\tau} &= (1-\alpha_0)\left(\delta_{\tau-1} - \hat{\mu}_{\delta,\tau-1}\right), \\
\delta_\tau - \mu_{\delta,\tau} + e_\tau &= (1-\alpha_0)\left(\delta_{\tau-1} - \mu_{\delta,\tau-1} + e_{\tau-1}\right), \\
V_\tau &= (1-\alpha_0)\, V_{\tau-1} - e_\tau + (1-\alpha_0)\, e_{\tau-1}, \\
V_\tau &= (1-\alpha_0)\, V_{\tau-1} - \sigma_{\mu,\tau}\nu_\tau, \qquad (3.28)
\end{aligned}$$

which is an AR(1) process in V_τ.

Considering the OU process in Equation (3.18), discretisation of the equation gives

$$V_\tau = (1-\theta_V)\, V_{\tau-1} + \sigma_V \epsilon_\tau, \qquad (3.29)$$

which is also an AR(1) process in V_τ. The equivalence between Equation (3.28) and Equation (3.29) when $\alpha_0 = \theta_V$ and $-\sigma_{\mu,\tau}\nu_\tau = \sigma_V \epsilon_\tau$ shows that the EMA procedure for estimating the dynamic hidden state of the artificial security's price naturally produces a stationary, mean-reverting series whose dynamics are known *a priori*, following correction of the EMA residual to account for return bias. Equivalently, the true hidden state of the stochastic process

$$\begin{aligned}
S_\tau &= \hat{M}_\tau + \delta_\tau, \\
&= \hat{M}_\tau + \mu_{\delta,\tau} + V_\tau,
\end{aligned}$$

is found to be $M_\tau = \hat{M}_\tau + \mu_{\delta,\tau}$—the EMA estimate of the mean requires a bias correction in order to reflect the true process mean. The modelling procedure first estimates the hidden process mean, $\hat{M}_\tau$, and its bias correction, $\mu_{\delta,\tau}$, subtracting both from the observable spread, S_τ, to derive the tradable spread process, V_τ. Figure 3.6 illustrates the difference between a TVHR estimate of the mean with and without correcting for the bias, $\mu_{\delta,\tau}$.

The model presented in Section 3.2.4 assumes that there is a hidden, unobservable process driving the evolution of the artificial security price, around which a stationary, mean-reverting process oscillates. The inability to directly observe the hidden dynamic mean necessitates the Kalman filter, or rather, EMA estimation procedure to approximate the hidden state. The inability of the estimation procedure to fully account for the hidden dynamics manifests in the bias, $\mu_{\delta,\tau}$, which must be corrected in order for the spread between observed and predicted security price to retain its model-enforced OU process behaviour. Knowledge of the true hidden state of the security price would make the bias correction unnecessary.

The bias term, $\mu_{\delta,\tau}$, is a function of the time-varying returns of the artificial security, $R_{S,\tau}$, thanks to their incorporation in the recursions of the difference series, $\delta_\tau = S_\tau - \hat{M}_\tau$.

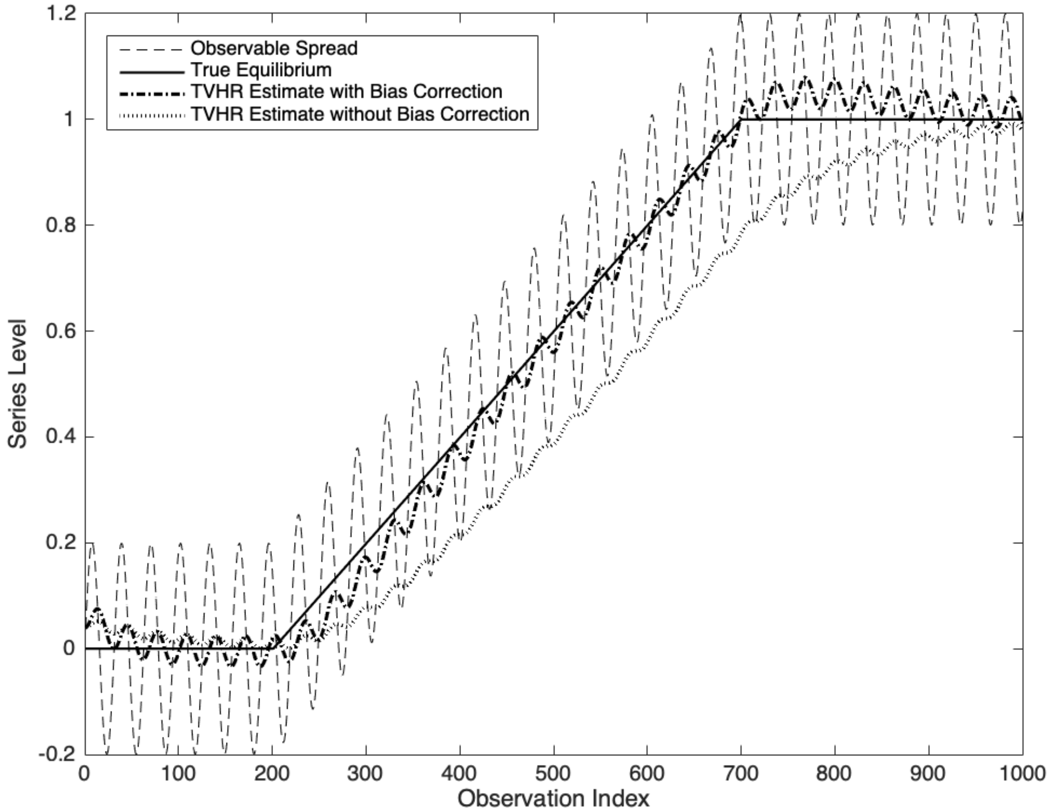

Figure 3.6: Comparison between the TVHR estimate of the process equilibrium with and without bias correction.

Since the artificial security price is defined as the natural logarithm of the ratio of two security prices, $\ln(Y_\tau / X_\tau)$, the bias is a natural consequence of the relative returns of the two securities constituting the bivariate statistical arbitrage portfolio. Rearranging the artificial security return gives

$$
\begin{aligned}
R_{S,\tau} &= S_\tau - S_{\tau-1}, \\
&= \ln\left(\frac{Y_\tau}{X_\tau}\right) - \ln\left(\frac{Y_{\tau-1}}{X_{\tau-1}}\right), \\
&= \ln\left(\frac{Y_\tau}{Y_{\tau-1}}\right) - \ln\left(\frac{X_\tau}{X_{\tau-1}}\right), \\
&= R_{Y,\tau} - R_{X,\tau},
\end{aligned} \tag{3.30}
$$

where $R_{Y,\tau}$ and $R_{X,\tau}$ are the observed log returns of securities Y_τ and X_τ, respectively. The bias exists only if, on average, one security's returns are greater than the other's. This result lends itself naturally to the concept of regimes, in which the returns of the securities vary over time according to the prevailing regime. If the two securities' returns were approximately equal across all time, then some EMA smoothing parameter, α_0, could be found such that the bias $\mu_{\delta,\tau} = 0, \ \forall \tau$.

If the bias term were constant and significantly different from zero—that is, if one security's mean return was greater or less than the other security's mean return by some constant amount across all time—then the statistical arbitrage opportunity would not be discovered by either of the distance or cointegration approaches. The SSD statistic used by the distance approach would deem the bivariate pair unsuitable for statistical arbitrage, owing to the pair's tendency to diverge over time. Similarly, a cointegration test would indicate a lack of stationarity in the residual of a cointegrating regression of the two securities' prices. The only condition that would indicate a statistical arbitrage opportunity under the conventional approaches would be a bias term equal to zero across all time. Such a condition would indicate that, on average, the return of the two securities were more or less equal, satisfying the requirement of a common stochastic trend.

Conversely, the modelling and estimation methodology presented in Sections 3.2 and 3.3 implicitly accounts for mean-averting behaviour of the dynamic spread by correcting for the bias attributable to mean-aversion. This mean-aversion shifts the equilibrium level of the artificial security's price, M_τ, but does not interfere with the superposed stationary, mean-reverting AR(1) process, V_τ. Consequently, standard statistical arbitrage trading can proceed in the presence of mean-aversion under the proposed methodology. Such mean-aversion is illustrated in Figure 3.6 between observations 200 and 700.

3.4 Practical Implementation

The discrete statistical arbitrage model presented in Section 3.3 models the difference between an artificial security price, S_τ, and a bias-corrected estimate of its equilibrium price, M_τ, in terms of a stationary, mean-reverting AR(1) process, V_τ. This result validates the model statistically, though there are a number of practical considerations that must be made in order for the model to be used for statistical arbitrage trading. The first concerns the generation of trading signals, and follows the conventional literature by specifying critical magnitudes of deviation which trigger trading opportunities.

Distance, cointegration and time series approaches to statistical arbitrage propose initiating trades once the spread between an artificial security and its equilibrium price has reached some significant level. A significant deviation can be quantified by normalising the spread with respect to its standard deviation or, more appropriately, its time-varying standard deviation. The bias-corrected spread, $V_\tau = S_\tau - \hat{M}_\tau - \hat{\mu}_{\delta,\tau}$, is assumed to be a zero-mean Gaussian random variable following a mean-reverting process, with some time-varying standard deviation, $\sigma_{V,\tau}$. The true time-varying standard deviation can be estimated using the Kalman filter, with the absolute deviation, $|V_\tau|$, serving as a proxy for the observable process whose hidden state at time τ can be estimated. In this random walk-plus-noise model, the Kalman filter is equivalent to an EMA, and so the standard deviation can be modelled as

$$\hat{\sigma}_{V,\tau} = \alpha_2 |V_\tau| + (1-\alpha_2)\,\hat{\sigma}_{V,\tau-1}, \tag{3.31}$$

with smoothing parameter $\alpha_2 \in (0,1)$, and EMA model initialised at $\hat{\sigma}_{V,0} = |V_0|$. This estimate, along with the bias-corrected spread, V_τ, can be used to compute the z-score of the spread's divergence, z_τ. The TVHR model recursions in their entirety are therefore

$$\begin{aligned}
S_\tau &= \left(\ln \frac{Y_\tau}{X_\tau}\right), &(3.32)\\
\hat{M}_\tau &= \alpha_0 S_\tau + (1-\alpha_0)\,\hat{M}_{\tau-1}, &(3.33)\\
\delta_\tau &= S_\tau - \hat{M}_\tau, &(3.34)\\
\hat{\mu}_{\delta,\tau} &= \alpha_1 \delta_\tau + (1-\alpha_1)\,\hat{\mu}_{\delta,\tau}, &(3.35)\\
V_\tau &= \delta_\tau - \hat{\mu}_{\delta,\tau}, &(3.36)\\
\hat{\sigma}_{V,\tau} &= \alpha_2 |V_\tau| + (1-\alpha_2)\,\hat{\sigma}_{V,\tau-1}, &(3.37)\\
z_\tau &= \frac{V_\tau}{\hat{\sigma}_{V,\tau}}. &(3.38)
\end{aligned}$$

Following the calculation of the standardised spread, critical z-scores indicating the point at which the spread has diverged significantly and a position can be opened, z_d, and the point at which the spread has once again converged sufficiently and the position can be closed, z_c, must be specified. The threshold values must satisfy $z_d > z_c \geq 0$, with common statistical arbitrage applications specifying $z_d = 2$ and $z_c = 0$. These values indicate that approximately 95% of observations of the standardised divergence, z_τ, will be insufficient to trigger a trading signal, with only the remaining 5% generating signals. The position to be taken in the artificial security in the next period, $\iota_{\tau+1} \in \{-1, 0, 1\}$, follows the piecewise function

$$\iota_{\tau+1} = \begin{cases} -\operatorname{sgn}(z_\tau), & |z_\tau| \geq z_d \\ \iota_\tau, & z_d > |z_\tau| > z_c \\ 0, & |z_\tau| \leq z_c \end{cases} \tag{3.39}$$

where $\iota_{\tau+1} = \iota_{\tau+1}(\alpha_0, \alpha_1, \alpha_2, z_d, z_c)$ is a function of five design parameters that can be optimised over some in-sample data partition. Similarly, the position to be taken in the artificial security in the presence of mean-aversion, $\iota^*_{\tau+1}$, follows

$$\iota^*_{\tau+1} = \begin{cases} \operatorname{sgn}(z_\tau), & |z_\tau| \geq z_d \\ \iota^*_\tau, & z_d > |z_\tau| > z_c \\ 0, & |z_\tau| \leq z_c \end{cases} \tag{3.40}$$

This inversion of trading rules allows the trading strategy to take full advantage of mean-averting behaviour attributable to the advance of the artificial security's equilibrium price to a new level.

Another practical consideration is the implementation of a stop-loss to limit excessive equity drawdowns in the event of a trading position that fails to revert within a reasonable amount of time, or one that is instigated before a volatile bias, $\mu_{\delta,\tau}$, quickly shifts the process to a new equilibrium. Such a consideration is analytically accommodated by the model presented in Sections 3.2 and 3.3, whose known mean-reversion rate, $(1 - \alpha_0)$ in Equation (3.28), allows for calculation of the expected first passage time and its confidence bounds. Though the effect of non-convergence of the spread is assuaged by the time-varying estimation procedure, the inclusion of a stop-loss based on the first passage time of the process grants further assurance against the accrual of losses in live trading.

The expected first passage time of an AR(1) process, $\tilde{\tau} = \sup\{\tau : z_\tau \leq z_c | z_0 \geq z_d\}$, can be calculated by

$$\mathbb{E}[\tilde{\tau}] = \sum_{i=1}^{\infty} i \cdot q_{\tilde{\tau}=i}, \tag{3.41}$$

where $q_{\tilde{\tau}=i}$ is the probability that the first passage time is i. Following Jaskowski and Van

Dijk (2015), these probabilities can be calculated by

$$\boldsymbol{q} = \Theta^{-1}\mathbf{A}, \tag{3.42}$$

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_\infty \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \theta_{2,1} & 1 & 0 & \dots & 0 \\ \theta_{3,1} & \theta_{3,2} & 1 & \dots & 0 \\ \dots & \dots & \dots & \ddots & \vdots \\ \theta_{\infty,1} & \theta_{\infty,2} & \theta_{\infty,3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \Phi\left(-\frac{z_c - \mathbb{E}_\tau[z_{\tau+1}]}{\mathbb{V}_\tau[z_{\tau+1}]}\right) \\ \Phi\left(-\frac{z_c - \mathbb{E}_\tau[z_{\tau+2}]}{\mathbb{V}_\tau[z_{\tau+2}]}\right) \\ \Phi\left(-\frac{z_c - \mathbb{E}_\tau[z_{\tau+3}]}{\mathbb{V}_\tau[z_{\tau+3}]}\right) \\ \vdots \\ \Phi\left(-\frac{z_c - \mathbb{E}_\tau[z_{\tau+\infty}]}{\mathbb{V}_\tau[z_{\tau+\infty}]}\right) \end{bmatrix}, \tag{3.43}$$

where the $(i,j)^{th}$ element of Θ is given by

$$\begin{aligned} \theta_{i,j} =& \frac{1}{\Phi\left(-\frac{z_c - \mathbb{E}_\tau[z_{\tau+j}]}{\mathbb{V}_\tau[z_{\tau+j}]}\right)} \\ & \int_{z_c}^{\infty} \phi\left(z_j = \tilde{z}; \mathbb{E}_\tau\left[z_{\tau+j}\right], \mathbb{V}_\tau\left[z_{\tau+j}\right]\right) \Phi\left(-\frac{z_c - \mathbb{E}_{\tau+i}\left[z_{\tau+i+j} | z_{\tau+i} = \tilde{z}\right]}{\mathbb{V}_{\tau+i}\left[z_{\tau+i+j} | z_{\tau+i} = \tilde{z}\right]}\right) \mathrm{d}\tilde{z}, \end{aligned}$$

with $\phi(\cdot)$ and $\Phi(\cdot)$ denoting the Gaussian density and distribution functions, respectively. Additionally, the i-period mean and variance of a zero-mean AR(1) process with unit variance, z_τ, are given by

$$\begin{aligned} \mathbb{E}_\tau\left[z_{\tau+i}\right] &= \left(1-\alpha_0\right)^i z_\tau, \\ \mathbb{V}_\tau\left[z_{\tau+i}\right] &= \sum_{k=0}^{i} \left(1-\alpha_0\right)^k. \end{aligned}$$

Given a trading signal generated at time τ to be implemented at time $\tau+1$, $\iota_{\tau+1}$, the expected first passage time allows the specification of the additional trading rule $\iota_{\tau+n} = 0$, where $n = \lceil \mathbb{E}[\tilde{\tau}] \rceil$, $n \in \mathbb{N}$. This rule liquidates any open positions that have not reverted within the expected amount of time. This stop-loss mechanism is only applicable to signals generated with implicit consideration of mean-aversion—that is, continued divergence of the spread in contradiction to the assumption of spread convergence. Signals generated under explicit consideration of mean-aversion, $\iota^*_{\tau+1}$, which does not follow an AR(1) process, cannot utilise the first passage time to establish when the position should be liquidated. A lack of stop-loss is acceptable for signals generated under a mean-averting regime as reversion to the mean z-score of zero acts as a natural stop-loss, terminating trades before they begin to accrue losses.

The specific characteristics of the TVHR model make consideration of the first passage time unnecessary in this thesis. The results reported in Chapter 4 indicate that the holding period for statistical arbitrage trades investigated under the backtesting procedure outlined in Section 3.6 are too brief for a time-based liquidation procedure to be warranted. This discussion is purely intended for the practical consideration of an arbitrageur.

3.5 Model Extensions

Extensions to the proposed TVHR model incorporate regime switching and statistical learning methodologies that attempt to identify and exploit the tendency for some statistical arbitrage opportunities to continue diverging. This continued divergence is the source of the non-convergence issue that has caused statistical arbitrage to decline in profitability over the years, though some research outlined in Section 2.1.4 indicates the possibility of exploiting such mean-averting behaviour. In contrast to traditional statistical arbitrage strategies, the exploitation of mean-aversion requires the inversion of trading rules when the observable spread reaches some significant level; while conventional statistical arbitrage would indicate entering a short position in the spread once it reaches a high positive value, with the expectation that the spread will revert to its mean level of zero, an expectation of mean-aversion will instead indicate the entry of a long position in the spread—this distinction can be observed in the position equations, Equation (3.39) and Equation (3.40). If the expectation of mean-aversion is met the spread will continue to diverge, moving to a new equilibrium level at which the arbitrageur will exit the long position in profit.

The literature in Section 2.1 does not offer consistent findings about the delineating feature that determines whether an arbitrage opportunity will continue to diverge. Huck (2015) found no evidence that volatility timing improves the profitability of statistical arbitrage, however, Do and Faff (2010) show that statistical arbitrage is more profitable during volatile bear markets. Bee and Gatti (2015) use spread volatility regime modelling to separate mean-averting from mean-reverting regimes, widening the trade entry thresholds during periods of high volatility. The approach achieves substantial profitability, supporting spread volatility as a variable on which to condition trade initiation. Krauss and Stübinger (2017) show that pairs can be found which are profitable when traded exclusively as mean-averting opportunities, achieving similar profitability to mean-reverting pairs under a copula-based modelling approach.

Section 2.2.1 details the use of regime switching models in finance, finding that there are relatively few applications to statistical arbitrage beyond specifying volatility regimes under which the spread is observed. Bock and Mestel (2009) model the observable spread under low and high volatility regimes, varying the trade initiation triggers for each respective regime. Similarly, Yang, Tsai, Shyu, and Chang (2016) model the spread as an OU process whose parameters are specified by a Markov switching model, finding evidence of a low-volatility regime corresponding with typical market dynamics, and a high-volatility regime corresponding with periods of market turmoil. As with the former study, trade initiation thresholds are widened during the high-volatility regime, indicating the tendency of market turmoil to precipitate mean-averting behaviour that would otherwise be unprofitable without consideration.

An alternative to the parametric Markov regime switching model is a statistical learning model that integrates non-linear, non-parametric universal function approximation, such as ANNs and their computationally optimal extensions, ELMs. The model would similarly indicate whether a mean-reverting or mean-averting trade would be optimal at any given time, but unlike the Markov regime switching model could optimise the functional mapping of the statistical learning method with respect to a metric of profitability. Under the Markov switching model, the model that most accurately reflects the true underlying process is not necessarily the one that optimises profitability. The ability for the arbitrageur to define the different regimes under the statistical learning approach affords it greater flexibility for exploitation of statistical arbitrage opportunities.

Section 2.2.2 discusses applications of statistical learning models to statistical arbitrage. There are relatively few such applications, though those that do exist indicate the ability of statistical learning models to identify profitable trading opportunities. While some unsupervised learning models are used for dimension reduction, such as PCA, the literature mostly details the use of supervised learning models to forecast either a class or scalar dependent variable. The use of classification models generally sees studies investigate whether statistical learning can predict profitability of stocks or arbitrage opportunities, while regression applications generally try to forecast the actual return at some horizon. Classification tasks are therefore more analogous to regime switching models, in which the time series is considered to be occupying a certain identifiable state at a given time.

The model extensions proposed in this thesis principally consider spread volatility for the identification of mean-averting behaviour. Given the inconsistency of findings in the literature regarding high and low volatility regimes, a simple logistic regression framework was chosen for the regime switching model due to its ability to model the probability of a trade being profitable, based on the level of volatility at the time of trade initiation, rather than modelling the different regimes directly. In this way, the logistic regression framework captures regime-dependent information in the level of volatility without introducing rigid modelling assumptions that might complicate the analysis. Similarly, an ELM framework was chosen as the statistical learning extension of the TVHR model due to its computational efficiency and classification ability. Applying both the regime switching logistic regression model and ELM statistical learning model to the task of profitability classification allows both models to be considered and their relative performance measured equitably, addressing the second and third research questions posited in this thesis.

3.5.1 Regime Switching with Logistic Regression

Logistic regression is a non-linear statistical model that performs multivariate classification by mapping its output to the interval $[0, 1]$, in which a label of 1 is considered a success possessing some desirable property, and a label of 0 is considered a failure. Given some vector of M input variables, $\boldsymbol{X} \in \mathbb{R}^{1\times M}$, logistic regression returns the probability, $Pr(\boldsymbol{Y}_n = 1|\boldsymbol{X})$ with $n = 1, 2, \ldots, N$ for $\boldsymbol{Y} \in \mathbb{R}^{1\times N}$, that the n^{th} output variable can be classified as a success—that is, that the observation belongs to the n^{th} class. For the remainder of Section 3.5, $\boldsymbol{X}$ and $\boldsymbol{Y}$ will refer to input and output variables, respectively, of statistical learning models. This notation should not be confused with the time series observations of security prices, $X(t)$ and $Y(t)$, used in the formation of statistical arbitrage stochastic spreads.

Given P distinct training samples, $\{\boldsymbol{X}_{p*}, \boldsymbol{Y}_{p*}|p = 1, 2, \ldots, P\}$, with $\boldsymbol{X} \in \mathbb{R}^{P\times M}$, $\boldsymbol{Y} \in \mathbb{R}^{P\times N}$, a mathematical expression for the estimate of the p^{th} training sample output, $\hat{\boldsymbol{Y}}_{p*}$, of a logistic regression model is given by

$$\hat{\boldsymbol{Y}}_{p*} = g\left(\boldsymbol{X}_{p*} \cdot \boldsymbol{\theta} + b\right), \tag{3.44}$$

where $\boldsymbol{\theta} \in \mathbb{R}^M$ is a vector of coefficients, $b \in \mathbb{R}$ is a bias term, and g is the logistic or sigmoid function, given by

$$g\left(z\right) = \frac{1}{1 + e^{-z}}. \tag{3.45}$$

The model parameters, $\boldsymbol{\theta}$ and b, are estimated numerically using optimisation procedures such as gradient descent.

The estimate of the training sample output, $\hat{\boldsymbol{Y}}_{p*}$, is the model-predicted probability that the p^{th} observation belongs to each of the N classes. The estimate of the output is therefore equivalent to the probability of belonging to each class, that is, $\hat{\boldsymbol{Y}}_{p*} = Pr(\boldsymbol{Y}_{p*} = 1|\boldsymbol{X}_{p*})$. The model is first trained using some subset of in-sample data before being applied to unseen, out-of-sample data, $p \geq P + 1$.

3.5.2 Statistical Learning with Extreme Learning Machines

Feedforward ANNs have been studied extensively for their ability to universally approximate complex non-linear mappings directly from sample input vectors, and their ability to provide models for a large class of natural and artificial phenomena that are difficult to handle using classical parametric techniques. Research into the function approximation ability of ANNs showed that a Single-Layer Feedforward Network (SLFN) with P hidden nodes can learn P distinct training samples with zero error (Huang and Babri, 1998). Though this serves to illustrate the impressive computational power of ANNs, it is not useful in practical applications because it causes overfitting of the ANN, leading to a mapping function which performs poorly on out-of-sample data.

Traditionally, all the transfer weights of the ANN are tuned through a gradient descent-based backpropagation learning algorithm, which can be slow to converge, and is not guaranteed to attain the global minimum cost function. Having several inter-dependent layers is what causes the ubiquitous computational burden and non-convex optimisation objective—if the network consisted only of an input layer and an output layer, the learning task would become a trivial convex optimisation problem.

A simplification of the process first requires understanding of the mapping capability of SLFNs with arbitrarily-chosen transfer weights between the input and hidden layers. In such situations, it was shown that a SLFN with P hidden nodes can learn P distinct training samples with arbitrarily small error (Tamura and Tateishi, 1997). This finding opens the possibility of assigning randomly sampled transfer weights between the input and hidden layers, and pursuing a simpler method of optimisation than backpropagation to tune the transfer weights between the hidden and output layers.

Huang, Zhu, and Siew (2004) showed that it is possible to construct a SLFN with uniformly-sampled transfer weights between the input and hidden layers, and almost any non-zero activation function, that can universally approximate any continuous function on any compact input set. This demonstrates that it is not necessary to tune the transfer weights between the input and hidden layers. A mathematical treatment of their method, named Extreme Learning Machine (ELM), is presented below following a brief discussion of SLFN approximation objectives.

Given P distinct training samples, $\{\boldsymbol{X}_{p*}, \boldsymbol{Y}_{p*} | p = 1, 2, \ldots, P\}$, with $\boldsymbol{X} \in \mathbb{R}^{P \times M}$ the matrix of P observations of M input variables, $\boldsymbol{Y} \in \mathbb{R}^{P \times N}$ the matrix of P observations of N output variables, P hidden nodes and activation function, g, a mathematical expression for the p^{th} training sample output, $\boldsymbol{Y}_{p*}$, of standard SLFNs can be expressed as

$$\hat{\boldsymbol{Y}}_{p*} = f(\boldsymbol{X}_{p*}), \tag{3.46}$$

$$= \sum_{h=1}^{H} g\left(\boldsymbol{X}_{p*}\boldsymbol{\Theta}^{(1)}_{*h} + \boldsymbol{B}_h\right)\boldsymbol{\Theta}^{(2)}_{h*}, \tag{3.47}$$

where $\boldsymbol{\Theta}^{(1)} \in \mathbb{R}^{M \times H}$ is the matrix of weights connecting the input layer to the hidden layer, $\boldsymbol{\Theta}^{(2)} \in \mathbb{R}^{H \times N}$ is the matrix of weights connecting the hidden layer to the output layer, and $\boldsymbol{B} \in \mathbb{R}^{1 \times H}$ is the vector of hidden node bias thresholds. In order to find a function approximation that allows the P training samples to be learned with zero error, there must exist $\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(2)}$ and $\boldsymbol{B}$ such that $\|\hat{\boldsymbol{Y}}_{p*} - \boldsymbol{Y}_{p*}\| = 0$, for $p = 1, 2, \ldots, P$. In typical SLFNs, the backpropagation algorithm, or a similar gradient-descent method, is responsible for this optimisation task.

The mathematical expression of the SLFN in Equation (3.46) can be written more succinctly as

$$\boldsymbol{Z}\boldsymbol{\Theta}^{(2)} = \boldsymbol{Y}, \tag{3.48}$$

where the hidden matrix, $\boldsymbol{Z}(\boldsymbol{X}, \boldsymbol{\Theta}^{(1)}, \boldsymbol{B})$, is the collection of the H hidden nodes expressed as a row vector, vertically stacked for each of the P training samples, given by

$$\boldsymbol{Z} = \begin{bmatrix} g\left(\boldsymbol{X}_{1*} \cdot \boldsymbol{\Theta}^{(1)}_{*1} + \boldsymbol{B}_1\right) & \dots & g\left(\boldsymbol{X}_{1*} \cdot \boldsymbol{\Theta}^{(1)}_{*H} + \boldsymbol{B}_H\right) \\ \vdots & \dots & \vdots \\ g\left(\boldsymbol{X}_{P*} \cdot \boldsymbol{\Theta}^{(1)}_{*1} + \boldsymbol{B}_1\right) & \dots & g\left(\boldsymbol{X}_{P*} \cdot \boldsymbol{\Theta}^{(1)}_{*H} + \boldsymbol{B}_H\right) \end{bmatrix}. \tag{3.49}$$

The hidden matrix, $\boldsymbol{Z}$, is a non-linear mapping of the input matrix, $\boldsymbol{X} \in \mathbb{R}^{P \times M} \rightarrow \boldsymbol{Z} \in \mathbb{R}^{P \times H}$. The non-linear mapping is fully specified by the activation function, g, the transfer weights between the input and hidden layers, $\boldsymbol{\Theta}^{(1)}$, and the hidden node bias threshold terms, $\boldsymbol{B}$.

As noted above, when the number of hidden nodes, H, is equal to the number of training samples, P, then the hidden matrix, $\boldsymbol{Z}$, is square and invertible, and the SLFN can approximate the P training samples with zero error. In most practical applications, however, $H \ll P$, which means that $\boldsymbol{Z}$ is no longer a square matrix, and there may not exist $\boldsymbol{\Theta}^{(1)}, \boldsymbol{\Theta}^{(2)}, \boldsymbol{B}$ that returns zero error.

When $\boldsymbol{Z}$ is unknown, gradient-based learning algorithms such as backpropagation are used to search the minimum of $\|\boldsymbol{Z}\boldsymbol{\Theta}^{(2)} - \boldsymbol{Y}\|$. The contribution of Huang, Zhu, and Siew (2004) was to show that the hidden matrix need not be unknown. Conventional SLFN training sees all transfer weights between each layer of the ANN iteratively optimised so that the network attains the minimum error. This leads to a costly computational burden, as well as several other problems relating to poor generalisation and sub-optimal approximations. ELMs dispense with iterative training by specifying a known hidden matrix, $\boldsymbol{Z}$, that remains constant throughout the entire training process. Arbitrary values drawn from $\mathcal{U}\left(-0.5, 0.5\right)$ need only be assigned to each element of $\boldsymbol{\Theta}^{(1)}$ and $\boldsymbol{B}$ at the beginning of the training process. The hidden matrix, $\boldsymbol{Z}$, can then be calculated as

$$\boldsymbol{Z} = g\left(\boldsymbol{X}\boldsymbol{\Theta}^{(1)} + \boldsymbol{B}\right), \tag{3.50}$$

$$= g\left(\begin{bmatrix} \boldsymbol{X}_{1*} \\ \boldsymbol{X}_{2*} \\ \vdots \\ \boldsymbol{X}_{P*} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Theta}^{(1)}_{*1} & \boldsymbol{\Theta}^{(1)}_{*2} & \dots & \boldsymbol{\Theta}^{(1)}_{*H} \end{bmatrix} + \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 & \dots & \boldsymbol{B}_H \\ \boldsymbol{B}_1 & \boldsymbol{B}_2 & \dots & \boldsymbol{B}_H \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_1 & \boldsymbol{B}_2 & \dots & \boldsymbol{B}_H \end{bmatrix}\right), \tag{3.51}$$

where g is the activation function, $\boldsymbol{X}$ is the $P \times M$ matrix of input variables, $\boldsymbol{\Theta}$ is the uniformly-sampled $M \times H$ matrix of transfer weights between the input and hidden layers, and $\boldsymbol{B}$ is the uniformly-sampled $1 \times H$ vector of threshold bias terms for each of the H hidden nodes. By specifying a fixed set of weights and biases and, consequently, a fixed hidden matrix across all training samples, the ELM training method reduces the problem to a simple linear system, given by

$$\boldsymbol{Z}\boldsymbol{\Theta}^{(2)} = \boldsymbol{Y}, \tag{3.52}$$

where $\boldsymbol{Y}$ represents the output matrix created by stacking the P transposed output vectors, giving $\boldsymbol{Y} \in \mathbb{R}^{P \times N}$. This linear system can be solved analytically by finding an OLS solution, $\hat{\boldsymbol{\Theta}}^{(2)}$, that minimises the error, such that

$$\|\boldsymbol{Z}\hat{\boldsymbol{\Theta}}^{(2)} - \boldsymbol{Y}\| = \min_{\boldsymbol{\Theta}^{(2)}} \|\boldsymbol{Z}\boldsymbol{\Theta}^{(2)} - \boldsymbol{Y}\|, \tag{3.53}$$

$$\hat{\boldsymbol{\Theta}}^{(2)} = \left(\boldsymbol{Z}^\intercal \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^\intercal \boldsymbol{Y}, \tag{3.54}$$

where the matrix inversion in Equation (3.54) is calculated by the Moore-Penrose pseudoinverse.

The estimate for $\hat{\boldsymbol{\Theta}}^{(2)}$ given by Equation (3.54) is known as the minimum norm least-squares solution. It has several interesting properties that make the ELM training method particularly attractive. It was discovered by Bartlett (1997) that, for feedforward networks, the magnitude of the transfer weights is what determines the generalisation ability of the network on unseen, out-of-sample data; the smaller the weights, the better the generalisation. The estimate for $\hat{\boldsymbol{\Theta}}^{(2)}$ therefore achieves the best generalisation performance, as it is known to have the smallest norm of all least-squares solutions, that is $\|\hat{\boldsymbol{\Theta}}^{(2)}\| = \|\left(\boldsymbol{Z}^\intercal \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^\intercal \boldsymbol{Y}\| \leq \|\boldsymbol{\Theta}^{(2)}\|$, $\forall \boldsymbol{\Theta}^{(2)} \in \{\boldsymbol{\Theta}^{(2)} \colon \|\boldsymbol{Z}\boldsymbol{\Theta}^{(2)} - \boldsymbol{Y}\| \leq \|\boldsymbol{Z}\boldsymbol{\Gamma} - \boldsymbol{Y}\|, \forall \boldsymbol{\Gamma} \in \mathbb{R}^{H \times N}\}$. Additionally, the estimate for $\boldsymbol{\Theta}^{(2)}$ is globally optimal and unique, owing to the convex optimisation objective of the linear system described by Equation (3.52).

The ELM training method is a very powerful way of non-linearly mapping sequences of input vectors to their corresponding output vectors, and approximating the functional relationship between them. The method only requires the arbitrary selection of transfer weights and biases for a hidden matrix that remains constant throughout training. Huang, Zhu, and Siew (2004) showed that it is not necessary to iteratively tune the transfer weights and biases in the hidden matrix, as is normally done in SLFN training, and that in fact no gains can be made from such a procedure. The ELM training method prescribes an analytical, globally-optimal and unique solution for the offline training of SLFNs, which is extremely computationally efficient and delivers the best possible generalisation performance on out-of-sample data.

3.5.3 Input and Output Variables

In this thesis, the input variables used by both the logistic regression regime switching model and ELM statistical learning model include the spread volatility level and the level of the VIX at the time of trade initiation. Given some pair whose spread, V_1, is observed to trigger a trade initiation opportunity by diverging significantly at time τ_0, the input variables are therefore the spread volatility level, σ_{V_1,τ_0}, and the VIX level at time τ_0. These two volatility variables are assumed to span the various sources of volatility to which a statistical arbitrage opportunity might be subject, both market-wide and idiosyncratic. Additionally, the mapping capability of ELMs allows the consideration of other variables, including the mean-reversion rate parameter, position of a given pair within the top 20, whether the constituent securities are from the same industry, as well as the forecast probabilities from the logistic regression model. These variables constitute all available information about a pair that can be discerned from its formation characteristics. The input vector size, M, is therefore two for the logistic regression model, and six for the ELM model.

Trade profitability has been chosen as the desired output variable, with profitable trades, $R_{\tau_0} \geq 0$, given the binary classification label 1 and unprofitable trades, $R_{\tau_0} < 0$, given the classification label 0, where R_{τ_0} is the return of the trade initiated at time τ_0. The output vector size, N, is therefore one. The output probability, $Pr(R_{\tau_0} \geq 0)$, for the logistic regression model is given by

$$\hat{\boldsymbol{Y}}_{\tau_0} = g\left(\boldsymbol{X}_{\tau_0 *} \cdot \boldsymbol{\theta} + b\right), \tag{3.55}$$

where $\boldsymbol{X}_{\tau_0 *} \in \mathbb{R}^{1\times 2}$ is the vector of market and idiosyncratic volatility levels at the time of trade initiation, $\hat{\boldsymbol{Y}}_{\tau_0} \in \mathbb{R}$ is the scalar model-predicted probability that the trade is profitable, and $\boldsymbol{\theta}$ and b are model parameters iteratively optimised by a logistic regression learning algorithm. The output probability for the ELM model is given by

$$\hat{\boldsymbol{Y}}^{\boldsymbol{*}}{}_{\tau_0} = g\left(\boldsymbol{X}^{\boldsymbol{*}}{}_{\tau_0 *}\boldsymbol{\Theta}^{(1)} + \boldsymbol{B}\right)\hat{\boldsymbol{\Theta}}^{(2)}, \tag{3.56}$$

where $\boldsymbol{X}_{\tau_0 *} \subsetneq \boldsymbol{X}^{\boldsymbol{*}}{}_{\tau_0 *} \in \mathbb{R}^{1\times 6}$ is the vector of market and idiosyncratic volatility levels at the time of trade initiation, in addition to the mean-reversion rate parameter, position of the pair within the top 20, a dummy variable indicating whether the securities are from the same industry, and the forecast probability from the logistic regression model, $\hat{\boldsymbol{Y}}_{\tau_0}$. Again, $\hat{\boldsymbol{Y}}^{\boldsymbol{*}}{}_{\tau_0}$ is a scalar model-predicted probability that the trade is profitable, and $\hat{\boldsymbol{\Theta}}^{(2)}$ is a matrix of transfer weights optimised by OLS regression. Unlike the logistic regression model, $\boldsymbol{\Theta}^{(1)}$ and $\boldsymbol{B}$ are not iteratively tuned, but rather sampled from a uniform distribution and held constant throughout the procedure. In both cases, g is the logistic or sigmoid function.

3.6 Empirical Evaluation

3.6.1 Backtesting Procedure

The backtesting procedure developed by Gatev, Goetzmann, and Rouwenhorst (2006) and later used by Do and Faff (2010) to evaluate the performance of the distance approach to statistical arbitrage is replicated in this thesis. The purpose of the backtesting procedure, as it was first described by Gatev, Goetzmann, and Rouwenhorst (2006), is to capture the essence of apocryphal statistical arbitrage trading approaches developed on Wall Street, while avoiding the dangers of overfitting and data-mining bias. The procedure therefore makes very few design choices that would require the optimisation of hyperparameters, instead relying on heuristics and common sense to direct the experiment.

Securities are first screened to ensure trading activity was observed on each day of a 12-month formation period. Securities are matched with each viable counter-security, avoiding pairs for which both securities come from the same company; different share classes issued by the same company, for example, constitute the most frequently-occurring pairs. Following Do and Faff (2010), these pairs are excluded to avoid any possible restrictions to which they may be subject.

Given N securities available for the 12-month formation period, all $N(N-1)/2$ bivariate pairings of securities are considered for inclusion in a portfolio of the top 20 pairs to be traded in the subsequent six-month trading period. Pairs are ranked according to the SSD metric, with those recording the 20 lowest SSD statistics selected for inclusion in a portfolio. The formation period data window is advanced by one month, and the process is repeated with a new set of viable securities to select the top 20 pairs to be traded in the subsequent six-month period. Since the portfolios are each traded for six months, there are six overlapping portfolios trading at any given time, each staggered by one month.

The pair selection procedure, as described by Gatev, Goetzmann, and Rouwenhorst (2006), first normalises prices of candidate securities at the beginning of the formation period. The SSD statistic is calculated for each viable pairing, and the top 20 pairs are selected. Selected pairs' prices are once again normalised at the beginning of the trading period, and a \$1 long-short position is entered once the normalised pair spread reaches a significant level of deviation, and closed once the spread reverts to its equilibrium level. The pair return over the period is considered to be the sum of long and short returns in the respective securities, with pair returns marked-to-market on a daily basis.

Pairs trading in this manner constitutes a dollar-neutral, self-financing trading and investment strategy. The short-sale of the overvalued security generates a positive cash flow of $1 which is then used to finance the long position in the undervalued security. As such, raw returns from the strategy are considered excess returns, and consequently the risk-free rate is assumed to be zero. Short-sale costs are similarly assumed to be zero, with no restrictions inhibiting the short-sale of any securities.

Two variations of return calculation are considered, namely those of committed and employed capital. Committed capital assumes that an equal allocation of capital to each of the 20 pairs is made at the beginning of the trading period, irrespective of whether a pair actually places a trade during the period. Employed capital, on the other hand, assumes the arbitrageur can mobilise capital in an efficient manner, allocating equal capital only to those pairs that place transactions during the trading period. Committed capital therefore prescribes a more conservative way of estimating returns, while employed capital more closely resembles the capital utilisation achievable by hedge funds.

In an extension of the framework developed by Gatev, Goetzmann, and Rouwenhorst (2006), a cointegration approach is considered in the empirical evaluation. While the cointegration approach of Do and Faff (2010) is heavily influenced by the distance approach, the cointegration approach investigated in this thesis does not condition its pair selection on having the smallest SSD statistic. As noted by Krauss (2017), the SSD metric is sub-optimal because it penalises pairs with a high degree of spread volatility. The author proposes an alternative selection criterion, in which the second summand in Equation (2.3) is minimised first, selecting pairs with the highest spread volatility among those that remain.

For the cointegration approach considered in this thesis, Section 3.2.4 illustrates that the second summand in Equation (2.3) is zero by construction when the spread is calculated as the difference between a stochastic variable and its mean, as in Equation (3.2). Only the spread volatility in the SSD decomposition remains, with Krauss (2017) proposing pairs be selected according to the highest spread volatility. Following Do and Faff (2010), the cointegration approach used in this thesis selects the 20 pairs with the greatest number of zero-crossings in the formation period as a proxy for spread volatility that is more relevant to the arbitrageur. In the absence of qualitative information about pair suitability, such as liquidity, market capitalisation and industry restrictions, the resultant pairs are typically quite volatile, sharing little or no arbitrage relationship. The resultant portfolios generate high absolute returns but inconsistent risk-adjusted performance. Their consideration in this thesis serves to illustrate the performance of selection guidelines advanced in the literature.

The proposed TVHR model is evaluated in this thesis relative to its static distance and cointegration counterparts. Following selection of pairs based on the above distance and cointegration algorithms, the TVHR model is evaluated with respect to these selected pairs. Consequently, the TVHR model is evaluated twice; first with pairs selected according to the distance approach, and then with pairs selected according to the cointegration approach. The TVHR model does not have its own pair selection procedure, but rather dictates how the identified pair should be traded. This comparative evaluation procedure allows for a more thorough investigation of the characteristics and dynamics of the TVHR model relative to the conventional approaches.

The principal difference separating the distance, cointegration, and TVHR modelling approaches lies in how the pair spreads are calculated. For the distance approach, security prices are normalised at the beginning of the formation period and again at the beginning of the trading period, which consequently informs the hedge ratio used to calculate the spread. The cointegration approach, by contrast, estimates the optimal hedge ratio during the formation period and applies it during the trading period. Both of these approaches maintain a static estimate of the hedge ratio through the trading period, while the TVHR model re-estimates the hedge ratio dynamically. Irrespective of how the hedge ratio and therefore the spread is calculated, all approaches enter a position once the spread reaches a statistically significant level of divergence.

In accordance with the literature surveyed in Section 2.1, two standard deviations is considered a statistically significant divergence sufficient to open a position in the pair. The variance of the spread is estimated during the formation period and used during the trading period for the distance and cointegration approach to trigger the opening of positions, while the TVHR model uses the formation period variance as an initial estimate that is re-estimated throughout the trading period.

A trade is initiated by investing $1 in a long position in the undervalued security, and $1 in a short position in the overvalued security. The total value of the trade is therefore $2 with 50% allocated to each security, which is subsequently unwound once the spread reverts to a level of zero. Like Gatev, Goetzmann, and Rouwenhorst (2006) and Do and Faff (2010), two variations of the trading rule are considered to assess the returns of the strategy in excess of any upward bias induced by the bid-ask bounce. The first variation opens and closes positions on the same day that signals are generated, while the second variation imposes a one-day delay on opening and closing positions once signals are generated.

3.6.2 Data

Data for the empirical evaluation consists of daily price observations for every ordinary stock in the CRSP database. The sample period begins in July 1962 and extends through June 2018. Stock codes 10 and 11 are used to retrieve ordinary stocks from the database, with daily close price, Standard Industrial Classification (SIC) code, delisting return, total daily return and daily volume retrieved for each stock over the period.

Candidate stocks are excluded over a given formation period if they have one or more missing prices, or one or more days of zero trading volume. Prices of the remaining candidate stocks are adjusted for dividends and splits at the beginning of the formation period using the total daily return information retrieved from CRSP, and adjusted again at the beginning of the trading period if selected for inclusion in a pair portfolio. This price adjustment should not be confused with the normalisation of prices at the beginning of formation and trading periods in accordance with the distance approach.

3.6.3 Parameter Estimation

Formation periods begin on the first trading day of every month, extending for a total of 12 months. All candidate stocks that have passed the previous step are matched with every other candidate stock and assessed for their viability as a pair portfolio. Under the distance approach, prices are normalised at the beginning of the formation period and their spread SSD calculated. The spread is calculated as

$$U_\tau = Y_\tau - \frac{Y_1}{X_1} X_\tau,$$

while the SSD is calculated as

$$\frac{\sum_{\tau=1}^{T} U_\tau^2}{Y_1^2},$$

given $\tau = 1, 2, \ldots, T$ days of price data during the formation period, and where the denominator is required for symmetry since $Y_\tau - (Y_1/X_1)\, X_\tau \neq X_\tau - (X_1/Y_1)\, Y_\tau$. Under the cointegration approach, however, prices do not require normalisation. The cointegration approach borrows the proposed unifying spread formulation, itself a time-invariant variation of the proposed TVHR model, whose spread is calculated as

$$V_\tau = \ln\left(\frac{Y_\tau}{X_\tau}\right) - \frac{1}{T}\sum_{\tau=1}^{T}\left(\ln\left(\frac{Y_\tau}{X_\tau}\right)\right),$$

where the second summand is simply the mean of the log-ratio, $\ln(Y_\tau/X_\tau)$, over the formation period. The SSD metric under the cointegration approach is therefore

$$\sum_{\tau=1}^{T} V_\tau^2,$$

with no adjustment necessary for symmetry thanks to the use of logarithms, which make the price ratio symmetrical about zero instead of being asymmetrical about one.

The distance approach estimate of the hedge ratio, β, during the formation period is therefore Y_1/X_1, though this estimate is discarded and the hedge ratio re-estimated at the beginning of the trading period. The cointegration approach estimate of β is

$$\frac{1}{T}\sum_{\tau=1}^{T}\left(\ln\left(\frac{Y_\tau}{X_\tau}\right)\right),$$

which, unlike the distance approach, is retained for the subsequent trading period. The standard deviation of the distance approach spread is calculated relative to U_τ, while the standard deviation of the cointegration approach spread is calculated relative to V_τ. These standard deviations are also retained for the trading period.

In addition to the above estimates of hedge ratio, SSD and spread standard deviation, the cointegration approach also requires an estimate of the number of spread zero-crossings, $\mathbb{E}[D]$, of each candidate pair during the formation period. A zero-crossing is defined as the change in sign of the spread, from positive to negative or from negative to positive, that is, $V_{\tau+1}V_\tau < 0$ for $\tau = 1, 2, \ldots, T-1$. A pair with a high number of zero-crossings is assumed to exhibit a favourably high level of volatility which the arbitrageur can profitably exploit.

Distance approach pairs are selected for the trading portfolio in order of increasing SSD, with only the 20 smallest SSD pairs being selected. Conversely, cointegration pairs are selected for the portfolio in order of decreasing zero-crossings, with only the 20 pairs with the greatest number of zero-crossings ultimately selected. Pairs selected for the distance and cointegration approach are traded in the subsequent trading period in accordance with the methodology specified by Gatev, Goetzmann, and Rouwenhorst (2006), where the normalised spread generates an entry signal in the pair portfolio if the z-score of the spread diverges by two or more standard deviations from zero. The proposed TVHR model trades pairs selected under the distance and cointegration approaches using the formation period statistics as initial estimates of the parameters. While the distance and cointegration approach use static parameter estimates, the TVHR model continues to re-estimate its hedge ratio and standard deviation parameters throughout the trading period.

The learning rate for the TVHR model is estimated from the number of zero-crossings of the pair observed during the formation period. Because the TVHR model assumes an AR(1) process in the spread, V_τ, Ylvisaker (1965) showed that the first-order autocorrelation, and consequently the reversion rate parameter, α, can be estimated directly from the number of zero-crossings as

$$\alpha = 1 - \rho_1, \tag{3.57}$$

$$= 1 - \cos\left(\frac{\pi\mathbb{E}\left[D\right]}{T-1}\right), \tag{3.58}$$

where $\rho_1 = \mathbb{E}\left[V_{\tau+1}V_\tau\right]/\mathbb{E}\left[V_\tau^2\right]$ is the first-order autocorrelation of the spread. This reversion rate determines the speed with which the estimates of the hedge ratio and spread standard deviation evolve, with a value close to one evolving smoothly and more closely resembling a static estimate, while a value close to zero resembles a white noise process. Though the TVHR model requires the specification of three reversion rates, namely α_0, α_1 and α_2, for simplicity this thesis assumes all reversion rates are identical to the above estimate, α. The adaptive nature of the TVHR model is what distinguishes it from the conventional distance and cointegration approaches, with the estimation of the reversion rate having a significant impact on the model's behaviour and performance. The reversion rate could therefore be tuned to optimise performance, so the above estimate based on the number of zero-crossings was chosen as an analytical way to select a parameter value free of forward-looking bias.

3.6.4 Performance Measures

Empirical performance of the distance and cointegration approaches, in addition to their TVHR variants, is evaluated with respect to monthly returns of the six overlapping pairs portfolios, each consisting of the top 20 pairs selected for a given formation period. In accordance with the methodology of Gatev, Goetzmann, and Rouwenhorst (2006), monthly returns are averaged across the six overlapping portfolios, in essence giving a uniform allocation of capital to each of the six portfolios. The returns of this aggregate portfolio are compounded monthly.

The results are reported for aggregate portfolio returns, hereafter referred to as portfolio returns, in addition to pair statistics and individual trade statistics. Pairs are divided into three groups: Group 1 describes pairs that do not place any trades during the trading period, Group 2 includes pairs that place a single non-convergent trade during the trading period, and Group 3 includes pairs that place one or more complete trades during the trading period. Similarly, trade statistics are segmented by non-convergent trades and convergent trades.

Portfolio return statistics include the mean, median, standard deviation, skewness, kurtosis, and proportion of negative monthly returns. Risk measures include the lower and upper semi-deviation, annualised Sharpe and Sortino ratios, proportion of Time In Market

(TIM), and mean return per unit TIM. The latter performance measures are intended to measure the efficiency of the various approaches; given equivalent monthly returns among the different methodologies, an approach that generates its return while spending less time in the market is generally less risky and more efficient than its alternatives.

Pair statistics cover the proportions observed in each of the three groups, as well as the proportion of profitable pairs per group, mean total return for the pairs over the trading period, average number of trades, TIM, the proportion of pairs that are industry-matched, the number of formation period zero-crossings under both the distance spread construction, U_τ, and cointegration spread construction, V_τ, and Beta Mean Difference (BMD). The final metric measures the mean absolute difference between the TVHR model's time-varying estimate of the hedge ratio, β_τ, and the formation period estimate that initialises it. The BMD is therefore calculated as

$$BMD = \frac{1}{T^* - T} \sum_{\tau=T+1}^{T^*} |\beta_\tau - \beta|, \tag{3.59}$$

where β is the formation period estimate used by the distance or cointegration approach, respectively, during the trading period, and T^* is the terminal time of the trading period. BMD is intended to measure the average deviation of the TVHR estimate from the formation period estimate of the hedge ratio.

Individual trade statistics report the proportion that are either convergent or non-convergent, as well as the profitable proportion for each, mean return, standard deviation, Sharpe ratio, Sortino ratio, mean profit, mean loss, mean long return, mean short return, mean trade length and median trade length. All Sharpe and Sortino ratio calculations refer to the annualised Sharpe and Sortino ratio, respectively. Given a sampling frequency, N, where N represents the number of times an observation is made in a year, the annualised Sharpe and Sortino ratios are calculated by multiplying the risk-adjusted mean return by $\sqrt{N}$. In the case of monthly return statistics, the Sharpe ratio is given by

$$\sqrt{N}\frac{\mu_R}{\sigma_R}, \tag{3.60}$$

where μ_R and σ_R are the monthly mean and standard deviation of returns, respectively, and $N = 12$. Similarly, the Sortino ratio is given by

$$\sqrt{N}\frac{\mu_R}{\sigma_R^-}, \tag{3.61}$$

where $\sigma_R^- = \left(\mathbb{E}\left[(R - \mathbb{E}[R])^2 \mathbf{1}_{\{R\leq 0\}}\right]\right)^{1/2}$ is the lower semi-deviation. $\mathbf{1}_{\{R\leq 0\}}$ is an indicator function returning 1 when the return, R, is negative, and zero when it is positive. The upper semi-deviation, σ_R^+, is calculated by changing this indicator variable to $\mathbf{1}_{\{R>0\}}$, which returns a 1 for positive returns and zero otherwise.

4 Results and Analysis

4.1 Proposed TVHR Model

4.1.1 Initial Study Period

The initial study period investigated in this thesis covers the sample period used by Do and Faff (2010), itself covering and extending the sample period used by Gatev, Goetzmann, and Rouwenhorst (2006). Employing daily stock prices retrieved from the CRSP database, Section 4.1.1 investigates the statistical and economic performance of conventional distance and cointegration variants of statistical arbitrage strategies, in addition to their TVHR analogues, throughout the sample period July 1962–June 2009. Pairs are selected according to either a distance (D) or cointegration (C) specification, the trading of which follows the procedure described by Gatev, Goetzmann, and Rouwenhorst (2006). The proposed TVHR model accepts the pairs selected under the distance and cointegration approaches, replacing the trading procedure used by Gatev, Goetzmann, and Rouwenhorst (2006) with a time-varying alternative. Time-varying distance (DTVHR) and time-varying cointegration (CTVHR) pairs are otherwise identical to their static counterparts, allowing for an unfettered investigation of their relative performance and characteristics.

Table 4.1 reports the excess return statistics for portfolios of the top 20 unrestricted pairs with immediate execution upon generation of a trading signal. Results are segmented by the capital allocation scheme used to calculate returns; committed capital uniformly allocates funds to each of the 20 pairs, while employed capital uniformly allocates funds only to those pairs that open a position during the trading period. Consequently, employed capital inflates the return relative to committed capital, though Gatev, Goetzmann, and Rouwenhorst (2006) argue that this more closely resembles the capital allocation schemes used by hedge funds. Unless otherwise stated, the remainder of this thesis will consider only employed capital in accordance with Gatev, Goetzmann, and Rouwenhorst (2006). Figure 4.1 displays the cumulative excess return of unrestricted distance and cointegration pairs under employed capital.

	Committed capital				Employed capital			
	D	DTVHR	C	CTVHR	D	DTVHR	C	CTVHR
Mean	0.0087	0.0082	0.0305	0.0093	0.0088	0.0084	0.0312	0.0172
t-Statistic	11.5583	19.0285	7.0356	9.5103	11.6668	18.8578	7.1511	11.0282
Median	0.0081	0.0081	0.0277	0.0053	0.0083	0.0082	0.0291	0.0103
Standard deviation	0.0110	0.0068	0.1071	0.0150	0.0111	0.0070	0.1081	0.0272
Skewness	0.9776	2.0047	-3.1104	2.0223	0.9281	1.9613	-3.0243	1.3757
Kurtosis	8.3016	21.0076	42.6631	11.8012	8.1550	19.7502	41.4049	8.6765
Minimum	-0.0375	-0.0169	-1.1993	-0.0537	-0.0375	-0.0169	-1.1993	-0.1112
Maximum	0.0717	0.0747	0.6854	0.0990	0.0720	0.0747	0.6854	0.1590
Observations < 0	0.1703	0.0634	0.2446	0.1087	0.1703	0.0634	0.2446	0.1087
Lower semi-deviation	0.0033	0.0015	0.0764	0.0041	0.0034	0.0015	0.0766	0.0085
Upper semi-deviation	0.0136	0.0106	0.0809	0.0171	0.0138	0.0108	0.0822	0.0310
Sharpe ratio	2.7570	4.1574	0.9849	2.1412	2.7597	4.1365	0.9995	2.1864
Sortino ratio	9.0691	19.0973	1.3818	7.7972	8.9477	19.1213	1.4102	7.0094
TIM	0.6269	0.1973	0.6729	0.0225	0.6269	0.1973	0.6729	0.0225
Return/TIM	0.0139	0.0416	0.0453	0.4123	0.0141	0.0424	0.0463	0.7653

Table 4.1: Excess return statistics for portfolios of top 20 unrestricted pairs, no execution delay, July 1962–June 2009.

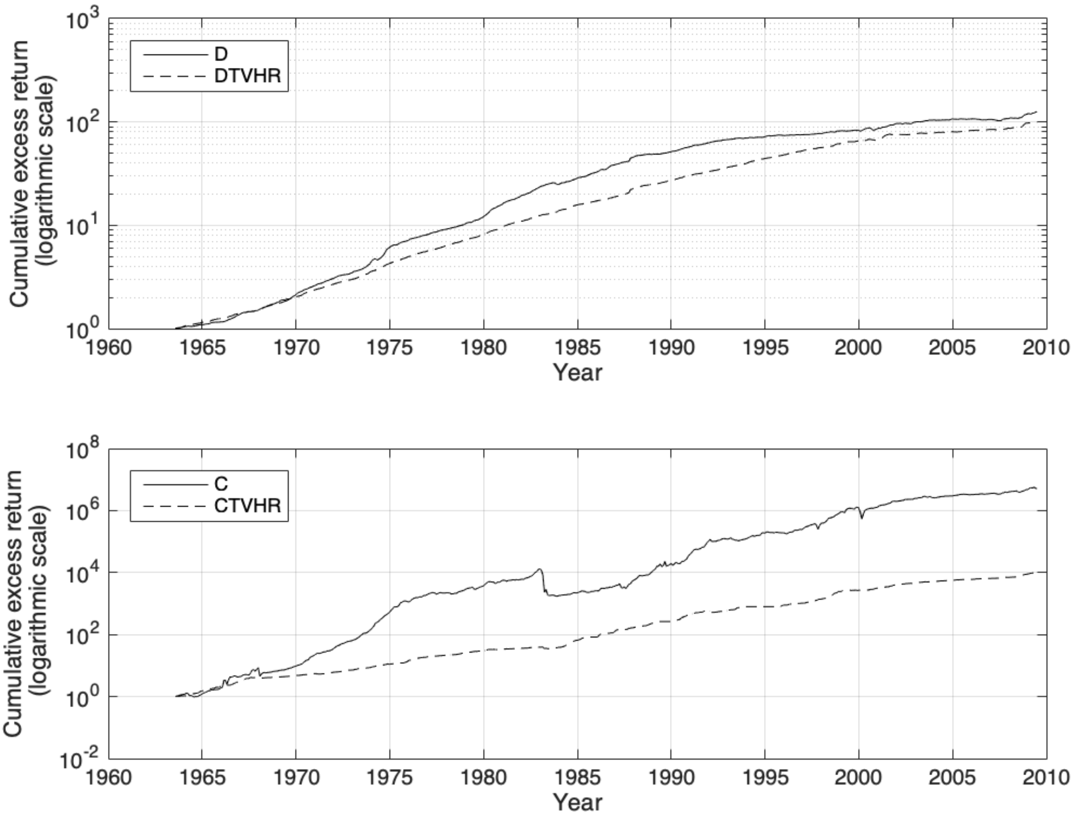

Figure 4.1: Cumulative excess return of top 20 unrestricted pairs, employed capital, no execution delay, July 1962–June 2009.

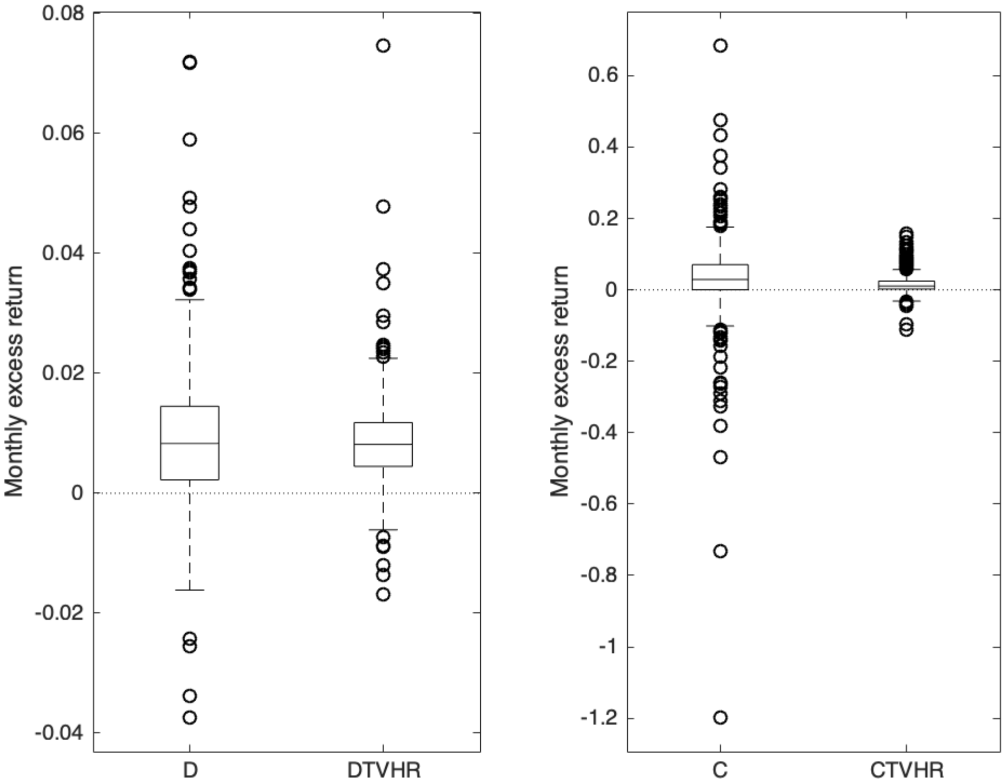

Figure 4.2: Distribution of monthly excess return of top 20 unrestricted pairs, employed capital, no execution delay, July 1962–June 2009.

The mean monthly committed capital return for the distance portfolio of 0.87%, and mean monthly employed capital return of 0.88%, closely align with the figures reported by Do and Faff (2010) of 0.89% and 0.90%, respectively. The DTVHR mean monthly returns under committed and employed capital are very similar to their static distance counterparts, at 0.82% and 0.84%, respectively. The C and CTVHR portfolios have no published reference figures to compare with, and their mean monthly employed capital returns are 3.12% and 1.72%, respectively. Figure 4.2 displays the distribution of monthly excess returns for distance and cointegration portfolios under employed capital.

The D and DTVHR portfolios realise remarkably similar mean monthly returns, but markedly different performance statistics. Under both committed and employed capital, the t-statistics, computed with Newey-West standard errors with six lags, are almost doubled under the TVHR specification relative to the static specification. Median monthly returns are similar, but return standard deviations are approximately 60% greater under the static D approach. Skewness, while positive for both strategy variants, is doubled under the TVHR model, and kurtosis is almost trebled. Both model specifications realise similar maximum monthly returns of 7.20% under the static D approach and 7.47% under

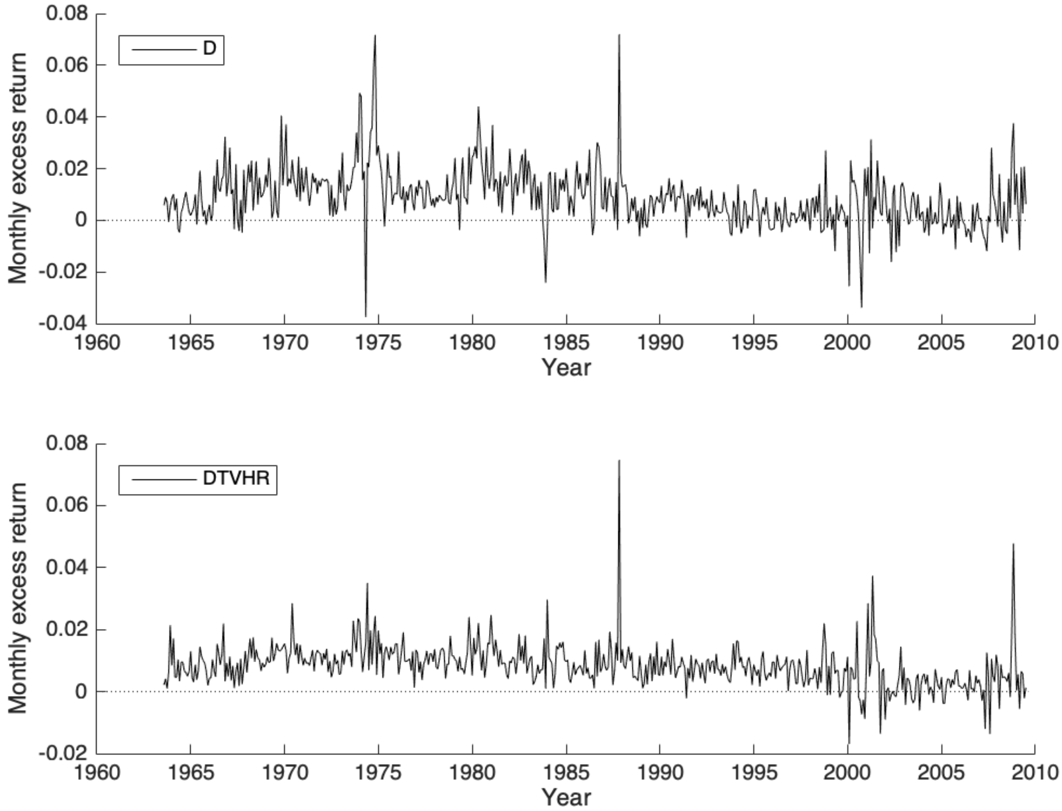

Figure 4.3: Monthly excess return of top 20 distance pairs, employed capital, no execution delay, July 1962–June 2009.

the DTVHR approach, though the DTVHR model's minimum monthly return of −1.69% is less than half that of the static model at −3.75%. The maximum monthly returns under both model specifications, however, are several times greater in magnitude than the minimum monthly returns. The proportion of DTVHR negative monthly returns across the sample period is almost one-third that of the static D model, at 6.34% and 17.03%, respectively. Figure 4.3 displays the monthly excess returns of static and TVHR distance portfolios under employed capital.

The annualised Sharpe and Sortino ratios of distance portfolios are substantially higher under the TVHR specification, experiencing little distortion across committed and employed capital calculations. TIM reports the proportion of time spent in the market—for example, a pair that trades 50 days in a 100-day period will report a TIM of 0.50, or 50%. The DTVHR model spends one-third the time in the market that the static D model does, inflating its return per unit TIM to 4.24% relative to the static model's 1.41%. These statistics paint a picture of the TVHR model as an efficient alternative to its conventional static analogue, mitigating risk by spending little time in the market pursuing only the most profitable arbitrage opportunities. While it delivers similar mean returns to the static distance model, its risk-adjusted returns are consistently better.

Mean monthly return of the static C model is more than three-times greater than both distance specifications under committed and employed capital, while the mean monthly return of the CTVHR model is on par with the distance models under committed capital, and more than double under employed capital. With the exception of the CTVHR model, all other specifications exhibit mean return stability across committed and employed capital. The near doubling of CTVHR mean return from committed to employed capital suggests a significant proportion of pairs do not trade during the trading period, in contrast to all other model specifications. The t-statistics of mean return are lower under the cointegration models than the distance models, but the TVHR specification is still higher than its static counterpart. The static model median return of 2.91% is similar to its mean return of 3.12% under employed capital, while the cointegration TVHR median return of 1.03% is almost half its mean of 1.72%.

Return standard deviation of the C and CTVHR models are an order of magnitude greater than their distance counterparts, explaining their deflated t-statistics despite their higher mean returns. Positive skewness reported for the CTVHR model is of similar magnitude to its distance equivalent. The static C model is the only model specification that delivers negative skewness of returns, in addition to exhibiting the highest magnitude of skewness. The static C model also exhibits the highest kurtosis and the greatest minimum and maxiumum monthly returns of −119.93% and 68.54%, respectively. It is the only model whose maximum return is smaller in magnitude than its minimum return. By contrast, the CTVHR model's minimum return of −11.12% is an order of magnitude lower than its static counterpart, and its maximum return of 15.90% duplicates the pattern exhibited by the distance models of maximum return being greater in magnitude than minimum return. The CTVHR model reports the second lowest proportion of negative monthly returns across all models at 10.87%, while the static C model reports the highest proportion of negative monthly returns at 24.46%. Figure 4.4 displays the monthly excess returns of static and TVHR cointegration portfolios under employed capital.

The static C model reports relatively low Sharpe and Sortino ratios, the lowest of all model specifications. Additionally, the model's Sharpe and Sortino ratios are quite similar, in contrast to all other model specifications whose Sortino ratios are several times greater than their Sharpe ratios. This reflects the approximate equivalence of the static model's lower and upper semi-deviation. The CTVHR model, by contrast, reports annualised Sharpe and Sortino ratios that are similar to the static D model. The static C model reports TIM of 67.29%, while the CTVHR model reports the lowest TIM of 2.25%. The CTVHR model therefore spends very little time in the market, opening and rapidly closing positions to extract profit from only the most extreme arbitrage opportunities.

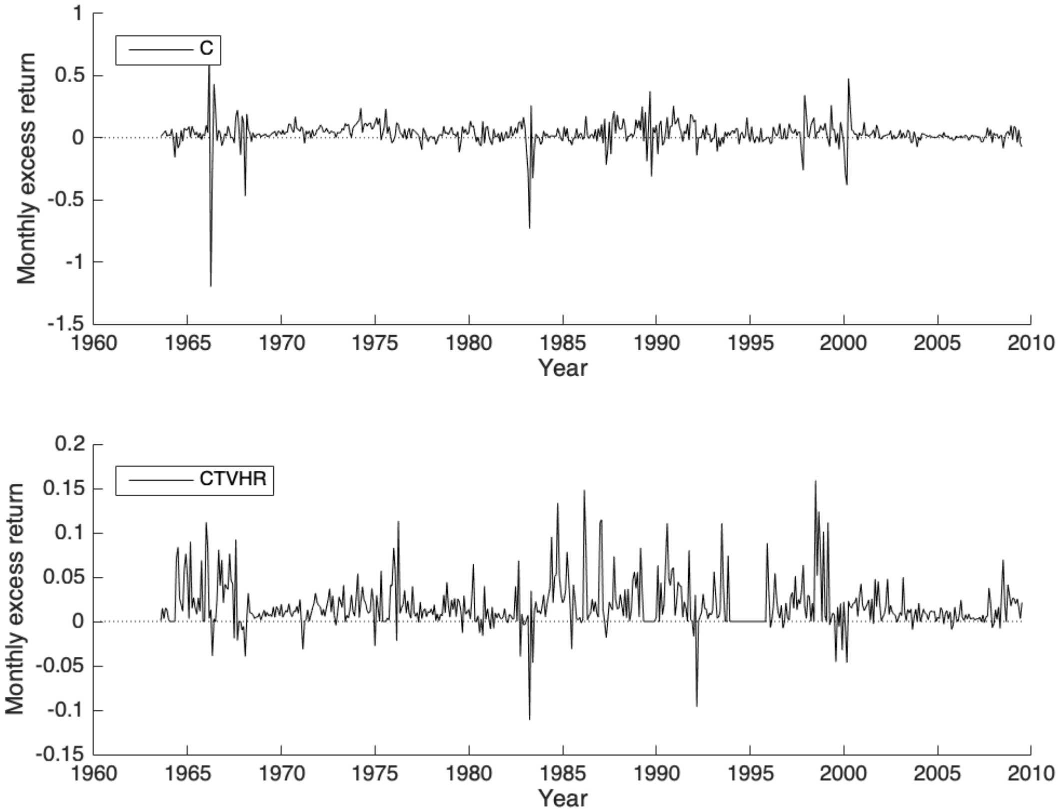

Figure 4.4: Monthly excess return of top 20 cointegration pairs, employed capital, no execution delay, July 1962–June 2009.

	D	DTVHR	C	CTVHR
Group 1 proportion	0.0164	0.0176	0.0214	0.4756
Group 2 proportion	0.2941	0.0034	0.3514	0.0035
Profitable proportion	0.2046	0.3784	0.2467	0.4211
Total return	-0.0609	-0.0294	-0.2440	-0.0256
Number of trades	1.0000	1.0000	1.0000	1.0000
TIM	0.7643	0.2628	0.8212	0.0342
Industry-matched	0.1328	0.1081	0.1279	0.1053
Distance zero-crossings	40.5928	32.5135	35.6269	36.2632
Cointegration zero-crossings	42.7078	30.4324	78.3379	72.5000
BMD	0.0000	0.7044	0.0000	4.1883
Group 3 proportion	0.6895	0.9790	0.6272	0.5209
Profitable proportion	0.9016	0.7913	0.9020	0.7964
Total return	0.1023	0.0509	0.4318	0.1068
Number of trades	2.7234	5.1043	2.7183	2.5129
TIM	0.5804	0.1998	0.6094	0.0422
Industry-matched	0.0997	0.1105	0.1349	0.1439
Distance zero-crossings	43.4118	42.2945	36.4415	32.4682
Cointegration zero-crossings	46.1197	44.7205	77.0592	71.0919
BMD	0.0000	0.5104	0.0000	1.3223

Table 4.2: Pair statistics for top 20 unrestricted pairs, no execution delay, July 1962–June 2009.

This efficiency facilitates the highest return per unit TIM of 76.53% under employed capital.

The characteristics and performance statistics of cointegration pairs are significantly different to those of conventional distance pairs. While the mean monthly returns delivered by the cointegration models are substantially greater than those of the distance models, they come at the price of greater return volatility and model instability. For example, the greatest one-month loss under the static D model of 3.75% pales in comparison to the greatest static C model loss of 119.93%. The volatility of the cointegration model can be attributed to its pair selection procedure which prioritises spread volatility among eligible pair constituents. The distance approach, by contrast, prioritises spread minimisation as a proxy for stochastic co-evolution of securities. These diametrically-opposed pair selection objectives create differing trade characteristics—the distance approach favours stability at the cost of higher returns, while the cointegration approach favours higher returns at the cost of stability. In both cases, however, the TVHR model serves to stabilise trade dynamics and improve risk-adjusted return.

Table 4.2 reports the pair statistics for portfolios of the top 20 unrestricted pairs with immediate execution upon generation of a trading signal. Pairs are divided into three

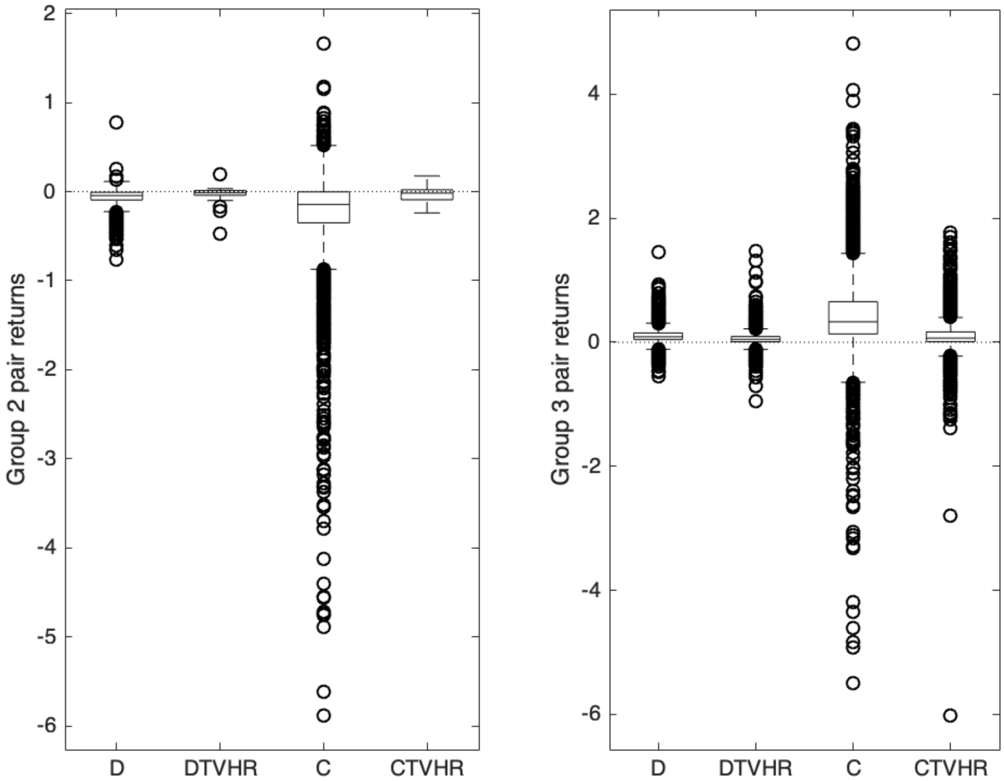

Figure 4.5: Distribution of pair excess return of top 20 unrestricted pairs, employed capital, no execution delay, July 1962–June 2009.

groups: Group 1 includes those pairs that do not open a single position during the trading period; Group 2 includes those pairs that open a single trade that remains open at the end of the trading period; Group 3 includes those pairs that open and close one or more trades. Group 2 therefore represents the manifestation of arbitrage risk, in which an open position fails to converge and realise a profit. Figure 4.5 displays the distribution of pair total returns for Group 2 and Group 3 pairs.

As stated in reference to Table 4.1, the marked increase in CTVHR mean monthly return from committed to employed capital can be attributed to a significant proportion of pairs that do not open a position during the trading period. This is further evidenced in the 47.56% Group 1 proportion of CTVHR pairs reported in Table 4.2. This statistic indicates that almost half of all CTVHR pairs fail to open a position during the trading period, in contrast to the ∼2% proportion of all other models whose pairs belong to Group 1. Assuming the arbitrageur were unable to deploy capital efficiently, the high proportion of Group 1 CTVHR pairs would represent a significant opportunity cost relative to the other models.

The proportions of Group 2 D and C pairs are similar at 29.41% and 35.14%, respectively. The proportions of DTVHR and CTVHR Group 2 pairs are also similar at 0.34% and 0.35%, respectively, though they are two orders of magnitude lower than their static D and C counterparts. Group 3 proportions are 68.95% for D pairs and 62.72% for C pairs, and 97.90% for DTVHR pairs and 52.09% for CTVHR pairs. The similarity of D and C Group 3 proportions is not mirrored in DTVHR and CTVHR pairs despite their almost identical Group 2 proportions. The arbitrageur would naturally favour a high proportion of Group 3 pairs over Group 2 pairs, as Group 3 represents the pairs that are able to identify and exploit multiple arbitrage opportunities, while Group 2 represents the pairs that enter a single non-convergent position during the trading period. Fortunately, all model specifications deliver a high proportion of Group 3 pairs, though the ratio relative to Group 2 pairs is greatest for TVHR variants.

The objective of the TVHR model is to deliver statistical arbitrage opportunities that do not succumb to spread non-convergence. The DTVHR and CTVHR Group 2 proportions, that are two orders of magnitude lower than their D and C counterparts, validate the fulfilment of this objective. Additionally, the TVHR variants exhibit a higher profitable proportion among Group 2 pairs, and a lower loss per non-convergent pair. TIM for TVHR variants is also much lower than the static alternatives, with CTVHR Group 2 pairs in the market 3.42% of the time compared to C Group 2 pairs being in the market 82.12% of the time. Industry matching is relatively consistent across all model variants. Formation period zero-crossings are reported under both distance and cointegration specifications, calculated with respect to the stochastic spreads governed by Equation (3.1) and Equation (3.2), respectively. Group 2 pairs under the DTVHR specification possess fewer zero-crossings than Group 2 D pairs. Similarly, Group 2 pairs under the CTVHR specification exhibit fewer formation period zero-crossings than Group 2 C pairs, though this observation is consistent with Group 3 pairs also.

The profitable proportion of Group 3 pairs is similar across D and C specifications at ~90%, and across DTVHR and CTVHR specifications at ~80%. Pair total returns are lower under TVHR specifications than static specifications, and average number of trades is fairly consistent across all specifications with the exception of DTVHR pairs which trade twice as frequently. As with Group 2 pairs, TIM for Group 3 pairs is much lower under TVHR specifications. However, with the exception of CTVHR pairs, Group 3 TIM is lower than Group 2 TIM, indicating that convergent pairs spend less time in the market than non-convergent pairs—an intuitive result owing to continued spread divergence among non-convergent pairs. Industry matching for Group 3 pairs is not too dissimilar to Group 2 pairs, nor are formation period zero-crossings with the exception of DTVHR pairs. Convergent Group 3 DTVHR pairs exhibit a greater number of formation period zero-crossings than non-convergent Group 2 pairs. This finding indicates that distance pairs benefit from having a greater number of formation period zero-crossings, a result that supports the augmented selection procedure proposed by Do and Faff (2010) in which pairs are first filtered in ascending order of spread SSD, and finally selected in descending order of zero-crossings.

BMD reports the average distance between trading period TVHR estimates and static formation period estimates of hedge ratio. The static D and C model variants naturally report a BMD of zero due to their retention of formation period hedge ratio estimates during the trading period. It is therefore expected that non-convergent pairs will exhibit a greater BMD than convergent pairs due to continued spread divergence as the hedge ratio shifts away from its formation period level. Empirically, Group 2 TVHR pairs do indeed report a higher BMD than Group 3 pairs, supporting the assertion that pair non-convergence can be attributed to a time-varying hedge ratio that is not accommodated by conventional static model specifications.

The characteristics of pairs under the various modelling approaches articulate the compromise between profitability and assurance of spread convergence. The TVHR variants almost always converge during the trading period, but deliver lower returns than conventional D and C model variants. Conversely, the conventional models deliver higher returns but are more susceptible to spread non-convergence. This presents a balancing act for the arbitrageur of maximising return while minimising non-convergence.

Table 4.3 reports the individual trade statistics for portfolios of the top 20 unrestricted pairs with immediate execution upon generation of a trading signal. Trades are further segmented by their outcome, with results reported for all trades, convergent trades, and non-convergent trades. The profitable proportion of all trades is approximately equal across all model variants at $\sim$75%. However, the TVHR variants deliver less than half the mean return of their static analogues. Despite their diminished mean returns, the TVHR models have much higher annualised Sharpe ratios—more than double the static distance pairs at 2.24, and approximately six-times the static cointegration pairs at 3.68—owing to their lower return standard deviations and brief trade lengths. Mean profit and loss is of approximately equal scale within model variants, though greater in magnitude for static models. All models, with the exception of CTVHR pairs, deliver a greater long return than short return, with only D trades making a loss on the short position. Mean and median trade lengths are approximately equal for static D and C models, and substantially greater than TVHR models. The median DTVHR trade length of four periods and CTVHR trade length of two periods illustrates the relative conservatism of the TVHR algorithm, completing round-trip transactions quickly and spending less time in the market than the conventional static models.

	D	DTVHR	C	CTVHR
All trades				
Profitable proportion	0.7539	0.7433	0.7421	0.7645
Mean return	0.0244	0.0099	0.0904	0.0423
Standard deviation	0.0670	0.0317	0.3663	0.1265
Sharpe ratio	0.9636	2.2405	0.6173	3.6830
Mean profit	0.0539	0.0219	0.2284	0.0772
Mean loss	-0.0661	-0.0249	-0.3068	-0.0712
Mean long return	0.0277	0.0073	0.0741	0.0201
Mean short return	-0.0033	0.0026	0.0162	0.0222
Mean trade length	36.0120	4.9116	40.2586	2.0742
Median trade length	22.0000	4.0000	24.0000	2.0000
Convergent trades	0.6403	0.9529	0.6140	0.9791
Profitable proportion	1.0000	0.7586	1.0000	0.7686
Mean return	0.0596	0.0108	0.2525	0.0431
Standard deviation	0.0235	0.0308	0.1689	0.1272
Sharpe ratio	8.5323	2.5166	5.0235	3.7334
Mean profit	0.0596	0.0221	0.2525	0.0776
Mean loss	0.0000	-0.0246	0.0000	-0.0715
Mean long return	0.0388	0.0076	0.1456	0.0205
Mean short return	0.0208	0.0032	0.1070	0.0226
Mean trade length	22.3278	4.9084	22.3208	2.0767
Median trade length	15.0000	4.0000	14.0000	2.0000
Non-convergent trades	0.3597	0.0471	0.3860	0.0209
Profitable proportion	0.3157	0.4359	0.3317	0.5728
Mean return	-0.0382	-0.0084	-0.1676	0.0037
Standard deviation	0.0732	0.0421	0.4402	0.0830
Sharpe ratio	-1.0668	-1.4224	-0.7287	0.5085
Mean profit	0.0221	0.0156	0.1127	0.0533
Mean loss	-0.0661	-0.0270	-0.3068	-0.0627
Mean long return	0.0079	0.0012	-0.0395	0.0003
Mean short return	-0.0461	-0.0096	-0.1281	0.0034
Mean trade length	60.3686	4.9756	68.7954	1.9570
Median trade length	59.0000	3.0000	70.0000	2.0000

Table 4.3: Trade statistics for top 20 unrestricted pairs, no execution delay, July 1962–June 2009.

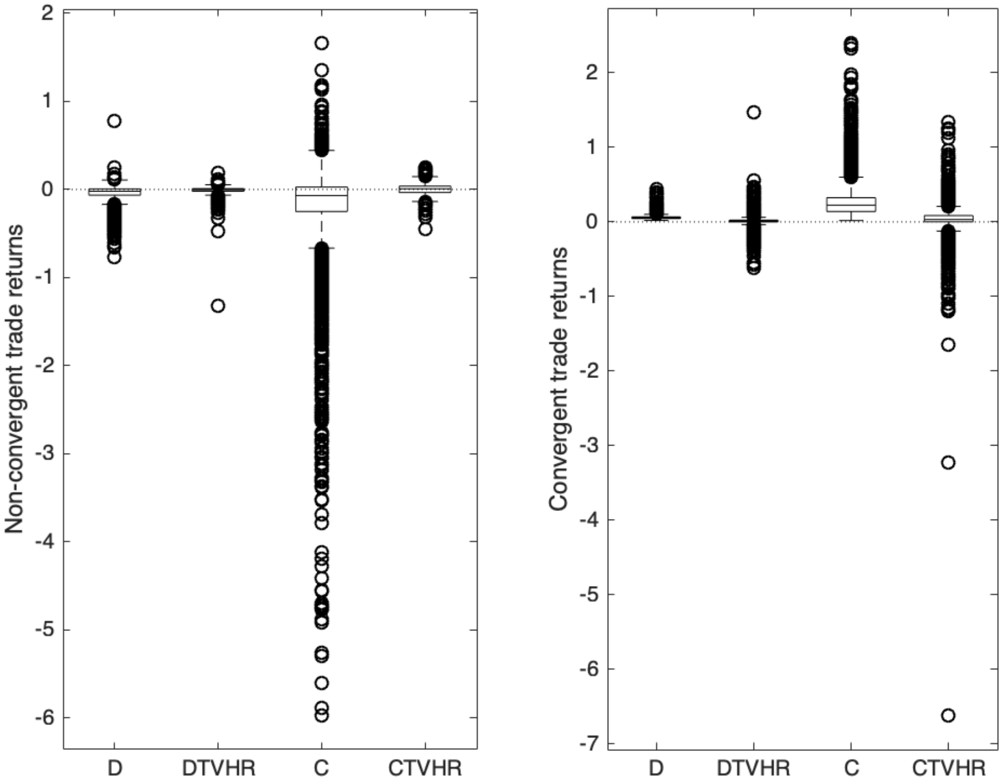

Figure 4.6: Distribution of trade excess return of top 20 unrestricted pairs, employed capital, no execution delay, July 1962–June 2009.

Figure 4.6 displays the distribution of trade returns for convergent and non-convergent trades. The proportion of convergent trades under the TVHR model is approximately equal for distance and cointegration pairs at ∼95%—approximately 50% greater than the proportion of convergent trades for static pairs. The profitable proportion under static pairs, however, is 100%, while the profitable proportion for TVHR pairs is ∼75%. The superiority of conventional static pairs in terms of profitable proportion is attributable to the construction of the tradable spread; if a conventional pair converges it will always return a profit. TVHR pairs, on the other hand, are not guaranteed to return a profit if spread convergence is facilitated chiefly by the time-varying correction of the hedge ratio. The standard deviation of convergent trade returns is lower under conventional D and C models than TVHR variants, leading to greater annualised Sharpe ratios. Mean loss is zero under static models, and mean profit is substantially greater than that reported for TVHR models. Mean and median trade length are shorter for convergent D and C trades than non-convergent trades, though not as short as DTVHR and CTVHR pairs which remain consistent across convergent and non-convergent trades.

The profitable proportion of non-convergent trades is higher under TVHR model specifications, while losses are not as pronounced. CTVHR pairs even manage to deliver a small positive return for non-convergent trades. Despite possessing a less extreme mean loss than non-convergent D trades, the non-convergent DTVHR trades delivered a less desirable Sharpe ratio due to the influence of its short mean trade length—the ratio of mean return to standard deviation is multiplied by $(252/4.9756)^{1/2}$ in order to annualise the Sharpe ratio, which is greater than the factor used to annualise the Sharpe ratio of D non-convergent trades. The mean loss of all model variants is greater in magnitude than the mean profit, and all models with the exception of CTVHR pairs saw non-convergent trade losses driven by the short leg of pairs. Mean and median trade length for D and C pairs is between three- and five-times greater than the length of convergent trades, while DTVHR and CTVHR trade lengths remain relatively consistent across trade outcomes.

The decomposition of trade statistics further confirms the nimble dynamics of the TVHR model. Time-varying estimation of the hedge ratio allows the model to exit trade positions immediately after extracting small amounts of profit, minimising arbitrage risk by spending little time in the market. The small proportion of non-convergent trades indicates the ability of the model to ameliorate the risk of spread non-convergence, albeit at the expense of greater profits. Results so far reported consider pairs that enter and exit a position immediately upon generation of a trading signal. Both Gatev, Goetzmann, and Rouwenhorst (2006) and Do and Faff (2010) impose a one-day execution delay to alleviate concerns regarding the potential upward bias in reported returns induced by bid-ask bounce. All remaining results in this thesis consider this execution delay and the detrimental effect it has on the proposed TVHR model.

	Committed capital				Employed capital			
	D	DTVHR	C	CTVHR	D	DTVHR	C	CTVHR
Mean	0.0059	0.0037	0.0117	0.0021	0.0060	0.0038	0.0122	0.0043
t-Statistic	10.2631	11.9812	3.2235	5.3471	10.3431	11.9502	3.2966	4.9680
Median	0.0054	0.0033	0.0120	0.0009	0.0055	0.0034	0.0127	0.0017
Standard deviation	0.0095	0.0051	0.1061	0.0094	0.0096	0.0052	0.1074	0.0206
Skewness	0.2406	1.5329	-3.2862	0.5003	0.2136	1.5493	-3.1689	1.5405
Kurtosis	7.6113	9.8317	40.7874	12.5947	7.4189	9.8133	39.2870	15.9644
Minimum	-0.0502	-0.0093	-1.2052	-0.0562	-0.0502	-0.0098	-1.2052	-0.0978
Maximum	0.0570	0.0364	0.4210	0.0605	0.0570	0.0367	0.4210	0.1503
Observations < 0	0.2482	0.1920	0.3551	0.3043	0.2482	0.1920	0.3551	0.3043
Lower semi-deviation	0.0039	0.0014	0.0809	0.0054	0.0040	0.0015	0.0814	0.0108
Upper semi-deviation	0.0105	0.0062	0.0695	0.0080	0.0106	0.0063	0.0709	0.0180
Sharpe ratio	2.1692	2.5032	0.3837	0.7763	2.1669	2.4913	0.3922	0.7176
Sortino ratio	5.2615	9.0504	0.5030	1.3533	5.2265	8.9871	0.5171	1.3718
TIM	0.6270	0.1973	0.6729	0.0225	0.6270	0.1973	0.6729	0.0225
Return/TIM	0.0094	0.0188	0.0175	0.0941	0.0096	0.0191	0.0181	0.1900

Table 4.4: Excess return statistics for portfolios of top 20 unrestricted pairs, execution delay, July 1962–June 2009.

Table 4.4 reports the excess return statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal. Under the imposition of the execution delay, mean monthly returns for TVHR models are less than half for distance pairs and less than one-quarter for cointegration pairs relative to the results reported in Table 4.1. Conventional static models, on the other hand, experience a more modest depreciation in mean return. While mean monthly return of the DTVHR model was comparable to that of the D model with immediate execution, there now exists a greater divide with a DTVHR mean return of 0.38% and a D mean return of 0.60%. Similarly, CTVHR mean return is now one-third the mean return of the C model, where it was previously only one-half with immediate execution. Figure 4.7 displays the cumulative excess return of unrestricted distance and cointegration pairs under employed capital with execution delay.

Mean return t-statistics still favour the TVHR model specification, though the magnitude of the outperformance has diminished. Return standard deviation, skewness, minimum, and proportion of negative monthly returns continue to favour the TVHR model, while median return reflects the depressive influence of the execution delay. Of particular note is the effect of the execution delay on the static C model mean return, which lost more than half its monthly return and now delivers 1.22% down from 3.12%. This indicates that a large proportion of the cointegration model's return is realised immediately following signal generation when spread reverts most strongly. Figure 4.8 displays the distribution of monthly excess returns for distance and cointegration portfolios under employed capital with execution delay.

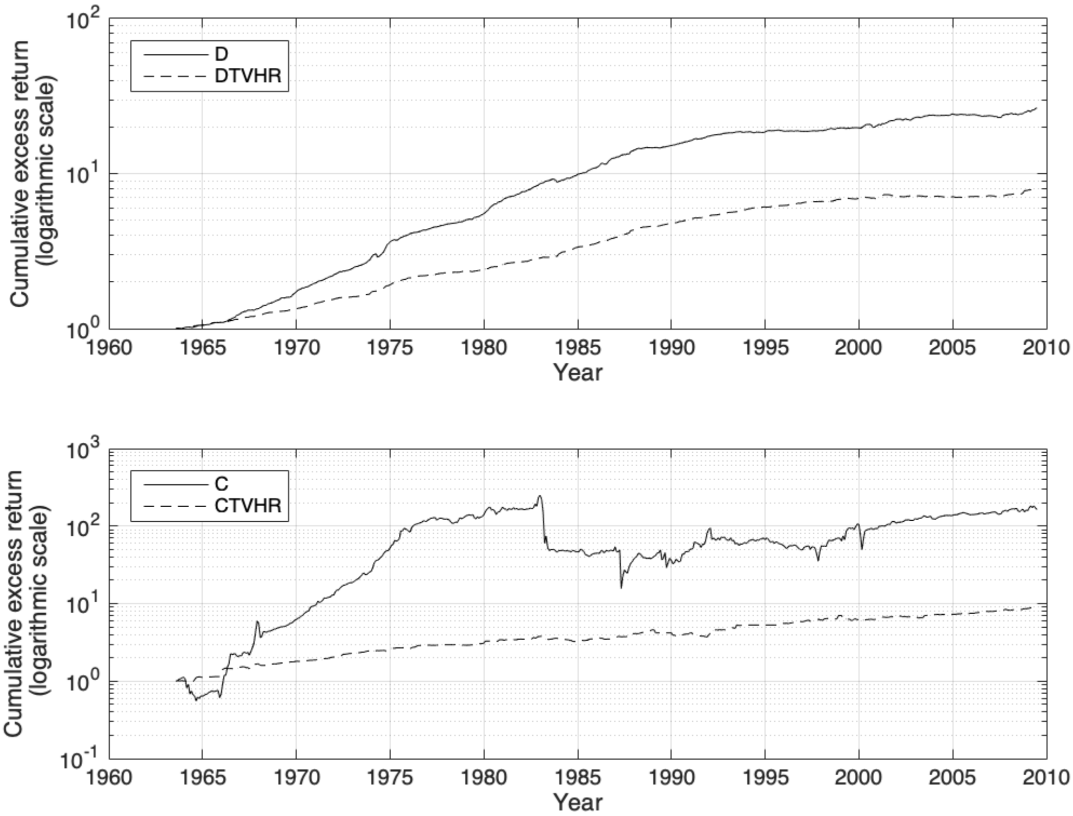

Figure 4.7: Cumulative excess return of top 20 unrestricted pairs, employed capital, execution delay, July 1962–June 2009.

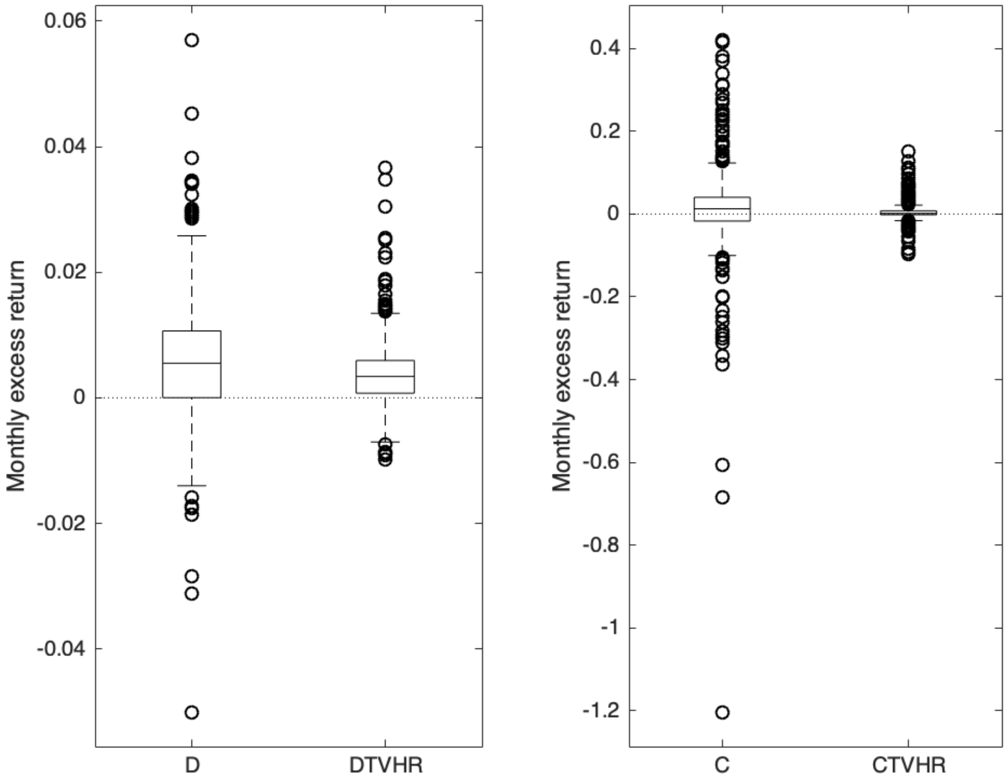

Figure 4.8: Distribution of monthly excess return of top 20 unrestricted pairs, employed capital, execution delay, July 1962–June 2009.

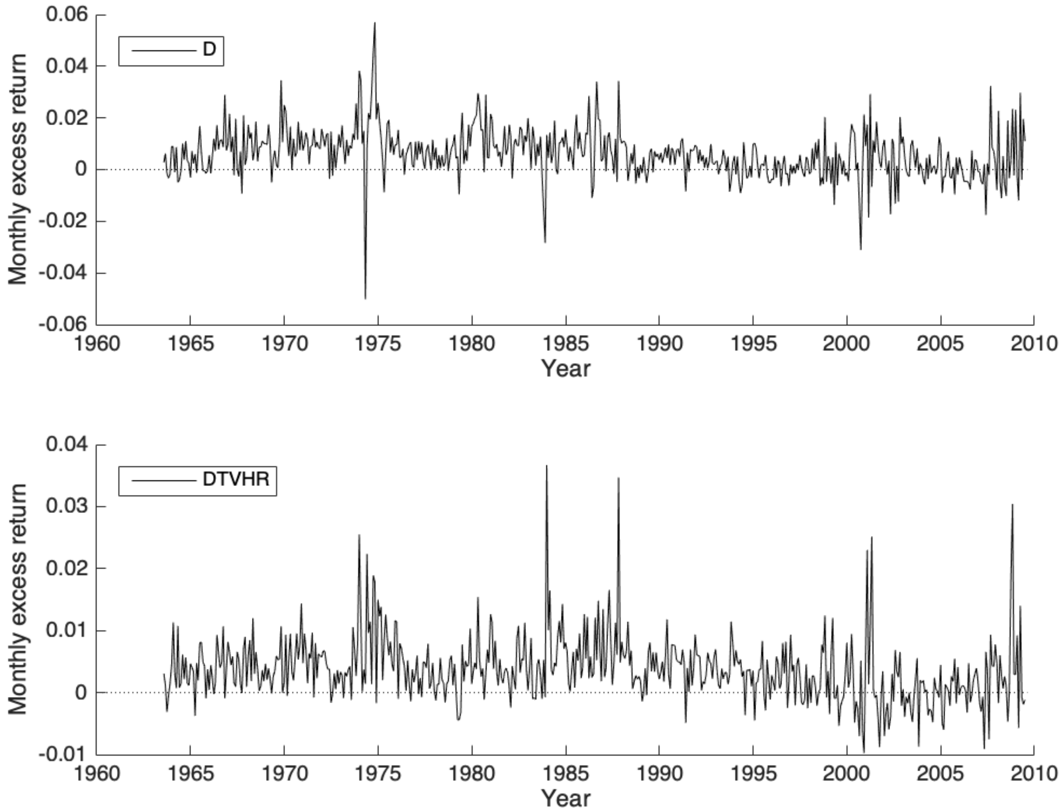

Figure 4.9: Monthly excess return of top 20 distance pairs, employed capital, execution delay, July 1962–June 2009.

The Sharpe and Sortino ratios continue to favour TVHR model variants, though the difference is not as substantial. DTVHR Sharpe ratio declined from 4.14 to 2.49 with the inclusion of an execution delay, and CTVHR Sharpe ratio declined from 2.19 to 0.72. More significant is the decline in Sortino ratio, which sees a decline from 19.12 to 8.99 for the DTVHR model, and from 7.01 to 1.37 for the CTVHR model. TIM remains identical since the opening and closing of positions is shifted back one day, but the lower return diminishes the return per unit TIM across all model specifications, with TVHR models most heavily affected. CTVHR portfolios, in particular, decline from a return per unit TIM of 76.53% to 19.00%. Figure 4.9 displays the monthly excess return for top 20 distance pairs, and Figure 4.10 displays the monthly excess return for top 20 cointegration pairs, both with execution delay.

Table 4.5 reports the pair statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal. The Group 1, 2 and 3 proportions differ marginally from those reported in Table 4.2, though the Group 2 proportions for TVHR variants remain two orders of magnitude lower than their conventional static counterparts. The imposition of the execution delay principally serves to deflate the total

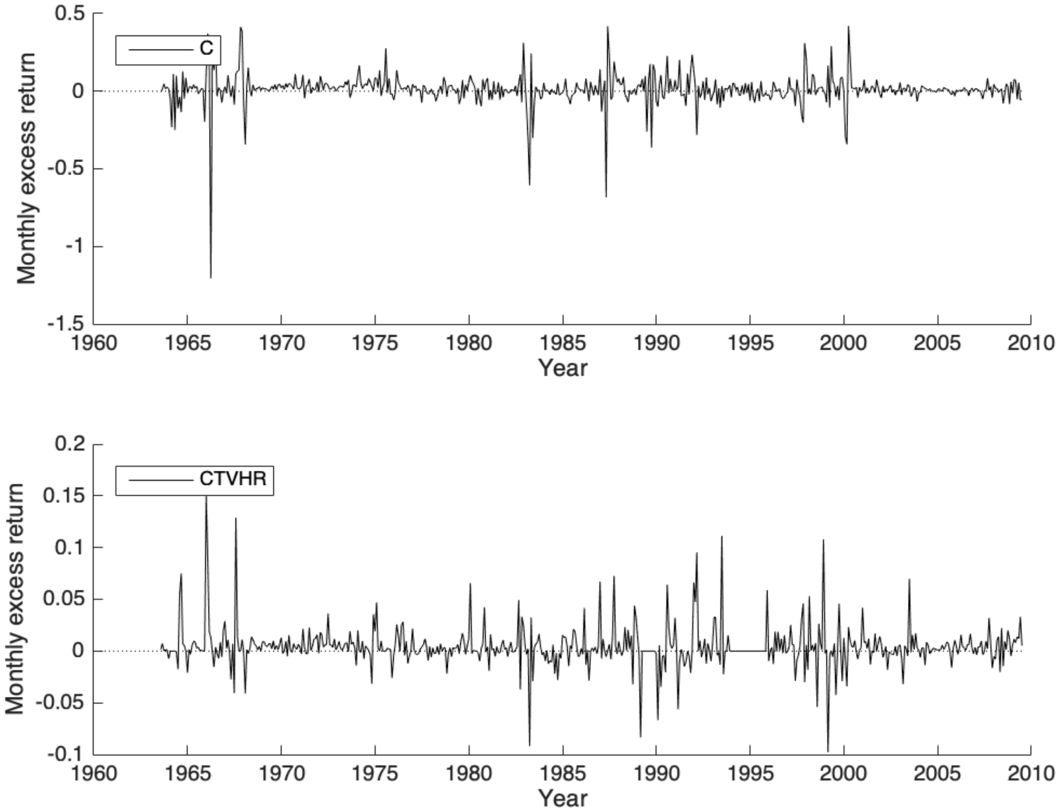

Figure 4.10: Monthly excess return of top 20 cointegration pairs, employed capital, execution delay, July 1962–June 2009.

	D	DTVHR	C	CTVHR
Group 1 proportion	0.0165	0.0177	0.0224	0.4757
Group 2 proportion	0.2942	0.0037	0.3503	0.0038
Profitable proportion	0.2101	0.4000	0.2224	0.5610
Total return	-0.0602	-0.0293	-0.2521	-0.0136
Number of trades	1.0000	1.0000	1.0000	1.0000
TIM	0.7638	0.2323	0.8220	0.0365
Industry-matched	0.1331	0.1000	0.1278	0.1220
Distance zero-crossings	40.5935	33.3250	35.5718	35.6829
Cointegration zero-crossings	42.7089	31.9000	78.3219	72.1463
BMD	0.0000	0.7122	0.0000	4.7811
Group 3 proportion	0.6893	0.9786	0.6273	0.5206
Profitable proportion	0.8575	0.6549	0.8104	0.5909
Total return	0.0773	0.0231	0.2551	0.0235
Number of trades	2.7233	5.1054	2.7183	2.5133
TIM	0.5805	0.1998	0.6094	0.0422
Industry-matched	0.0996	0.1106	0.1349	0.1438
Distance zero-crossings	43.4153	42.2964	36.4415	32.4754
Cointegration zero-crossings	46.1226	44.7210	77.0592	71.0931
BMD	0.0000	0.5103	0.0000	1.3221

Table 4.5: Pair statistics for top 20 unrestricted pairs, execution delay, July 1962–June 2009.

return of Group 3 pairs across all model variants. While the total return of Group 3 D pairs dropped from an average of 10.23% to 7.73%, DTVHR pair total returns more than halved from 5.09% to 2.31%. Group 3 C pairs exhibit a substantial reduction in total return from an average of 43.18% to 25.51%, and CTVHR pairs lose more than three-quarters of their Group 3 total return, going from 10.68% to 2.35% under imposition of the execution delay. Accordingly, the profitable proportion of Group 3 pairs across all model variants is reduced relative to the results reported in Table 4.2, with the most significant reduction apparent in CTVHR pairs.

Group 2 proportion statistics remain relatively consistent with those reported in Table 4.2. The total return and, consequently, the profitable proportion under D, DTVHR, and C model variants are stable, though CTVHR Group 2 pairs become marginally less unprofitable reporting a change in total return from -2.56% to -1.36%, and an increase in profitable proportion from 42.11% to 56.10%. Group 2 DTVHR pairs are approximately half as unprofitable as Group 2 D pairs, but a substantial reduction in Group 3 profitability finds DTVHR pairs only one-third as profitable as D pairs. Similarly, while CTVHR pairs possessed one-quarter the profitability of C pairs without execution delay, CTVHR Group 3 pairs are one-tenth as profitable as C pairs with the execution delay. Figure 4.11 displays the distribution of pair total returns for Group 2 and Group 3 pairs with one-day execution

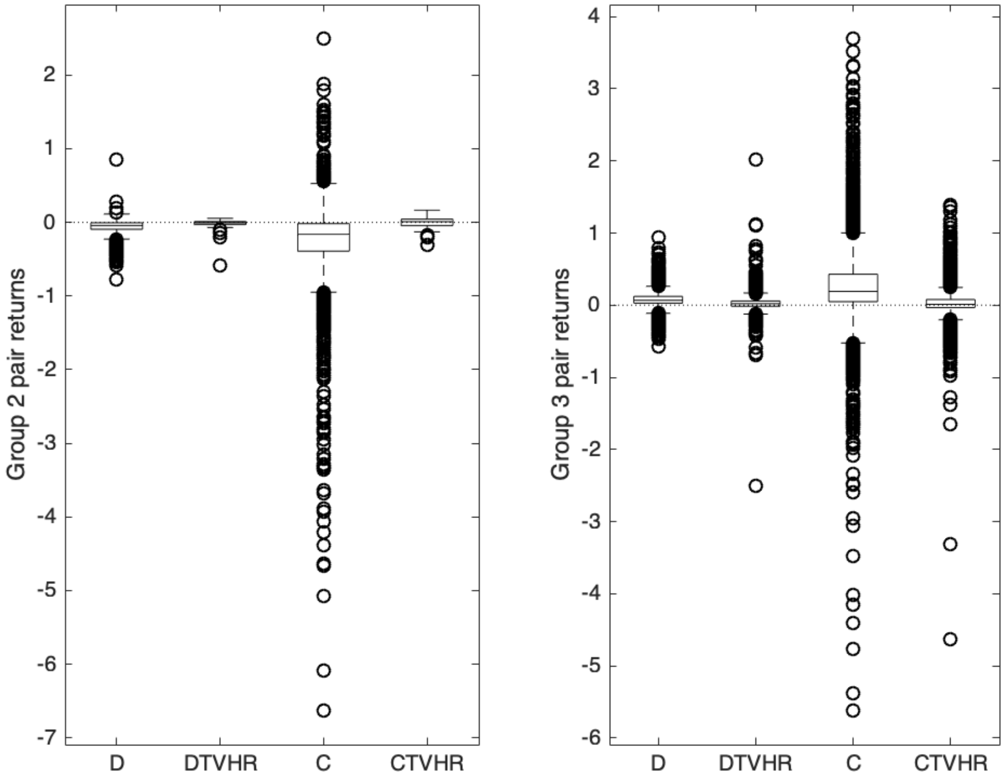

Figure 4.11: Distribution of pair excess return of top 20 unrestricted pairs, employed capital, execution delay, July1962–June 2009.

delay.

Table 4.6 reports the individual trade statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal. The proportions of convergent and non-convergent trades are almost identical to those reported in Table 4.3, however the reduction in mean return among convergent trades has a significant effect on pair profitability. The profitable proportion for all trades is reduced by ∼2% for D pairs, ∼13% for DTVHR pairs, ∼8% for C pairs, and ∼19% for CTVHR pairs. The profitable proportion among all trades is now highest for D pairs at a rate of 72.94%, and lowest under CTVHR pairs at a rate of 57.32%. The mean return across all trades under imposition of the execution delay most significantly impacts TVHR model specifications, with DTVHR mean return dropping from 0.99% to 0.45%, and CTVHR mean return dropping from 4.23% to 0.96%. Static D pair mean return is less severely impacted, dropping from 2.44% to 1.66%, while static C pair mean return drops from 9.04% to 3.53%. Standard deviation of returns remains relatively consistent with and without the execution delay, so the reduction in Sharpe ratio is attributable to the reduction in mean return. The highest Sharpe ratio is observed under DTVHR pairs, having reduced from

2.24 to 0.98, while the smallest Sharpe ratio is observed under static C pairs, having reduced from 0.62 to 0.24. All model specifications exhibit a reduction in both long and short mean return, though TVHR model variants return a slight profit for short positions while static variants make a loss on short positions.

Figure 4.12 displays the distribution of trade returns for convergent and non-convergent trades with one-day execution delay. While no execution delay previously allowed static D and C pairs to achieve a profitable proportion of 100% among convergent trades, the imposition of the execution delay reduces D profitable proportion to 96.57% and C profitable proportion to 90.24%. TVHR pairs experience a more substantial reduction in profitable proportion, with DTVHR pairs dropping from 75.86% to 62.30%, and CTVHR pairs dropping from 76.86% to 57.36%. Mean return for convergent trades is lower for all model variants under the execution delay, with D pairs dropping from 5.96% to 4.74%, DTVHR pairs dropping from 1.08% to 0.51%, C pairs dropping from 25.25% to 16.82%, and CTVHR pairs dropping from 4.31% to 0.97%. Standard deviation of convergent trade returns increased for static pairs but remained stable for TVHR pairs, with Sharpe ratios updated accordingly. The greatest reduction in convergent trade Sharpe ratio is observed under CTVHR pairs, going from 3.73 under no execution delay to 0.94. All pairs experience a reduction in mean long and short return for convergent trades, though the greatest reduction in percentage terms is observed under DTVHR short positions, declining from 0.32% to 0.05%.

The profitable proportion of non-convergent trades is similar to the proportions reported in Table 4.3 for D and DTVHR pairs, though C and CTVHR pairs experience a slight reduction in profitable proportion of 2–4%. D, DTVHR, and C pairs report similar mean returns, standard deviations, and Sharpe ratios relative to those reported without execution delay. CTVHR pairs, however, exhibit a substantial increase in mean return from 0.37% to 0.63%, accompanied by a reduction in return standard deviation and consequent increase in Sharpe ratio. Mean profit and loss across all model variants is stable for non-convergent trades relative to Table 4.3, as are mean long and short return with the exception of C and CTVHR long return.

The imposition of a one-day execution delay has a significant impact on the profitability of the proposed TVHR model. The short trade lengths that typify the model's dynamics extract small amounts of profit in the few days following the generation of a trading signal, closing positions shortly thereafter as hedge ratios adaptively move toward new equilibrium levels. The execution delay is intended to account for the potential upward bias induced by bid-ask bounce. As discussed by Gatev, Goetzmann, and Rouwenhorst (2006), it is difficult to quantify what proportion of the reduction in profit can be attributed to bid-ask bounce and what can be attributed to the opportunity cost of

	D	DTVHR	C	CTVHR
All trades				
Profitable proportion	0.7294	0.6142	0.6676	0.5732
Mean return	0.0166	0.0045	0.0353	0.0096
Standard deviation	0.0655	0.0327	0.3624	0.1128
Sharpe ratio	0.6689	0.9821	0.2436	0.9421
Mean profit	0.0456	0.0202	0.1854	0.0603
Mean loss	-0.0617	-0.0205	-0.2662	-0.0584
Mean long return	0.0240	0.0044	0.0408	0.0059
Mean short return	-0.0074	0.0001	-0.0055	0.0037
Mean trade length	36.0138	4.9109	40.2651	2.0742
Median trade length	22.0000	4.0000	24.0000	2.0000
Convergent trades	0.6402	0.9528	0.6144	0.9791
Profitable proportion	0.9657	0.6230	0.9024	0.5736
Mean return	0.0474	0.0051	0.1682	0.0097
Standard deviation	0.0326	0.0320	0.2156	0.1134
Sharpe ratio	4.8781	1.1405	2.6207	0.9432
Mean profit	0.0495	0.0203	0.1948	0.0605
Mean loss	-0.0134	-0.0201	-0.0780	-0.0586
Mean long return	0.0327	0.0046	0.1014	0.0059
Mean short return	0.0146	0.0005	0.0668	0.0038
Mean trade length	22.3312	4.9085	22.3208	2.0768
Median trade length	15.0000	4.0000	14.0000	2.0000
Non-convergent trades	0.3598	0.0472	0.3856	0.0209
Profitable proportion	0.3089	0.4353	0.2937	0.5530
Mean return	-0.0383	-0.0078	-0.1764	0.0063
Standard deviation	0.0731	0.0429	0.4399	0.0775
Sharpe ratio	-1.0694	-1.2985	-0.7670	0.9177
Mean profit	0.0237	0.0160	0.1393	0.0520
Mean loss	-0.0660	-0.0261	-0.3076	-0.0503
Mean long return	0.0085	0.0010	-0.0558	0.0040
Mean short return	-0.0467	-0.0088	-0.1206	0.0022
Mean trade length	60.3595	4.9597	68.8517	1.9503
Median trade length	59.0000	3.0000	70.0000	2.0000

Table 4.6: Trade statistics for top 20 unrestricted pairs, execution delay, July 1962–June 2009.

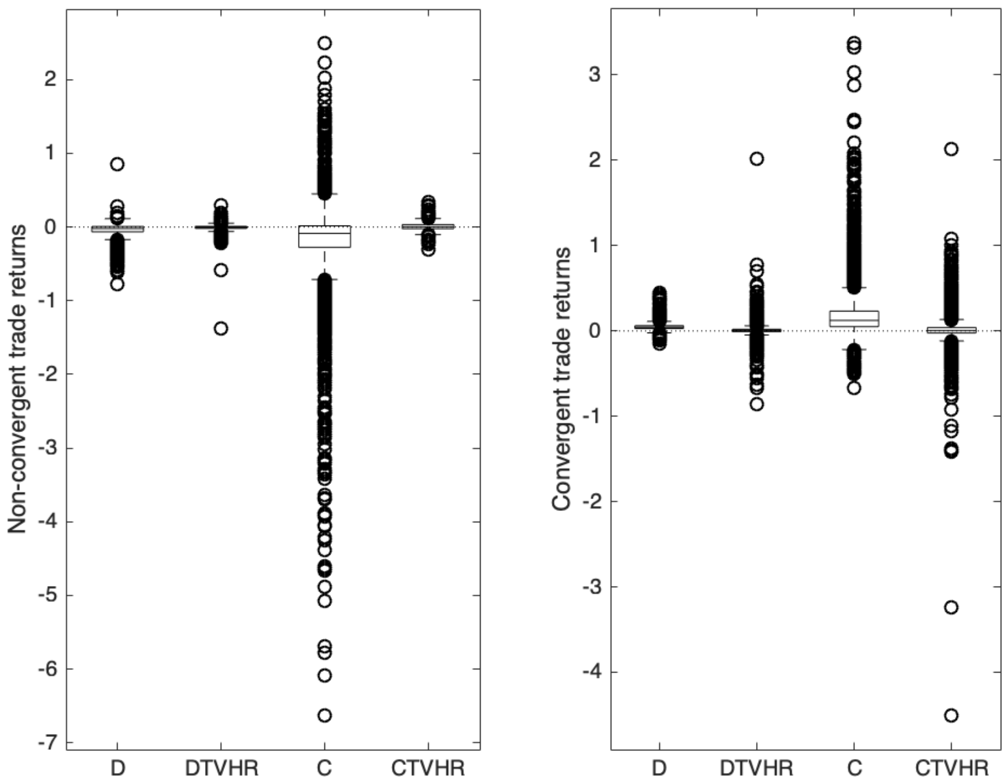

Figure 4.12: Distribution of trade excess return of top 20 unrestricted pairs, employed capital, execution delay, July 1962–June 2009.

	Industrials				Transportation			
	D	DTVHR	C	CTVHR	D	DTVHR	C	CTVHR
Mean	0.0065	0.0098	0.0105	0.0093	0.0027	0.0095	0.0018	0.0125
t-Statistic	7.8270	10.9706	3.8353	8.7890	2.4404	9.6298	0.9072	10.2433
Median	0.0064	0.0089	0.0115	0.0068	0.0030	0.0080	0.0031	0.0105
Standard deviation	0.0180	0.0144	0.0720	0.0242	0.0260	0.0197	0.0419	0.0256
Skewness	0.2612	0.1129	-0.9177	1.1270	-0.1575	-0.0422	-0.3658	3.1930
Kurtosis	4.4865	11.3486	17.8865	12.3020	6.1268	6.9479	6.6781	41.6931
Minimum	-0.0724	-0.0935	-0.5834	-0.1072	-0.1337	-0.0984	-0.1848	-0.1330
Maximum	0.0810	0.1021	0.4572	0.1922	0.1198	0.0883	0.2073	0.3183
Observations < 0	0.3569	0.2029	0.3967	0.2862	0.4293	0.2699	0.4565	0.2554
Lower semi-deviation	0.0090	0.0059	0.0494	0.0115	0.0174	0.0095	0.0300	0.0096
Upper semi-deviation	0.0169	0.0163	0.0533	0.0232	0.0195	0.0197	0.0292	0.0268
Sharpe ratio	1.2513	2.3572	0.5079	1.3355	0.3557	1.6741	0.1502	1.6956
Sortino ratio	2.4988	5.7107	0.7403	2.8186	0.5318	3.4744	0.2096	4.5139
TIM	0.6068	0.2539	0.6809	0.0900	0.5048	0.2595	0.6237	0.2233
Return/TIM	0.0107	0.0385	0.0155	0.1035	0.0053	0.0366	0.0029	0.0561

Table 4.7: Excess return statistics for portfolios of top 20 restricted Industrials and Transportation pairs, employed capital and execution delay, July 1962–June 2009.

foregoing strong spread reversion immediately following generation of a trading signal, though the influence of bid-ask bounce is likely non-trivial. The execution delay therefore offers a very conservative estimate of statistical arbitrage profitability, though one which still demonstrates the favourable risk-adjusted returns of the proposed TVHR model.

Results reported until now have considered the profitability of statistical arbitrage opportunities offered by the unrestricted matching of pair candidates. A bank stock, for example, might be matched with a fashion label stock, or mining stock with an information technology stock. Matching can instead be restricted to consider only those pairs whose constituents come from the same industry group. Do and Faff (2010) consider restricted pairs coming from either Industrial, Transportation, Utility, or Financial industry groups, classified as such by the first two digits of each security's SIC code: 15–17 and 30–39 for Industrials, 40–47 for Transportation, 49 for Utilities, and 60–67 for Financials. Table 4.7 reports the excess return statistics for portfolios of the top 20 restricted Industrials and Transportation pairs with one-day execution delay upon generation of a trading signal.

The excess return statistics for portfolios of top 20 Industrials pairs demonstrates the utility of pair restriction. Mean monthly return has risen from 0.60% to 0.65% for D pairs, from 0.38% to 0.98% for DTVHR pairs, and from 0.43% to 0.93% for CTVHR pairs. C pairs exhibit a slight decline in mean monthly return from 1.22% to 1.05%, accompanied by a decline in return standard deviation and modest increase in Sharpe and Sortino ratios. Median return and standard deviation are higher for restricted Industrials pairs across all but the static C model, whose skewness also increased substantially from −3.17 to

−0.92. The minimum monthly return—that is, the maximum one-month loss—increased in absolute terms for the D, DTVHR, and CTVHR models, with the DTVHR model maximum loss increasing by one order of magnitude from −0.98% to −9.35%. Maximum monthly return also increased for all model variants, while the proportion of negative monthly return observations increased for all but the CTVHR model. Sharpe and Sortino ratios declined for the D and DTVHR models but increased for the C and CTVHR models, while return per unit TIM increased for the D and DTVHR models but declined for the C and CTVHR models. The effect of restricting pairs to Industrials is to inflate monthly mean return at the cost of risk-adjusted return for the D, DTVHR and CTVHR models, and vice versa for the C model.

Restriction to Transportation pairs facilitates substantial outperformance of static D and C models by their TVHR counterparts. Mean and median monthly return are greater for DTVHR pairs than D pairs, and for CTVHR pairs than C pairs, with CTVHR pairs delivering the greatest portfolio excess return of 1.25% per month. Return t-statistics reflect the magnitude of portfolio excess return, with DTVHR and CTVHR model variants exhibiting greater t-statistics than their static model analogues. Return standard deviation is higher among Transportation restricted pairs than Industrials pairs for all model variants with the exception of the static C model, and skewness is more positive for C and CTVHR models but more negative for D and DTVHR models. Relative to unrestricted pairs, the outperformance of DTVHR and CTVHR Transportation pairs is largely attributable to a reduction in profitability observed for D and C Transportation pairs. Risk-adjusted returns for D, DTVHR and C model variants declined relative to unrestricted pair portfolios, though DTVHR pairs report a substantially greater absolute return. CTVHR pairs, by contrast, are more profitable in both absolute and risk-adjusted terms under restriction to Transportation securities.

Table 4.8 reports the excess return statistics for portfolios of the top 20 restricted Utilities and Financials pairs with one-day execution delay upon generation of a trading signal. Mean monthly return is greater for both Utilities and Financials restricted pairs under D, DTVHR, and CTVHR model variants, while static C pairs experience a decline in profitability relative to unrestricted pairs. Return t-statistics of Utilities pairs are higher for all model variants, while t-statistics for Financials pairs are lower for all model variants relative to unrestricted pairs. This is despite Financials pairs delivering higher mean returns than Utilities pairs, indicating an increase in return standard deviation for Financials pairs. Skewness, kurtosis, minimum and maximum monthly return are all greater in magnitude among Financials pairs than Utilities pairs; for example, the kurtosis of CTVHR returns is 20.22 for Utilities pairs and 188.37 for Financials pairs, and the minimum monthly return of C returns is −40.43% for Utilities pairs and −116.40%

	Utilities				Financials			
	D	DTVHR	C	CTVHR	D	DTVHR	C	CTVHR
Mean	0.0066	0.0056	0.0078	0.0054	0.0075	0.0074	0.0046	0.0091
t-Statistic	11.3380	13.8706	4.8557	7.3777	7.9619	10.1920	1.2702	3.9730
Median	0.0058	0.0050	0.0064	0.0043	0.0064	0.0052	0.0064	0.0051
Standard deviation	0.0132	0.0097	0.0421	0.0167	0.0216	0.0139	0.0931	0.0515
Skewness	0.2035	0.8552	2.5205	1.0538	1.5597	1.5832	-5.8526	9.9109
Kurtosis	5.2371	8.3605	65.7720	20.2247	41.5562	18.2315	74.4270	188.3360
Minimum	-0.0482	-0.0305	-0.4043	-0.1051	-0.1917	-0.0709	-1.1640	-0.2974
Maximum	0.0694	0.0611	0.4982	0.1482	0.2111	0.1252	0.4966	0.9279
Observations < 0	0.2699	0.1938	0.3442	0.2663	0.2717	0.2029	0.3841	0.2790
Lower semi-deviation	0.0060	0.0040	0.0239	0.0088	0.0112	0.0055	0.0781	0.0224
Upper semi-deviation	0.0134	0.0105	0.0355	0.0151	0.0199	0.0148	0.0506	0.0472
Sharpe ratio	1.7382	2.0223	0.6392	1.1166	1.2098	1.8401	0.1719	0.6103
Sortino ratio	3.7816	4.9329	1.1259	2.1082	2.3351	4.6279	0.2048	1.4010
TIM	0.5251	0.2344	0.5848	0.1750	0.5915	0.2436	0.6572	0.1286
Return/TIM	0.0126	0.0241	0.0133	0.0307	0.0128	0.0304	0.0070	0.0705

Table 4.8: Excess return statistics for portfolios of top 20 restricted Utilities and Financials pairs, employed capital and execution delay, July 1962–June 2009.

for Financials pairs. Proportion of negative monthly return observations is relatively consistent across Utilities, Financials, and unrestricted pairs.

Despite the higher mean monthly return delivered by most model specifications across the different industry group restrictions, neither D nor DTVHR model variants delivered a higher Sharpe or Sortino ratio than unrestricted pairs, though this is likely due to the loss of diversification caused by industry restriction. Only the CTVHR model was able to deliver both higher mean returns and Sharpe and Sortino ratios across all industry restrictions with the exception of Financials, whose Sharpe ratio was marginally lower. The selection algorithm for cointegration pairs favours pairings that exhibit high spread volatility during the formation period. This criteria assumes but does not ensure a shared stochastic relationship among pair candidates, with spurious relationships indistinguishable from real ones. The improved performance of the CTVHR model is likely attributable to the presence of stochastic commonality among pair candidates, that would not otherwise be present among unrestricted pairings, imposed by industry restriction. However, that does not explain the relatively poor performance of the static C model. A possible explanation is the inability of the static C model to trade broader market noise due to industry restriction. Pair restriction removes some of the most volatile inter-industry pairings from consideration, and the static C model lacks the time-varying infrastructure that allows the CTVHR model to exploit the more stable restricted pairs.

	D			DTVHR		
	1962–1988	1989–2002	2003–2009	1962–1988	1989–2002	2003–2009
Mean	0.0089	0.0027	0.0019	0.0050	0.0027	0.0012
t-Statistic	11.9712	4.5860	1.8062	13.2189	5.8252	1.4931
Median	0.0087	0.0027	0.0001	0.0042	0.0030	0.0006
Standard deviation	0.0097	0.0076	0.0092	0.0050	0.0047	0.0056
Skewness	-0.0306	-0.3560	1.1965	2.0817	0.8147	2.1519
Kurtosis	9.8806	5.7279	4.8271	11.8581	6.9620	11.8357
Minimum	-0.0502	-0.0311	-0.0175	-0.0044	-0.0098	-0.0091
Maximum	0.0570	0.0292	0.0324	0.0367	0.0251	0.0304
Observations < 0	0.1340	0.3393	0.5000	0.1046	0.2560	0.3974
Lower semi-deviation	0.0037	0.0043	0.0042	0.0006	0.0019	0.0025
Upper semi-deviation	0.0126	0.0068	0.0083	0.0071	0.0051	0.0052
Sharpe ratio	3.1574	1.2307	0.7014	3.4251	2.0047	0.7234
Sortino ratio	8.2510	2.1651	1.5491	29.8587	5.0788	1.6484
TIM	0.6093	0.6462	0.6549	0.1750	0.2124	0.2523
Return/TIM	0.0146	0.0042	0.0028	0.0285	0.0129	0.0047

Table 4.9: Excess return statistics for portfolios of top 20 unrestricted distance pairs, employed capital and execution delay, July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009.

4.1.2 Initial Study Sub-Periods

The initial study sub-periods investigated in this thesis cover the sample periods used by Do and Faff (2010), specifically July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009. Both Gatev, Goetzmann, and Rouwenhorst (2006) and Do and Faff (2010) sought to understand the evolution of statistical arbitrage profitability, from its unfettered exploitation of market inefficiencies after 1962, its challenges in an increasingly competitive and hedge fund-saturated market after 1989, through its out-of-sample investigation after 2003. Table 4.9 reports the excess return statistics for portfolios of the top 20 D and DTVHR unrestricted pairs across the three sub-periods.

Mean monthly return declines for both D and DTVHR models between subsequent sub-periods. The greatest mean return of 0.89% is achieved by D pairs between 1962 and 1988, declining by 70% to 0.27% between 1989 and 2002, and a further 30% to 0.19% between 2003 and 2009. The declining profitability of the DTVHR model is marginally more modest between the first two sub-periods, declining 46% from a mean return of 0.50% to 0.27%, and a further 56% to 0.12% in the final sub-period. Return t-statistics favour the DTVHR model for all but the final sub-period, while median return favours the DTVHR model in all but the first sub-period. Standard deviation, skewness, minimum monthly return, and proportion of negative monthly returns all favour the DTVHR model specification to those statistics of the static D model in corresponding sub-periods. Kurtosis is higher for the DTVHR model than the static D model for corresponding sub-

	C			CTVHR		
	1962–1988	1989–2002	2003–2009	1962–1988	1989–2002	2003–2009
Mean	0.0144	0.0118	0.0039	0.0052	0.0026	0.0042
t-Statistic	2.4623	2.1628	1.4271	4.5922	1.4151	3.1810
Median	0.0151	0.0075	0.0057	0.0027	0.0000	0.0031
Standard deviation	0.1238	0.0978	0.0314	0.0197	0.0248	0.0126
Skewness	-3.7703	-0.0587	-0.3127	2.4116	0.6905	1.6193
Kurtosis	38.8626	7.6973	4.0041	20.4358	9.7897	11.9498
Minimum	-1.2052	-0.3633	-0.0825	-0.0919	-0.0978	-0.0320
Maximum	0.4190	0.4210	0.0841	0.1503	0.1113	0.0698
Observations < 0	0.3235	0.4226	0.3333	0.3007	0.3274	0.2692
Lower semi-deviation	0.0982	0.0635	0.0214	0.0088	0.0150	0.0058
Upper semi-deviation	0.0765	0.0749	0.0231	0.0183	0.0199	0.0118
Sharpe ratio	0.4037	0.4197	0.4326	0.9124	0.3677	1.1549
Sortino ratio	0.5091	0.6466	0.6349	2.0407	0.6086	2.4945
TIM	0.6527	0.6828	0.7309	0.0191	0.0127	0.0565
Return/TIM	0.0221	0.0173	0.0054	0.2711	0.2071	0.0743

Table 4.10: Excess return statistics for portfolios of top 20 unrestricted cointegration pairs, employed capital and execution delay, July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009.

periods, reflecting a higher proportion of extreme return observations. The maximum returns, for example, are of similar scale to those of the static D model despite the DTVHR model possessing a lower return standard deviation.

Sharpe and Sortino ratios are higher in all corresponding sub-periods for the DTVHR model than the D model, though both models experience declining Sharpe and Sortino ratios between subsequent sub-periods. The greatest Sortino ratio of 29.86, which relates mean return to the lower semi-deviation of returns, is delivered by DTVHR pairs between 1962 and 1988—more than three-times greater than the corresponding D model Sortino ratio of 8.25. The outperformance of the DTVHR model narrows substantially between 2003 and 2009, reflecting an increase in lower semi-deviation and decline in mean return. TIM grows for both model specifications between subsequent sub-periods while return per unit TIM declines.

Table 4.10 reports the excess return statistics for portfolios of the top 20 C and CTVHR unrestricted pairs across the three sub-periods. As with the static D model, the static C model delivers progressively smaller mean returns between subsequent sub-periods, declining from 1.44% between 1962 and 1988 to 0.39% between 2003 and 2009. The CTVHR model, however, delivers a ∼60% higher mean return during the final sub-period relative to the preceding sub-period, delivering an excess monthly return of 0.42%. The CTVHR model is the only model specification to reclaim some of its former profitability during the final sub-period following a sharp decline in profitability during the second sub-period. Median CTVHR returns indicate a high degree of variability among sub-

	D			DTVHR		
	1962–1988	1989–2002	2003–2009	1962–1988	1989–2002	2003–2009
Group 1 proportion	0.0077	0.0314	0.0213	0.0195	0.0167	0.0144
Group 2 proportion	0.2353	0.3547	0.3920	0.0002	0.0084	0.0088
Profitable proportion	0.2174	0.2104	0.1965	1.0000	0.4286	0.2857
Total return	-0.0541	-0.0619	-0.0703	0.0432	-0.0353	-0.0167
Number of trades	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TIM	0.7771	0.7554	0.7442	0.0160	0.2406	0.1961
Industry-matched	0.0390	0.0850	0.4409	0.0000	0.0357	0.2143
Distance zero-crossings	44.6402	38.9806	34.2891	64.0000	34.8929	28.3571
Cointegration zero-crossings	46.9030	41.2980	35.7316	75.0000	32.3571	29.5714
BMD	0.0000	0.0000	0.0000	0.6284	0.6259	4.6903
Group 3 proportion	0.7570	0.6139	0.5867	0.9803	0.9749	0.9768
Profitable proportion	0.8789	0.8332	0.8058	0.6925	0.6395	0.5436
Total return	0.0878	0.0596	0.0658	0.0304	0.0172	0.0081
Number of trades	2.8688	2.5238	2.4461	5.3026	5.0521	4.4583
TIM	0.5674	0.5949	0.6143	0.1791	0.2115	0.2544
Industry-matched	0.0337	0.0783	0.4707	0.0338	0.0851	0.4583
Distance zero-crossings	46.1708	40.7072	35.8666	45.4466	39.9467	35.1423
Cointegration zero-crossings	48.8745	43.3531	38.7311	47.8631	42.4781	37.4038
BMD	0.0000	0.0000	0.0000	0.4228	0.5104	1.2237

Table 4.11: Pair statistics for top 20 unrestricted distance pairs, execution delay, July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009.

periods, with median return dropping to 0.00% between 1989 and 2002 before achieving its highest median return of 0.31% between 2003 and 2009.

Both C and CTVHR models exhibit a greater proportion of negative monthly returns between 1989 and 2002, but while that proportion increased by ∼10% for the static C model from the first sub-period, it only increased by ∼3% for the CTVHR model. TIM steadily increased for the static C model, while initially declining for the CTVHR model before achieving its highest proportion in the final sub-period of 5.65%, diminishing the return per unit TIM accordingly. Sharpe and Sortino ratio are greater for the CTVHR model than the static C model in corresponding sub-periods with the exception of the second sub-period, for which both ratios are inferior for the CTVHR specification.

Table 4.11 reports the pair statistics for portfolios of the top 20 D and DTVHR unrestricted pairs with one-day execution delay upon generation of a trading signal. Group 1 proportions, those pairs that place no trades during the trading period, are relatively stable for the DTVHR model across subsequent sub-periods. Static D model Group 1 proportions, however, increase four-fold between the first and second sub-periods. Group 2 proportions indicate the number of non-convergent pairs increases between subsequent sub-periods for the static D model but remains below 1% for the DTVHR model across all sub-periods. Nearly 40% of all D pairs are non-convergent between 2003 and 2009, up from ∼24% between 1962 and 1988.

The profitable proportion of Group 2 pairs remains steady between subsequent sub-periods for the static D model, but declines significantly for the DTVHR model. For example, 100% of Group 2 DTVHR pairs were profitable between 1962 and 1988. This counter-intuitive result, in which all non-convergent pairs were ultimately profitable, is a function of the negligible 0.02% proportion of DTVHR pairs belonging to Group 2. The magnitude of Group 2 losses increases between subsequent sub-periods for the static D model, beginning with a total return of −5.41% and ending with a total return of −7.03%. The Group 2 total return for the DTVHR model is 4.32% in the first sub-period, reflecting once again the negligible sample size of Group 2 pairs in that sub-period. In the following two sub-periods, however, the magnitude of losses declines from −3.53% to −1.67%.

The proportion of convergent Group 3 pairs declines for the static D model but remains at ∼98% for the DTVHR model between subsequent sub-periods. The profitable proportion declines for both models, as does the total return for DTVHR pairs. The total return for static D pairs fluctuates between the sub-periods, beginning at 8.78% in the first sub-period and ending at 6.58% in the final sub-period. By contrast, the total return of Group 3 DTVHR pairs is only 0.81% in the final sub-period. The number of trades declines under both model specifications while TIM increases. This result indicates that both models spend progressively more time in the market per trade, though those trades are becoming less profitable for the DTVHR model.

BMD, calculated according to Equation (3.59), reports the mean absolute difference between the hedge ratio estimated during the formation period and the time-varying hedge ratio estimated during the trading period. Non-convergent Group 2 DTVHR pairs are expected and observed to have a higher BMD than Group 3 pairs. While this observation is consistent across all sub-periods, the magnitude of both Group 2 and Group 3 BMD statistics is substantially higher in the final sub-period than the two preceding sub-periods. This indicates that while the hedge ratio for non-convergent Group 2 pairs differs substantially between formation and trading periods, the hedge ratio of convergent Group 3 pairs is also becoming less stable as the equilibrium relationship of arbitrage opportunities evolves though time. This is further evidenced by the length of time spent in the market for each trade growing across subsequent sub-periods; the further a tradable spread diverges following the opening of a position, the greater the BMD and length of the trade.

Table 4.12 reports the pair statistics for portfolios of the top 20 C and CTVHR unrestricted pairs with one-day execution delay upon generation of a trading signal.

	C			CTVHR		
	1962–1988	1989–2002	2003–2009	1962–1988	1989–2002	2003–2009
Group 1 proportion	0.0269	0.0182	0.0138	0.4889	0.6341	0.0920
Group 2 proportion	0.3459	0.3492	0.3742	0.0037	0.0036	0.0044
Profitable proportion	0.2349	0.2182	0.1890	0.6818	0.2500	0.7143
Total return	-0.2432	-0.2813	-0.2190	0.0116	-0.0603	-0.0129
Number of trades	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TIM	0.8056	0.8336	0.8596	0.0206	0.0218	0.1115
Industry-matched	0.0986	0.1330	0.2341	0.0909	0.1667	0.1429
Distance zero-crossings	36.7704	37.9744	26.7391	40.1818	33.3333	25.5714
Cointegration zero-crossings	79.4500	81.3069	68.4080	72.5000	73.3333	69.0000
BMD	0.0000	0.0000	0.0000	0.9516	22.4516	4.7811
Group 3 proportion	0.6272	0.6326	0.6120	0.5074	0.3623	0.9036
Profitable proportion	0.8291	0.7807	0.8016	0.6122	0.5472	0.5824
Total return	0.2699	0.2595	0.1860	0.0276	0.0153	0.0214
Number of trades	2.7833	2.6184	2.6871	2.3727	2.2030	3.0554
TIM	0.5947	0.6090	0.6687	0.0371	0.0335	0.0597
Industry-matched	0.1234	0.1120	0.2321	0.1130	0.1397	0.2168
Distance zero-crossings	37.7315	38.3798	27.0746	34.6758	33.6672	26.7175
Cointegration zero-crossings	76.7369	81.6706	68.2025	71.8138	73.4043	67.6385
BMD	0.0000	0.0000	0.0000	0.9862	2.5141	3.2352

Table 4.12: Pair statistics for top 20 unrestricted cointegration pairs, execution delay, July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009.

Group 1 proportions decline between subsequent sub-periods for the static C model beginning at 2.69% between 1962 and 1988 and ending at 1.38% between 2003 and 2009. Group 1 proportions of the CTVHR model, however, are substantially higher in all sub-periods than the static C model and actually constitute the majority of pairs between 1989 and 2002—that is, 63.41% of pairs fail to open a single position during the trading period. The reversion rate parameter used in the TVHR model is estimated through a transformation of the number of formation period zero-crossings—the greater the number of zero-crossings, the faster the reversion rate. Cointegration pairs are selected for their high number of formation period zero-crossings which naturally leads to a comparatively fast reversion parameter. Group 1 proportions are therefore much higher under the CTVHR model than the static C model because the hedge ratio is updated too quickly for a significant spread divergence to form and trigger a statistical arbitrage opportunity.

Group 2 proportions remain relatively consistent between subsequent sub-periods for both models, though those proportions remain at ∼35% for static C pairs and below 0.5% for CTVHR pairs. The profitable proportion declines for static C pairs and fluctuates for CTVHR pairs. The first sub-period delivers a positive total return for CTVHR pairs while the final sub-period delivers a negative total return, despite a very high proportion of profitable pairs. Both C and CTVHR Group 2 pairs deliver their greatest loss between 1989 and 2002, corresponding with pairs that delivered the greatest number of formation period zero-crossings under the cointegration specification. This is accompanied by a high BMD statistic for the CTVHR model, indicating a high degree of continued spread divergence during the trading period.

Group 3 proportions are stable between sub-periods for the static C model but vary substantially for the CTVHR model due to the influence of high Goup 1 proportions. The greatest profitable proportion under both models is observed between 1962 and 1988, while the lowest profitable proportion under both models is observed between 1989 and 2002. Total return declines between sub-periods for the C model but significantly outperforms those of the CTVHR model by one order of magnitude. Number of trades are similar for both models, and TIM achieves its highest proportion in the final sub-period, though the models again differ by an order of magnitude. Industry matching progressively increases for both Group 2 and Group 3 pairs, and BMD increases for Group 3 CTVHR pairs, both statistics reflecting changing market dynamics toward a state of greater intra-industry volatility.

Table 4.13 reports the individual trade statistics for portfolios of the top 20 D and DTVHR unrestricted pairs with one-day execution delay upon generation of a trading signal. Profitable proportion, mean return, and Sharpe ratio all decline between subsequent sub-periods for both model specifications. In the final sub-period, a higher proportion

	D			DTVHR		
	1962–1988	1989–2002	2003–2009	1962–1988	1989–2002	2003–2009
All trades						
Profitable proportion	0.7671	0.6792	0.6513	0.6223	0.6114	0.5818
Mean return	0.0223	0.0081	0.0066	0.0057	0.0033	0.0015
Standard deviation	0.0579	0.0709	0.0836	0.0266	0.0393	0.0398
Sharpe ratio	1.0814	0.2822	0.1883	1.6555	0.5823	0.2304
Mean profit	0.0457	0.0428	0.0522	0.0203	0.0191	0.0224
Mean loss	-0.0546	-0.0653	-0.0785	-0.0183	-0.0215	-0.0275
Mean long return	0.0249	0.0255	0.0165	0.0049	0.0041	0.0025
Mean short return	-0.0025	-0.0174	-0.0099	0.0008	-0.0008	-0.0009
Mean trade length	31.9762	41.2805	44.9021	4.2403	5.2656	7.1784
Median trade length	19.0000	27.0000	32.0000	3.0000	4.0000	5.0000
Convergent trades	0.6863	0.5790	0.5431	0.9616	0.9468	0.9272
Profitable proportion	0.9606	0.9749	0.9771	0.6281	0.6221	0.5993
Mean return	0.0469	0.0452	0.0563	0.0061	0.0041	0.0028
Standard deviation	0.0301	0.0349	0.0402	0.0265	0.0377	0.0397
Sharpe ratio	5.4364	4.1375	4.2276	1.7586	0.7536	0.4148
Mean profit	0.0493	0.0466	0.0584	0.0204	0.0192	0.0228
Mean loss	-0.0122	-0.0104	-0.0332	-0.0181	-0.0208	-0.0271
Mean long return	0.0320	0.0342	0.0340	0.0050	0.0043	0.0028
Mean short return	0.0148	0.0110	0.0223	0.0010	-0.0002	0.0000
Mean trade length	20.6686	24.6403	27.6185	4.2581	5.2661	7.1635
Median trade length	14.0000	16.0000	20.0000	3.0000	4.0000	5.0000
Non-convergent trades	0.3137	0.4210	0.4569	0.0384	0.0532	0.0728
Profitable proportion	0.3439	0.2724	0.2642	0.4772	0.4208	0.3598
Mean return	-0.0314	-0.0429	-0.0524	-0.0030	-0.0109	-0.0141
Standard deviation	0.0672	0.0760	0.0835	0.0272	0.0595	0.0379
Sharpe ratio	-0.9855	-1.1183	-1.2309	-0.9133	-1.2702	-2.1735
Mean profit	0.0235	0.0237	0.0250	0.0172	0.0142	0.0157
Mean loss	-0.0602	-0.0679	-0.0801	-0.0216	-0.0291	-0.0308
Mean long return	0.0091	0.0134	-0.0043	0.0023	0.0007	-0.0015
Mean short return	-0.0405	-0.0563	-0.0481	-0.0054	-0.0116	-0.0125
Mean trade length	56.7146	64.1686	65.4436	3.7920	5.2554	7.3678
Median trade length	53.0000	66.0000	68.0000	3.0000	3.0000	5.0000

Table 4.13: Trade statistics for top 20 unrestricted distance pairs, execution delay, July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009.

of D model trades are profitable than DTVHR model trades, also delivering more than four-times the mean return. Despite this, the DTVHR model delivers a higher Sharpe ratio than the D model in all sub-periods due to its lower return standard deviation. Mean profit and mean loss are of a similar scale within each model specification, though the magnitude of mean profit under the DTVHR model is approximately half that of the static D model, and mean loss is approximately one-third under the DTVHR model. The static D model generates negative returns on its short positions, the greatest of which being a 1.74% loss between 1989 and 2002. The DTVHR model was only able to generate a positive return on the short position during the first sub-period before making losses of the same scale during the two subsequent sub-periods. Mean and median trade length increase under both model specifications.

The proportion of convergent trades declines between subsequent sub-periods for both model specifications, though the DTVHR model's lowest proportion of 92.72% in the final sub-period is higher than the D model's greatest proportion of 68.63% in the first sub-period. In both cases, the declining proportion of convergent trades is a function of the increasing mean trade length—the longer a trade takes to converge, the more likely a position will remain open at the end of the trading period. The profitable proportion of convergent trades declines slightly under the DTVHR model specification to ∼60% in the final sub-period. By contrast, the profitable proportion of convergent D trades is 97.71% in the final sub-period; the static D model would deliver a convergent trade profitable proportion of 100% were the execution delay not imposed. Mean return declines between subsequent sub-periods for DTVHR convergent trades to its lowest value of 0.28% in the final sub-period, while the static D model delivers returns of 5.63% in the same sub-period. Return standard deviation is similar between both models in each sub-period, leading to substantial outperformance from the static D model in terms of Sharpe ratio. For convergent trades, the static D model delivers a positive return on its short positions in every sub-period, while the DTVHR model only delivers a positive return on its short positions in the first sub-period. Additionally, the DTVHR model delivers declining positive returns on its long positions between subsequent sub-periods.

Both model specifications deliver a declining proportion of profitable non-convergent trades, though the DTVHR model maintains a greater proportion in each sub-period. Non-convergent trades are becoming progressively more unprofitable for both models with static D non-convergent trades realising a loss of 5.24% in the final sub-period compared to DTVHR non-convergent trades realising a loss of 1.41%. With the exception of the final sub-period, losses are entirely driven by the short position of both model specifications with the long position contributing small additional losses in the final sub-period. For the static D model, the mean trade length is two- to three-times greater than that observed for convergent trades in corresponding sub-periods. The mean trade length of non-convergent DTVHR trades, by contrast, is relatively consistent with mean convergent trade length within corresponding trade periods. This illustrates the temporal cost of non-convergent static D trades relative to non-convergent DTVHR trades—while conventional static models bind investment capital in loss-making trades for a relatively long time, the DTVHR model spends no more time in a loss-making trade than a profit-making trade, allowing capital to be redeployed soon after.

Table 4.14 reports the individual trade statistics for portfolios of the top 20 C and CTVHR unrestricted pairs with one-day execution delay upon generation of a trading signal. The profitable proportion of trades declines marginally for both models, while mean return declines more substantially for the static C model and fluctuates for the CTVHR model between subsequent sub-periods. The static C model maintains a greater profitable proportion and mean return than the CTVHR model in corresponding sub-periods at the expense of a higher standard deviation, leading to outperformance by the CTVHR model in terms of Sharpe ratio. While the CTVHR model delivers higher Sharpe ratios than the static C model in corresponding sub-periods, it delivers an abnormally low Sharpe ratio during the second sub-period. This low Sharpe ratio is caused by a combination of a comparatively low mean return and high standard deviation. Mean profit and loss are of a similar scale within each sub-period for the CTVHR model, while the magnitude of mean loss is greater than mean profit for the static C model.

The proportion of convergent trades declines between sub-periods for the static C model to its lowest proportion of 57.89%, but remains above 97% in all sub-periods for the CTVHR model. The proportion of convergent trades is slightly higher for the CTVHR model than the DTVHR model due to the estimation of a faster reversion rate parameter facilitating shorter trade lengths. Due to the overwhelming proportion of CTVHR trades converging during the trading period, convergent trade statistics are almost identical to all trade statistics. Mean convergent trade return for the static C model fluctuates between sub-periods, delivering its highest return of 19.19% during the second sub-period. Despite mean return declining to 14.91% during the final sub-period, the static C model delivers its highest Sharpe ratio of 3.10 thanks largely to a reduction in return standard deviation. Both models deliver their highest return standard deviation during the second sub-period, indicating the presence of an unfavourable market regime for cointegration pairs. Both models deliver positive long and short position returns for all sub-periods with the exception of the second sub-period for the CTVHR model, which delivers a negative return for its short positions. Mean trade length is substantially shorter for static C convergent trades than non-convergent trades with convergent trades completing in one-third the time of non-convergent trades. There is no appreciable difference in mean trade length between convergent and non-convergent CTVHR trades.

	C			CTVHR		
	1962–1988	1989–2002	2003–2009	1962–1988	1989–2002	2003–2009
All trades						
Profitable proportion	0.6824	0.6440	0.6586	0.5907	0.5532	0.5580
Mean return	0.0406	0.0340	0.0176	0.0120	0.0056	0.0084
Standard deviation	0.3591	0.3983	0.2842	0.0947	0.1715	0.0930
Sharpe ratio	0.2900	0.2096	0.1468	1.4446	0.3784	0.9221
Mean profit	0.1798	0.2158	0.1452	0.0585	0.0803	0.0513
Mean loss	-0.2584	-0.2949	-0.2285	-0.0552	-0.0868	-0.0457
Mean long return	0.0461	0.0396	0.0213	0.0064	0.0079	0.0040
Mean short return	-0.0055	-0.0057	-0.0036	0.0056	-0.0023	0.0044
Mean trade length	38.3706	41.7014	45.0635	1.9251	1.8988	2.4227
Median trade length	22.0000	26.0000	30.0000	2.0000	2.0000	2.0000
Convergent trades	0.6265	0.6076	0.5789	0.9817	0.9779	0.9756
Profitable proportion	0.9010	0.8790	0.9597	0.5905	0.5566	0.5574
Mean return	0.1604	0.1919	0.1491	0.0119	0.0061	0.0085
Standard deviation	0.2172	0.2342	0.1508	0.0950	0.1728	0.0935
Sharpe ratio	2.5661	2.6758	3.1031	1.4311	0.4050	0.9209
Mean profit	0.1860	0.2309	0.1573	0.0585	0.0806	0.0515
Mean loss	-0.0720	-0.0917	-0.0470	-0.0554	-0.0874	-0.0458
Mean long return	0.1018	0.1126	0.0740	0.0063	0.0084	0.0040
Mean short return	0.0586	0.0793	0.0751	0.0056	-0.0024	0.0045
Mean trade length	20.8849	23.6349	25.5559	1.9283	1.8980	2.4285
Median trade length	12.5000	15.5000	17.0000	2.0000	2.0000	2.0000
Non-convergent trades	0.3735	0.3924	0.4211	0.0183	0.0221	0.0244
Profitable proportion	0.3159	0.2800	0.2446	0.6015	0.4000	0.5818
Mean return	-0.1603	-0.2107	-0.1630	0.0151	-0.0139	0.0065
Standard deviation	0.4492	0.4700	0.3227	0.0740	0.0952	0.0686
Sharpe ratio	-0.6883	-0.8523	-0.9460	2.4472	-1.6631	1.0179
Mean profit	0.1502	0.1420	0.0802	0.0548	0.0645	0.0430
Mean loss	-0.3036	-0.3478	-0.2418	-0.0449	-0.0661	-0.0443
Mean long return	-0.0473	-0.0733	-0.0513	0.0132	-0.0160	0.0036
Mean short return	-0.1130	-0.1374	-0.1118	0.0019	0.0021	0.0029
Mean trade length	67.6962	69.6797	71.8810	1.7519	1.9333	2.1909
Median trade length	68.0000	71.0000	78.0000	1.0000	1.0000	2.0000

Table 4.14: Trade statistics for top 20 unrestricted cointegration pairs, execution delay, July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009.

Non-convergent static C model trades grew in proportion to constitute 42.11% of all trades in the final sub-period compared to 2.44% for the CTVHR model. As with the distance model variants, the increasing proportion of non-convergent C and CTVHR trades is attributable to the increasing mean trade length between subsequent sub-periods. The profitable proportion of non-convergent static C trades is declining, while the CTVHR model maintains a proportion of ∼60% in the first and final sub-periods. The second sub-period delivers a profitable non-convergent CTVHR trade proportion of only 40.00%, along with the model's only negative non-convergent trade return of −1.39%. The static C model also delivers its greatest loss on non-convergent trades in the second sub-period. As with convergent trades, the standard deviation of non-convergent trade returns is greatest in the second sub-period for both models, with Sharpe ratios reflecting the greater return variability. Mean loss is more than double mean profit for the static C model, while profit and loss are of similar magnitudes for the CTVHR model in all sub-periods. Non-convergent trade losses are largely driven by short positions for the static C model, while only the long return of the CTVHR model delivers a loss during the second sub-period, with long positions in the other two sub-periods and short positions in all sub-periods delivering profits.

The sub-period analysis conducted in Section 4.1.2 reveals a trend of declining profitability, as identified by Do and Faff (2010), for all model specifications with the exception of cointegration-based TVHR portfolios. Static D and C model portfolios both suffer a greater proportional decline in profitability than DTVHR model portfolios, though the latter delivers lower mean returns in absolute terms in the final sub-period. Sharpe and Sortino ratios for the D and DTVHR models decline between subsequent sub-periods, though the DTVHR model outperforms the D model in every sub-period.

The declining profitability of the static D model is attributable to a combination of an increasing proportion of non-convergent Group 2 pairs, increasing losses on non-convergent trades, and the declining profitability of Group 3 pairs. This is accompanied by an increasing mean trade length for both convergent and non-convergent trades, indicating slowing reversion rates for distance pairs. This is further confirmed by the growing mean trade length of both convergent and non-convergent DTVHR trades between subsequent sub-periods. Despite the construction of the TVHR algorithm ensuring convergence given sufficient time, growing BMD for Group 2 and Group 3 DTVHR pairs and longer convergent and non-convergent trade lengths indicate that the market is gradually becoming less amenable to the exploitation of distance-based statistical arbitrage opportunities.

Cointegration pairs behave differently to their distance counterparts, with the static C model delivering consistent Sharpe and Sortino ratios between subsequent sub-periods while delivering diminishing returns. The cointegration pair selection procedure prioritises pair volatility without requiring some common stochastic relationship, so the return of static C pairs can be considered a proxy for broad market volatility with both the mean and standard deviation of monthly returns declining. Risk-adjusted returns remain consistent for cointegration pairs despite declining market volatility, in contrast to distance pairs which are more adversely affected by declining market volatility. This is supported by CTVHR pairs delivering their greatest risk-adjusted returns in the final sub-period despite declining market volatility.

The CTVHR portfolio returns are unique in their delivery of their greatest risk-adjusted returns in the final sub-period following their poorest risk-adjusted and absolute returns in the second sub-period. Examination of the statistics reported in Table 4.12 reveals that the second sub-period had the greatest proportion of Group 1 pairs in addition to the highest number of cointegration zero-crossings for both Group 2 and Group 3 pairs. The high number of formation period zero-crossings leads to the estimation of relatively fast reversion rate parameters in the TVHR model which, if there is a great disparity between formation and trading period spread volatility, will cause the time-varying estimation of the hedge ratio to update before the spread can diverge significantly enough to generate a trading signal. The cooling effect that reduced market volatility has on the estimation of the reversion rate parameter leads to a greater number of pairs opening a position during the trading period, thereby reducing Group 1 proportions to their lowest level in the final sub-period. While the assertion of declining pairs trading profitability made by Do and Faff (2010) is confirmed in this thesis, declining risk-adjusted returns are found to be confined to distance pairs while cointegration pairs maintain or improve their risk-adjusted returns.

4.1.3 Extended Study Period

Section 4.1.3 extends the sample period beyond that considered by Do and Faff (2010) to include market data between July 2009 and June 2018. The extended study period allows for the out-of-sample investigation of the conventional distance approach in addition to the static cointegration approach developed in this thesis, along with both their TVHR extensions. Table 4.15 reports the excess return statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal. Figure 4.13 displays the cumulative excess return of unrestricted distance and cointegration pairs with one-day execution delay, and Figure 4.14 displays the distribution of monthly excess returns for distance and cointegration portfolios with one-day execution delay.

	D	DTVHR	C	CTVHR
Mean	0.0000	0.0010	0.0067	0.0056
t-Statistic	0.0110	2.3458	1.1380	3.4205
Median	-0.0002	0.0012	0.0008	0.0026
Standard deviation	0.0061	0.0036	0.0622	0.0200
Skewness	-0.0863	0.4847	7.0213	4.7455
Kurtosis	2.6490	4.5150	64.4062	35.8846
Minimum	-0.0156	-0.0073	-0.0919	-0.0335
Maximum	0.0129	0.0155	0.5719	0.1610
Observations < 0	0.5278	0.3889	0.4815	0.3241
Lower semi-deviation	0.0043	0.0019	0.0187	0.0055
Upper semi-deviation	0.0043	0.0032	0.0594	0.0199
Sharpe ratio	0.0031	0.9877	0.3745	0.9766
Sortino ratio	0.0044	1.8833	1.2468	3.5398
TIM	0.6293	0.2634	0.6985	0.0682
Return/TIM	0.0000	0.0039	0.0096	0.0826

Table 4.15: Excess return statistics for portfolios of top 20 unrestricted pairs, employed capital and execution delay, July 2009–June 2018.

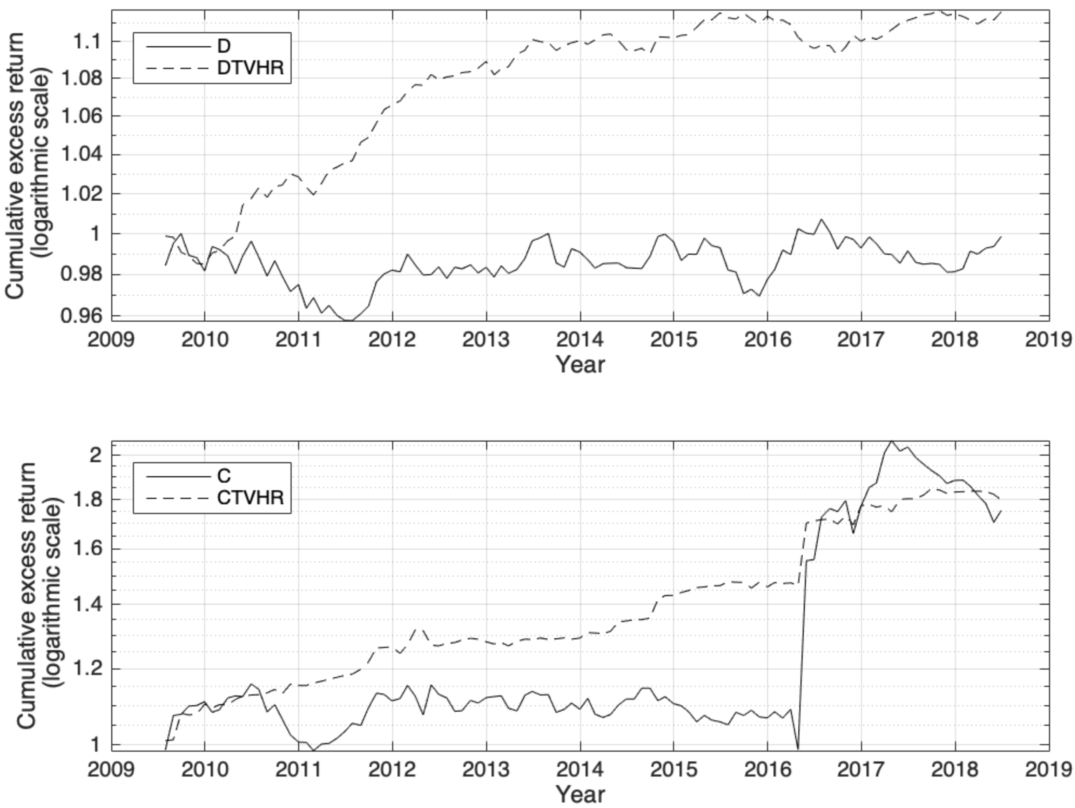

Figure 4.13: Cumulative excess return of top 20 unrestricted pairs, employed capital, execution delay, July 2009–June 2018.

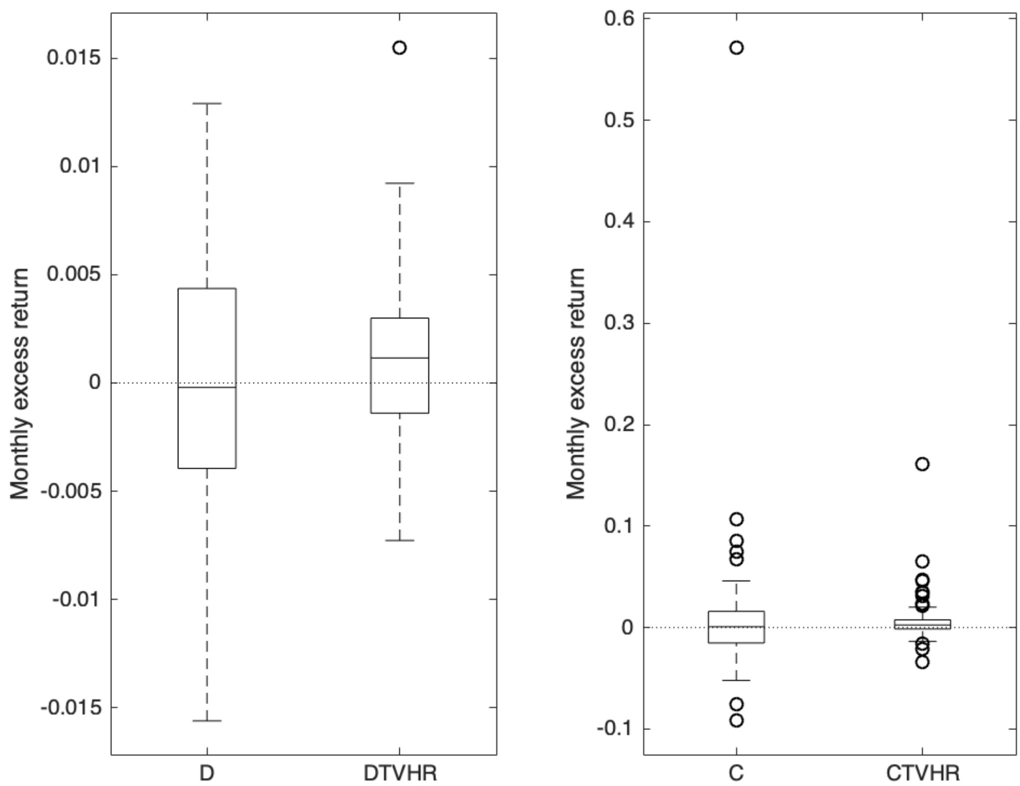

Figure 4.14: Distribution of monthly excess return of top 20 unrestricted pairs, employed capital, execution delay, July 2009–June 2018.

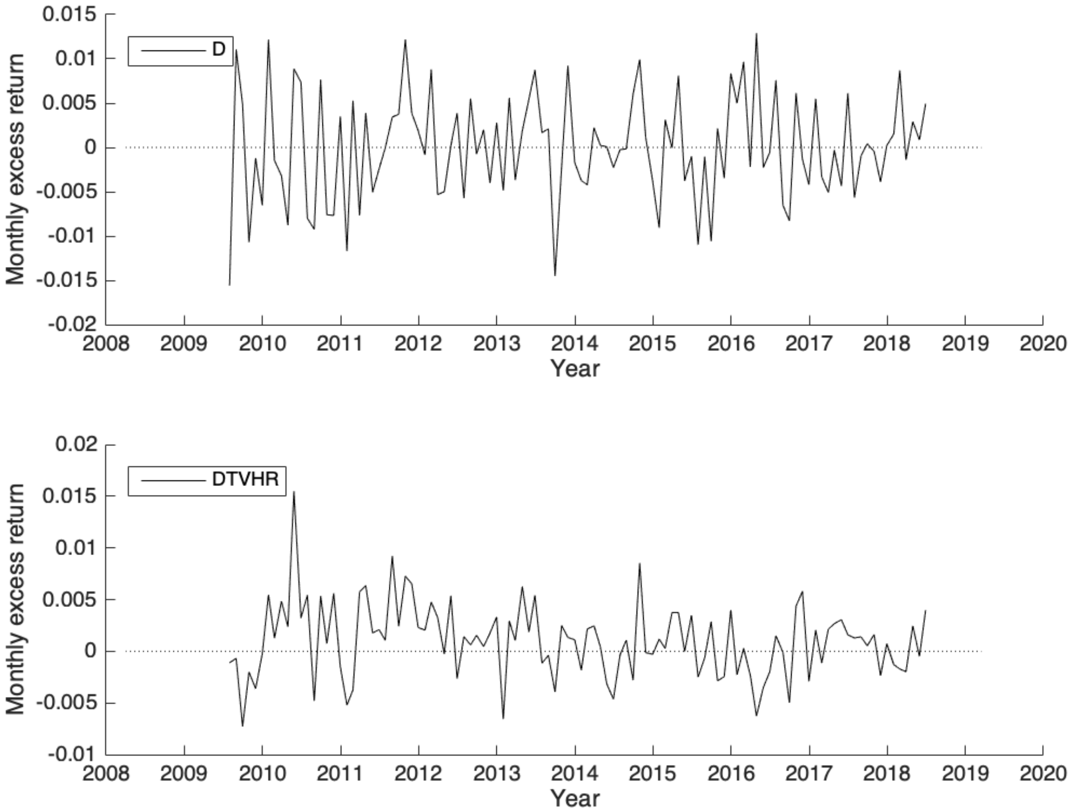

Figure 4.15: Monthly excess return of top 20 distance pairs, employed capital, execution delay, July 2009–June 2018.

Monthly mean returns declined for both distance models and increased for both cointegration models relative to the results reported in Section 4.1.2 for the final sub-period. Of particular note is the 0.00% mean monthly return of the static D model, and its median monthly return of −0.02%. The static D model is outperformed by its TVHR counterpart in terms of every statistic reported in Table 4.15 with the possible exception of kurtosis and TIM, depending on the objectives of the arbitrageur. Mean monthly return of the DTVHR model is 0.10% with a median monthly return of 0.12%. Additionally, the DTVHR model's positive skewness contrasts the static D model's negative skewness, while its maximum loss is less than half that of its static counterpart, its maximum gain is greater than that of its static counterpart, and its proportion of negative monthly returns is lower. The DTVHR model's Sharpe ratio is two orders of magnitude greater than that of the static D model, and its Sortino ratio is three orders of magnitude greater. The DTVHR model also delivers its returns more efficiently, spending less than half the time in the market that the static D model spends. Figure 4.15 displays the monthly excess return for top 20 distance pairs.

Cointegration pairs improve on their performance from the January 2003–June 2009

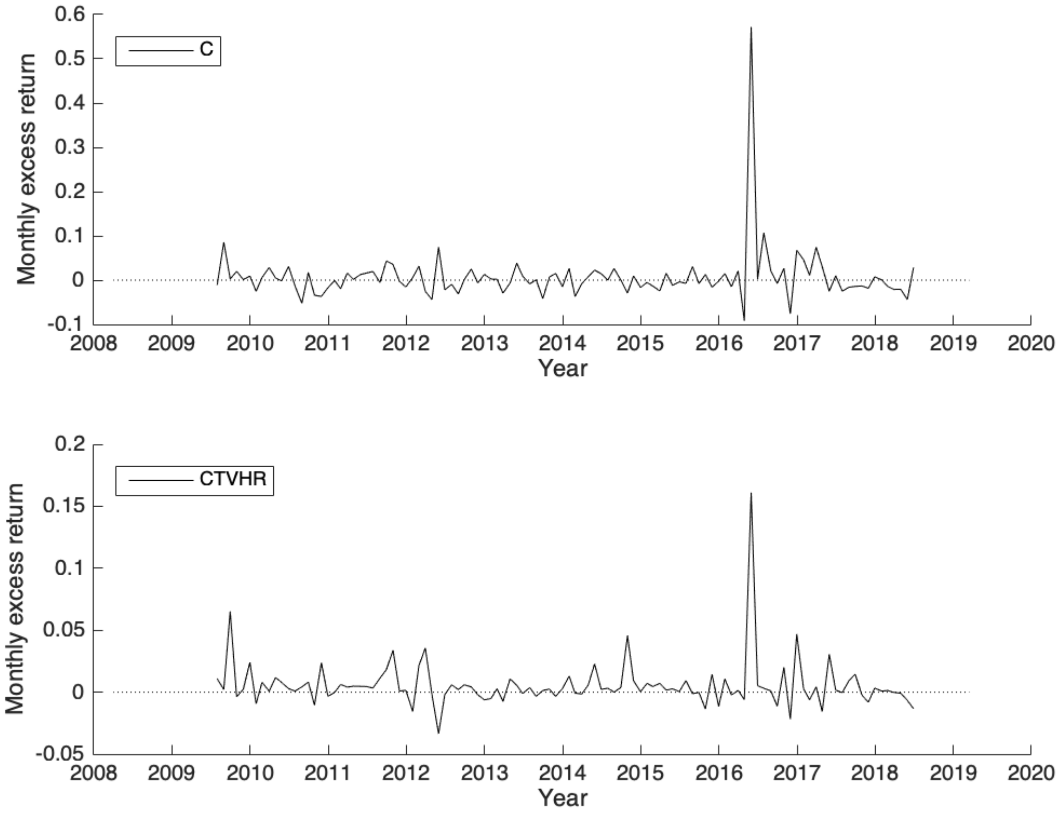

Figure 4.16: Monthly excess return of top 20 cointegration pairs, employed capital, execution delay, July 2009–June 2018.

sub-period, with mean monthly return increasing from 0.39% to 0.67% under the static C model, and from 0.42% to 0.56% under the CTVHR model. This represents the greatest mean monthly return in both absolute and risk-adjusted terms for the CTVHR model of any sub-period. The CTVHR model also delivers the greatest median monthly return and lowest proportion of negative monthly returns out of any model specification. The CTVHR model additionally achieves the second-highest Sharpe ratio behind the DTVHR model, and the highest Sortino ratio of 3.54. The C and CTVHR models achieve greater capital efficiency than their distance counterparts, delivering higher returns per unit TIM. The CTVHR model in particular delivers the greatest return per unit TIM of 8.26% thanks largely to its very limited TIM of 6.82%. Figure 4.16 displays the monthly excess return for top 20 cointegration pairs.

Abnormally large positive monthly returns of 57.19% for the C model, and 16.10% for the CTVHR model, are observed in May 2016. While both model specifications realise their greatest monthly returns in May 2016, the C model specification return is conspicuously large relative to its other monthly returns in the extended study period. Despite its statistical appearance as an outlier that should be excluded from the dataset, this return remains unaltered in the analysis for a number of reasons. The first reason for its inclusion is that the large return is not attributable to any single trade during the month. The C model placed 164 trades in May 2016, whose mean of 31.63% is lower than the median return among all trades of 38.94%, indicating that trade returns were negatively skewed during the month. The second reason for the return's inclusion is that the CTVHR model also realised its greatest return in May 2016, indicating that the state of the market was particularly amenable to the cointegration specifications during the month. The third reason for its inclusion is that neither Gatev, Goetzmann, and Rouwenhorst (2006) nor Do and Faff (2010) removed any outliers from their respective analyses. The final reason for its inclusion is that it serves to illustrate the variability and volatility of pairs selected under the cointegration specification, as further evidenced by the minimum monthly return of -120.52% and maximum monthly return of 42.10% realised by the C model during the initial study period as reported in Table 4.4. This large positive return is therefore attributable to the idiosyncrasies of the C model and the weighting of overlapping portfolio returns under the employed capital allocation scheme, and its exclusion along with its other extreme returns would unfairly bias the analysis.

The excess return statistics of distance and cointegration portfolios between July 2009 and June 2018 demonstrate the continued decline of out-of-sample profitability of distance pairs. By contrast, cointegration pairs halted and reversed their declining profitability, with the CTVHR model delivering its highest mean monthly return of any sub-period. For both distance and cointegration pairs, the TVHR model outperformed its static hedge ratio analogue in risk-adjusted terms, and in absolute return terms for distance pairs. To more thoroughly investigate the outperformance delivered by cointegration pairs and the TVHR model, Table 4.16 reports the pair statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal.

Group 1 proportions for the D, DTVHR, and C models are relatively consistent with those observed in previous sub-periods, while the Group 1 proportion for the CTVHR model is at its lowest level of all sub-periods, declining from 9.20% between 2003 and 2009 to 4.59% between 2009 and 2018. While this proportion is still the highest among alternative model specifications, it is substantially lower than the Group 1 proportion typically observed for the CTVHR model. This reduction in the proportion of non-trading pairs is accompanied by an increase in the proportion of convergent Group 3 pairs, while non-convergent Group 2 pairs are observed in similar proportion to previous sub-periods. Both the DTVHR and CTVHR models deliver substantially lower Group 2 pair proportions than their static counterparts, with the DTVHR Group 2 proportion being one order of magnitude lower than the static D model, and CTVHR Group 2 proportion being two orders of magnitude lower than the static C model. In particular, the CTVHR Group 2 proportion is negligible at 0.14%.

	D	DTVHR	C	CTVHR
Group 1 proportion	0.0255	0.0116	0.0153	0.0459
Group 2 proportion	0.4606	0.0186	0.4096	0.0014
Profitable proportion	0.2105	0.2250	0.2072	0.6667
Total return	-0.0573	-0.0282	-0.2236	0.0018
Number of trades	1.0000	1.0000	1.0000	1.0000
TIM	0.7211	0.2869	0.8278	0.0162
Industry-matched	0.2276	0.2000	0.1370	0.0000
Distance zero-crossings	30.6526	22.0500	24.0362	21.3333
Cointegration zero-crossings	32.2326	17.3500	66.4247	69.3333
BMD	0.0000	0.7517	0.0000	697.3834
Group 3 proportion	0.5139	0.9699	0.5751	0.9527
Profitable proportion	0.8538	0.5887	0.8145	0.5920
Total return	0.0540	0.0064	0.2275	0.0375
Number of trades	2.1787	3.8733	2.4919	3.4946
TIM	0.5739	0.2638	0.6271	0.0709
Industry-matched	0.3014	0.2673	0.1605	0.1470
Distance zero-crossings	31.9287	31.3840	27.2597	25.7843
Cointegration zero-crossings	34.1408	33.4931	66.7411	66.3690
BMD	0.0000	1.3426	0.0000	4.5593

Table 4.16: Pair statistics for top 20 unrestricted pairs, execution delay, July 2009–June 2018.

The profitable proportion of Group 2 pairs is similar for D, DTVHR, and C models at ∼21%, while the profitable proportion of CTVHR Group 2 pairs is more than three-times greater at ∼67%. Additionally, the CTVHR model is the only model variant that delivers a positive total return among its Group 2 pairs, though the limited sample size likely skews this statistic. The negative total return observed for DTVHR Group 2 pairs is approximately half that of the static D model, and approximately one-eighth that of the static C model. TVHR model Group 2 pairs also spend substantially less time in the market than their static counterparts, with the CTVHR model delivering a TIM of only 1.62%, compared with the static C model TIM of 82.78%.

The DTVHR Group 3 proportion is the highest of all model variants at 96.99%, with the CTVHR model delivering a slightly lower proportion of 95.27%. The static D and C model Group 3 proportions are much lower at 51.39% and 57.51%, respectively. Almost half of all D and C model pairs, therefore, either fail to open a position, or open a single non-convergent trade during the trading period. The profitable proportion of Group 3 pairs is higher for static models than TVHR models, offsetting some of the lost performance attributable to a high proportion of Group 2 pairs. Additionally, the positive total return of static Group 3 pairs is one order of magnitude greater than their respective TVHR variants, with the static C model delivering the greatest total return of 22.75%, and

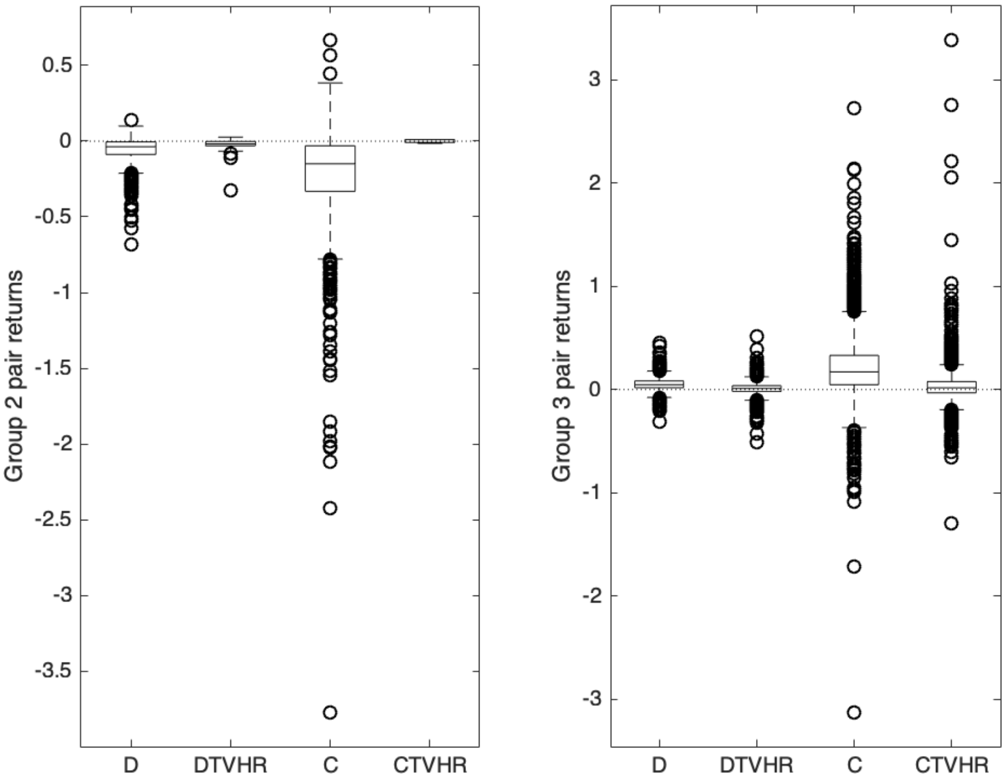

Figure 4.17: Distribution of pair excess return of top 20 unrestricted pairs, employed capital, execution delay, July 2009–June 2018.

DTVHR model delivering the lowest total return of 0.64%. DTVHR and CTVHR Group 3 pairs place more trades than their static counterparts during the trading period despite spending less time in the market. Figure 4.17 displays the distribution of pair total returns for Group 2 and Group 3 pairs with one-day execution delay.

BMD is very high for CTVHR Group 2 pairs, illustrating the extreme level of continued divergence caused by the time-varying hedge ratio shifting to a new equilibrium level. As expected, BMD for CTVHR Group 3 pairs is substantially lower than that observed for Group 2 pairs, though a hitherto unobserved inversion of Group 2 and Group 3 BMD magnitudes is reported for DTVHR pairs. BMD is typically greater for Group 2 pairs due to the tendency of trade non-convergence to be caused by evolving hedge ratios. A possible explanation for the relatively low BMD of DTVHR Group 2 pairs, despite the relative stability indicated by BMD, is the estimation of a slow reversion rate parameter for the TVHR model, calculated according to Equation (3.58). The estimate of the reversion rate parameter is informed by the number of formation period zero-crossings, $\mathbb{E}[D]$. Given the limited number of distance zero-crossings observed for DTVHR Group 2 pairs, it is possible that the estimated reversion rate parameter was too slow to facilitate convergence in a timely manner. This illustrates the limitations of using a relatively short formation period to estimate the reversion rate parameter of the TVHR model, whose estimation is inconsistent with the empirical features of the tradable spread observed during the trading period.

Table 4.17 reports the individual trade statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal. All models deliver a profitable proportion of trades exceeding 50%, with the static C model delivering the greatest profitable proportion of 65.46%, and the CTVHR model delivering the lowest profitable proportion of 55.49%. Despite the relatively low profitable proportion of the CTVHR model, its mean return is the second highest at 0.96%, and its Sharpe ratio is the highest at 0.92. Unlike other model variants, the average profitable CTVHR trade is greater in magnitude than the average unprofitable trade, contributing to its high mean return despite its relatively low profitable proportion. The static D model is the only model specification whose trades deliver a negative mean return, driven by its relatively large average loss on unprofitable trades that is almost 60% greater in magnitude than its average gain on profitable trades. Both the DTVHR and static C models deliver greater losses on unprofitable trades than gains on profitable trades, though the ratio is sufficiently low on DTVHR trades, and the profitable proportion sufficiently high on static C trades, negating this imbalance to deliver positive mean returns for each. Figure 4.18 displays the distribution of trade returns for convergent and non-convergent trades with one-day execution delay.

The proportion of convergent trades is lowest for the static D model at 47.62%, and highest for the CTVHR model at 97.76%. The extended study period from July 2009 to June 2018 is the only sub-period in which any model variant delivers fewer than 50% convergent trades. Both TVHR models deliver greater than 90% convergent trade proportions, while the static C model convergent trade proportion lies below 55%. The static models exhibit greater profitable proportions, mean returns, and Sharpe ratios among convergent trades than their TVHR counterparts. Additionally, the average gain on profitable trades is more than three-times greater in magnitude than the average loss on unprofitable trades for static D and C pairs, while the average loss on unprofitable trades is greater in magnitude than the average gain on profitable trades for DTVHR pairs. Across all model specifications, only the DTVHR model delivers a negative return on its short positions.

Among non-convergent trades, the CTVHR model is the only model specification that delivers a profitable proportion greater than 50%. Static D and C pairs exhibit similar profitable proportions at ∼27%, while the DTVHR model exhibits a profitable proportion of ∼37%. In addition to higher profitable proportions of non-convergent trades, the TVHR

	D	DTVHR	C	CTVHR
All trades				
Profitable proportion	0.6090	0.6095	0.6546	0.5549
Mean return	-0.0004	0.0015	0.0161	0.0096
Standard deviation	0.0714	0.0296	0.2759	0.1035
Sharpe ratio	-0.0137	0.2777	0.1354	0.9227
Mean profit	0.0407	0.0165	0.1491	0.0510
Mean loss	-0.0645	-0.0219	-0.2359	-0.0421
Mean long return	0.0334	0.0069	0.0353	0.0024
Mean short return	-0.0338	-0.0054	-0.0192	0.0071
Mean trade length	49.3328	8.7279	46.8306	2.5276
Median trade length	39.0000	6.0000	32.0000	2.0000
Convergent trades	0.4762	0.9153	0.5487	0.9776
Profitable proportion	0.9916	0.6316	0.9598	0.5559
Mean return	0.0464	0.0027	0.1568	0.0100
Standard deviation	0.0285	0.0295	0.1474	0.1044
Sharpe ratio	4.7731	0.4872	3.3057	0.9518
Mean profit	0.0470	0.0169	0.1654	0.0516
Mean loss	-0.0137	-0.0217	-0.0501	-0.0421
Mean long return	0.0426	0.0078	0.0944	0.0026
Mean short return	0.0038	-0.0051	0.0624	0.0074
Mean trade length	29.3807	8.6432	26.0970	2.5355
Median trade length	22.0000	6.0000	17.0000	2.0000
Non-convergent trades	0.5238	0.0847	0.4513	0.0224
Profitable proportion	0.2612	0.3705	0.2835	0.5092
Mean return	-0.0430	-0.0107	-0.1549	-0.0078
Standard deviation	0.0719	0.0272	0.2983	0.0512
Sharpe ratio	-1.1566	-2.0212	-0.9710	-1.6458
Mean profit	0.0190	0.0094	0.0819	0.0263
Mean loss	-0.0650	-0.0226	-0.2486	-0.0432
Mean long return	0.0250	-0.0029	-0.0365	-0.0046
Mean short return	-0.0680	-0.0078	-0.1184	-0.0033
Mean trade length	67.4710	9.6425	72.0353	2.1840
Median trade length	71.0000	6.0000	75.0000	2.0000

Table 4.17: Trade statistics for top 20 unrestricted pairs, execution delay, July 2009–June 2018.

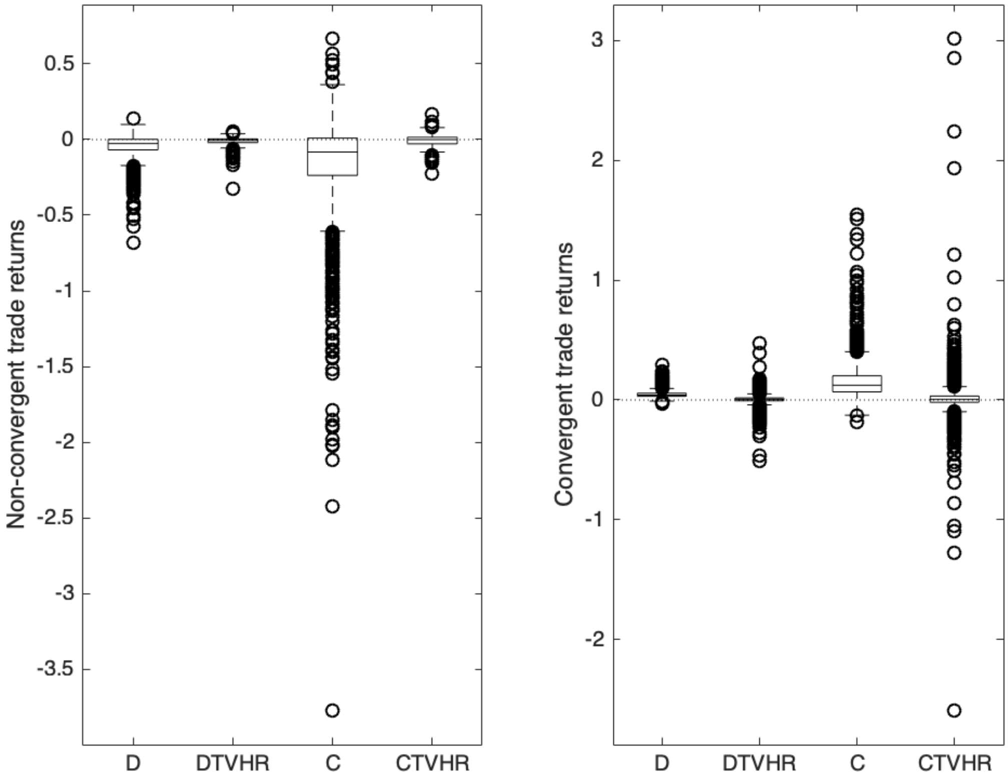

Figure 4.18: Distribution of trade excess return of top 20 unrestricted pairs, employed capital, execution delay, July 2009–June 2018.

models realise less extreme losses than their static counterparts accompanied by lower standard deviations. All models with the exception of static D pairs deliver negative returns on both their long and short positions, while D pairs manage to deliver a positive return on their long positions. Losses on non-convergent trades are largely driven by the short position for all models with the exception of the CTVHR model, which delivers the majority of its negative returns through its long positions. Mean trade length of non-convergent trades is greater than that of convergent trades for all model specifications but the CTVHR model, indicating that non-convergent CTVHR trades likely did not have sufficient time to converge at the end of the trading period before being liquidated.

The performance statistics of the statistical arbitrage model variants surveyed in Section 4.1.3 reveal the continuing decline of distance pair profitability in the first out-of-sample investigation of the conventional distance approach since Do and Faff (2010). By contrast, the extended study period reveals the continued profitability of the cointegration approach proposed in this thesis, with both the static and TVHR variants delivering greater excess returns than those observed in the final sub-period investigated in Section 4.1.2. Despite the continued decline of distance pair profitability, the TVHR model outperforms

	$z_d = 1$		$z_d = 2$		$z_d = 3$	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Mean	0.0068	0.0192	0.0033	0.0045	0.0011	0.0007
t-Statistic	12.2249	3.4986	11.4336	5.8259	6.5113	1.5452
Median	0.0056	0.0138	0.0029	0.0019	0.0008	0.0000
Standard deviation	0.0098	0.1595	0.0051	0.0205	0.0043	0.0115
Skewness	1.0869	-1.1207	1.5347	2.0185	0.5767	4.1585
Kurtosis	6.6403	29.4678	9.8503	18.9636	19.3559	104.0996
Minimum	-0.0203	-1.6709	-0.0098	-0.0978	-0.0339	-0.1140
Maximum	0.0635	1.0477	0.0367	0.1610	0.0384	0.1666
Observations < 0	0.2333	0.3985	0.2242	0.3076	0.3606	0.0258
Lower semi-deviation	0.0029	0.1070	0.0015	0.0101	0.0024	0.0062
Upper semi-deviation	0.0116	0.1197	0.0059	0.0184	0.0037	0.0096
Sharpe ratio	2.3828	0.4173	2.2514	0.7592	0.8922	0.1992
Sortino ratio	8.1407	0.6220	7.5051	1.5397	1.6151	0.3664
TIM	0.6536	0.6106	0.2081	0.0299	0.0281	0.0001
Return/TIM	0.0104	0.0315	0.0159	0.1500	0.0394	4.6793

Table 4.18: Excess return statistics for portfolios of top 20 unrestricted pairs, employed capital and execution delay, July 1962–June 2018. Sensitivity of returns to the choice of divergence parameter, z_d, is assessed across three levels—the default value of $z_d = 2$, in addition to $z_d = 1$ and $z_d = 3$.

the static model in terms of both absolute and risk-adjusted returns, while the TVHR specification of cointegration pairs outperforms the static model in risk-adjusted terms. Though the static cointegration model outperformed the TVHR model in absolute return terms, this outperformance was driven by a single monthly return of 57.19%, skewing the return statistics in its favour. A more statistically robust measure of monthly return, the monthly median return, reveals that the TVHR model outperforms its static counterparts for both distance and cointegration pairs.

4.1.4 Risk and Sensitivity Analysis

Parameter sensitivity with respect to spread divergence threshold and reversion rate parameter for the TVHR model is investigated in Section 4.1.4. Both static and TVHR models open a position in a statistical arbitrage opportunity once the tradable spread has diverged by a significant amount, z_d, normalised for spread volatility. The conventional static models use a divergence threshold of $z_d = 2$, indicating that a trading position is opened once the spread has diverged by two standard deviations. This threshold was adopted by the TVHR model for its implementation in Sections 4.1.1, 4.1.2, and 4.1.3. Table 4.18 reports the excess return statistics for TVHR portfolios of the top 20 unrestricted pairs with divergence parameters $z_d = \{1, 2, 3\}$ across the entire sample period, July 1962–June 2018.

The standard divergence parameter, $z_d = 2$, delivers a mean monthly return of 0.33% for DTVHR pairs and 0.45% for CTVHR pairs across the entire sample period. Despite its higher mean monthly returns, the CTVHR model delivers lower median returns than the DTVHR model, indicating that a relatively large proportion of the CTVHR mean return is attributable to extreme positive returns, further confirmed by its greater return kurtosis and positive skew. Increasing or decreasing the divergence parameter has a monotonic effect on mean monthly return, with $z_d = 1$ delivering the greatest mean monthly returns of 0.68% for the DTVHR model, and 1.92% for the CTVHR model. Unlike the original divergence parameter, $z_d = 1$ sees the CTVHR model delivering both greater mean and median returns than the DTVHR model, with mean return more than four-times greater than under $z_d = 2$. Conversely, a divergence parameter of $\mathrm{z}_d = 3$ delivers the lowest monthly mean returns, with the DTVHR model outperforming the CTVHR model in terms of both mean and median return. Median return and return standard deviation also monotonically increase as the divergence parameter decreases, while the proportion of negative monthly returns is lowest for the DTVHR model when $z_d = 2$, and lowest for the CTVHR model when $z_d = 3$. Of particular note is the very small proportion of negative monthly returns observed under the CTVHR model when $z_d = 3$ of 2.58%—one order of magnitude lower than the alternative model specifications.

Sharpe and Sortino ratios monotonically increase for the DTVHR model as the divergence parameter decreases, realising its highest ratios of 2.38 and 8.14, respectively, when $z_d = 1$. TIM monotonically increases and return per unit TIM monotonically decreases as the divergence parameter moves lower. In particular, the very low TIM of 0.01% and very high return per unit TIM of 467.93% of the CTVHR model when $z_d = 3$ indicates a high degree of efficient capital utilisation for high divergence parameters despite delivering relatively low portfolio mean returns. Conversely, low divergence parameters deliver high mean returns but utilise capital less efficiently.

Table 4.19 reports the pair statistics for TVHR portfolios of the top 20 unrestricted pairs with divergence parameters $z_d = \{1, 2, 3\}$. The greatest difference between choice of divergence parameter is realised in the proportion of Group 1 pairs. This proportion is lowest for both the DTVHR and CTVHR models when $z_d = 1$, and highest when $z_d = 3$. The smallest Group 1 proportion of 0.13% is observed under the DTVHR model, while the highest proportion of 99.37% is observed under the CTVHR model. This result is expected because a lower divergence threshold causes models to open positions with greater frequency given the greater occurrence of such deviations. The greatest difference in Group 1 proportions between DTVHR and CTVHR pairs is observed when $z_d = 2$, though Group 1 proportions for DTVHR pairs rise substantially from 1.67% when $z_d = 2$ to 63.95% when $z_d = 3$.

	$z_d = 1$		$z_d = 2$		$z_d = 3$	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Group 1 proportion	0.0013	0.0030	0.0167	0.4047	0.6395	0.9937
Group 2 proportion	0.0024	0.0011	0.0061	0.0034	0.0230	0.0002
Profitable proportion	0.1875	0.5714	0.3125	0.5682	0.4067	0.6667
Total return	-0.0500	0.0573	-0.0287	-0.0126	-0.0112	0.0147
Number of trades	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TIM	0.5666	0.3859	0.2596	0.0351	0.0727	0.0382
Industry-matched	0.2188	0.2143	0.1500	0.1136	0.2500	0.0000
Distance zero-crossings	28.2500	37.5000	27.6875	34.7045	32.1733	30.6667
Cointegration zero-crossings	27.2813	71.3571	24.6250	71.9545	31.8433	62.6667
BMD	0.6678	2.5907	0.7140	4.9396	0.7078	25.8744
Group 3 proportion	0.9962	0.9959	0.9772	0.5919	0.3375	0.0061
Profitable proportion	0.6468	0.5750	0.6441	0.5912	0.6201	0.5696
Total return	0.0412	0.1122	0.0204	0.0272	0.0071	0.0443
Number of trades	19.0797	37.6083	4.9036	2.7742	1.2647	1.3165
TIM	0.6533	0.6100	0.2103	0.0498	0.0777	0.0228
Industry-matched	0.1369	0.1342	0.1362	0.1447	0.1803	0.0886
Distance zero-crossings	40.6179	34.3704	40.5088	30.6968	36.2933	55.4304
Cointegration zero-crossings	43.0586	75.7424	42.8816	69.8373	37.6235	91.4684
BMD	0.5811	2.2562	0.5775	1.6626	0.6283	16.7419

Table 4.19: Pair statistics for top 20 unrestricted pairs, employed capital and execution delay, July 1962–June 2018. Sensitivity of returns to the choice of divergence parameter, z_d, is assessed across three levels—the default value of $z_d = 2$, in addition to $z_d = 1$ and $z_d = 3$.

Group 2 proportions remain relatively low across all model specifications with a minimum proportion of 0.02% for CTVHR pairs, and a maximum proportion of 2.30% for DTVHR pairs, both under a divergence parameter of $z_d = 3$. Profitable proportion and total return both monotonically increase for the DTVHR model as divergence parameter increases, while total return of the CTVHR model fluctuates for different divergence parameters. The CTVHR model delivers positive total returns for non-convergent Group 2 pairs when $z_d = 1$ and $z_d = 3$, with a substantial positive total return of 5.73% when $z_d = 1$. TIM decreases for DTVHR pairs as the divergence parameter increases, while TIM decreases to a minimum of 3.51% for CTVHR pairs and maintains that approximate proportion as the divergence parameter increases.

Group 3 proportions are most heavily influenced by Group 1 proportions rather than non-convergent Group 2 proportions. The greatest Group 3 proportion of 99.62% for the DTVHR model with divergence parameter $z_d = 1$ contrasts the lowest Group 3 proportion of 0.61% for the CTVHR model with divergence parameter $z_d = 3$. All Group 3 proportions monotonically decrease as the divergence parameter increases. Both TVHR models maintain relatively consistent profitable proportions of Group 3 pairs, with the DTVHR model profitable proportion at ∼63% and the CTVHR model profitable proportion at ∼57%. DTVHR total returns decline from their maximum of 4.12% when $z_d = 1$ to 0.71% when $z_d = 3$. By contrast, the CTVHR model delivers its maximum total return of 11.22% when $z_d = 1$ and its minimum total return of 2.72% when $z_d = 2$. Both number of trades and TIM monotonically decrease for DTVHR and CTVHR Group 3 pairs as the divergence parameter decreases. BMD is greater for all Group 2 pairs than Group 3 pairs, a finding consistent with the assertion that spread non-convergence is attributable to time-varying hedge ratios. In general, BMD is greatest under both TVHR specifications when the divergence parameter is greatest, $z_d = 3$.

Table 4.20 reports the individual trade statistics for TVHR portfolios of the top 20 unrestricted pairs with divergence parameters $z_d = \{1, 2, 3\}$. The profitable proportion of trades does not differ substantially across parameter specifications, though the mean return per trade monotonically increases for both distance and cointegration pairs as the divergence parameter increases. Sharpe ratio monotonically increases for the CTVHR model as the divergence parameter increases, delivering its greatest Sharpe ratio of 1.69 when $z_d = 3$. By contrast, the DTVHR model achieves its greatest Sharpe ratio of 0.87 when $z_d = 2$. For both model specifications, mean profit and mean loss are greatest when $z_d = 3$. Similarly, the mean returns of long and short positions realise their greatest magnitude when $z_d = 3$, though the DTVHR short position delivers a negative return which is roughly one-third the magnitude of its long position. Trade length monotonically increases for the DTVHR model, but remains relatively consistent for the CTVHR model as the divergence parameter increases.

	$z_d = 1$		$z_d = 2$		$z_d = 3$	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
All trades						
Profitable proportion	0.5819	0.5443	0.6136	0.5671	0.6086	0.5926
Mean return	0.0021	0.0030	0.0041	0.0096	0.0047	0.0334
Standard deviation	0.0273	0.1018	0.0323	0.1098	0.0493	0.2138
Sharpe ratio	0.6009	0.3254	0.8659	0.9320	0.5444	1.6915
Mean profit	0.0167	0.0592	0.0197	0.0573	0.0261	0.1401
Mean loss	-0.0181	-0.0642	-0.0207	-0.0528	-0.0286	-0.1219
Mean long return	0.0031	0.0037	0.0047	0.0048	0.0075	0.0191
Mean short return	-0.0010	-0.0008	-0.0006	0.0049	-0.0028	0.0143
Mean trade length	4.2928	2.0083	5.4060	2.2257	7.7407	2.1481
Median trade length	3.0000	2.0000	4.0000	2.0000	6.0000	2.0000
Convergent trades	0.9633	0.9831	0.9480	0.9786	0.9165	0.9630
Profitable proportion	0.5885	0.5446	0.6241	0.5677	0.6268	0.5962
Mean return	0.0026	0.0030	0.0048	0.0098	0.0061	0.0354
Standard deviation	0.0268	0.1017	0.0317	0.1105	0.0497	0.2172
Sharpe ratio	0.7357	0.3297	1.0340	0.9426	0.7039	1.7587
Mean profit	0.0168	0.0592	0.0199	0.0576	0.0267	0.1432
Mean loss	-0.0178	-0.0642	-0.0203	-0.0529	-0.0284	-0.1237
Mean long return	0.0033	0.0037	0.0050	0.0048	0.0082	0.0198
Mean short return	-0.0007	-0.0007	-0.0002	0.0050	-0.0021	0.0156
Mean trade length	4.2608	2.0091	5.3762	2.2300	7.7162	2.1635
Median trade length	3.0000	2.0000	4.0000	2.0000	6.0000	2.0000
Non-convergent trades	0.0367	0.0169	0.0520	0.0214	0.0835	0.0370
Profitable proportion	0.4090	0.5226	0.4216	0.5376	0.4089	0.5000
Mean return	-0.0090	0.0008	-0.0084	0.0013	-0.0108	-0.0185
Standard deviation	0.0364	0.1072	0.0401	0.0697	0.0425	0.0882
Sharpe ratio	-1.7232	0.0832	-1.3685	0.2096	-1.4261	-2.5174
Mean profit	0.0136	0.0577	0.0147	0.0435	0.0163	0.0464
Mean loss	-0.0246	-0.0615	-0.0253	-0.0477	-0.0295	-0.0834
Mean long return	0.0000	0.0052	0.0002	0.0010	-0.0004	0.0028
Mean short return	-0.0090	-0.0044	-0.0086	0.0003	-0.0104	-0.0213
Mean trade length	5.1317	1.9643	5.9484	2.0323	8.0101	1.7500
Median trade length	3.0000	2.0000	4.0000	2.0000	5.0000	2.0000

Table 4.20: Trade statistics for top 20 unrestricted pairs, employed capital and execution delay, July 1962–June 2018. Sensitivity of returns to the choice of divergence parameter, z_d, is assessed across three levels—the default value of $z_d = 2$, in addition to $z_d = 1$ and $z_d = 3$.

The proportion of convergent trades declines for both model specifications as the divergence parameter increases, though it still remains high with convergent trade proportions above 90% for all model specifications. The profitable proportion, mean return, and standard deviation of returns monotonically increase for both models as the divergence parameter increases, with both models delivering their highest mean return of 0.61% for the DTVHR model and 3.54% for the CTVHR model when $z_d = 3$. Given the high proportion of convergent trades, convergent trade statistics closely resemble the statistics for all trades. Non-convergent trade statistics differ from those of convergent trades, delivering their greatest profitable proportion, mean return, and Sharpe ratio when $z_d = 2$. Mean trade length for non-convergent DTVHR trades is greater than corresponding convergent trades under the various divergence parameters, while the trade length of non-convergent CTVHR trades is shorter than corresponding convergent trades.

The choice of divergence parameter has a significant effect on the characteristics of TVHR trades and portfolios. While greater absolute portfolio returns are delivered with lower values of the divergence parameter, they come at the expense of reduced efficiency, with a greater number of trades and diminished return per trade for both model specifications. The arbitrageur therefore has a choice to make, between a high portfolio return with low return per trade and high number of transactions, or a low portfolio return with high return per trade and low number of transactions—too low a divergence parameter, z_d, will diminish per-trade returns beyond the point of feasibility after consideration of transaction costs. In this thesis, transaction costs are treated implicitly following Gatev, Goetzmann, and Rouwenhorst (2006), a discussion of which is presented in Section 4.1.5.

As an alternative to the variation of divergence parameter, the arbitrageur may choose to tune the speed with which the TVHR model updates its time-varying estimate of the hedge ratio. The reversion rate parameter, α, is estimated from the number of spread zero-crossings, $\mathbb{E}[D]$, observed during the formation period in accordance with Equation (3.58). As discussed in Sections 4.1.1, 4.1.2, and 4.1.3, the short duration of TVHR distance and cointegration trades indicates that the reversion rate parameter may be too fast, estimated from in-sample observations that do not correspond with out-of-sample realisations. Consequently, Table 4.21 reports the excess return statistics for TVHR portfolios of the top 20 unrestricted pairs with $\alpha = f(\{0.1\mathbb{E}[D], 0.5\mathbb{E}[D], \mathbb{E}[D]\})$ across the entire sample period, July 1962–June 2018; that is, the number of formation period zero-crossings is artificially reduced to one-half and one-tenth their observed number, respectively.

The results reported under the $\mathbb{E}[D]$ columns correspond with the reversion rate parameter estimated and used in the preceding sections of this thesis. By deflating the number

	$0.1\mathbb{E}[D]$		$0.5\mathbb{E}[D]$		$\mathbb{E}[D]$	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Mean	0.0053	0.0131	0.0057	0.0145	0.0033	0.0045
t-Statistic	9.0697	3.8813	11.1739	5.2223	11.4336	5.8259
Median	0.0046	0.0109	0.0048	0.0114	0.0029	0.0019
Standard deviation	0.0098	0.0990	0.0082	0.0685	0.0051	0.0205
Skewness	0.7074	-2.2603	1.2377	4.2864	1.5347	2.0185
Kurtosis	7.8725	42.4244	8.1602	86.1670	9.8503	18.9636
Minimum	-0.0480	-0.9819	-0.0217	-0.5314	-0.0098	-0.0978
Maximum	0.0624	0.8150	0.0527	1.0381	0.0367	0.1610
Observations < 0	0.2848	0.3394	0.2136	0.2970	0.2242	0.3076
Lower semi-deviation	0.0040	0.0716	0.0024	0.0372	0.0015	0.0101
Upper semi-deviation	0.0104	0.0694	0.0097	0.0592	0.0059	0.0184
Sharpe ratio	1.8502	0.4595	2.4355	0.7317	2.2514	0.7592
Sortino ratio	4.5906	0.6349	8.1487	1.3478	7.5051	1.5397
TIM	0.5805	0.5465	0.3416	0.2315	0.2081	0.0299
Return/TIM	0.0091	0.0240	0.0168	0.0625	0.0159	0.1500

Table 4.21: Excess return statistics for portfolios of top 20 unrestricted pairs, employed capital and execution delay, July 1962–June 2018. Sensitivity of returns to the choice of reversion parameter, $\mathbb{E}[D]$, is assessed across three levels—the default value of $\mathbb{E}[D]$, in addition to $0.1\mathbb{E}[D]$ and $0.5\mathbb{E}[D]$.

of zero-crossings observed in the formation period, the time-varying estimate of the hedge ratio updates more slowly, approaching the performance and trade characteristics of static distance and cointegration models. The mean return of the DTVHR model almost doubles, increasing from a monthly mean of 0.33% to 0.57% as the number of formation period zero-crossings decreases to $0.5\mathbb{E}[D]$. Similarly, the mean return of the CTVHR model more than trebles, from 0.45% to 1.45%. The mean return of both models declines marginally as formation period zero-crossings move from $0.5\mathbb{E}[D]$ to $0.1\mathbb{E}[D]$, but still remain substantially more profitable than when $\mathbb{E}[D]$ is used. Median return is also highest for both model variants when using $0.5\mathbb{E}[D]$, though standard deviation of returns continues to increase as zero-crossings move toward $0.1\mathbb{E}[D]$.

Skewness and kurtosis monotonically decline for the DTVHR model, while both skewness and kurtosis for the CTVHR model achieve their highest values when $0.5\mathbb{E}[D]$ is used. Both minimum and maximum monthly DTVHR returns monotonically increase in magnitude as zero-crossings approach $0.1\mathbb{E}[D]$, while the maximum monthly CTVHR return is delivered under $0.5\mathbb{E}[D]$. Both model specifications also deliver their smallest proportion of negative monthly returns when $0.5\mathbb{E}[D]$ is used. The DTVHR model delivers its greatest Sharpe and Sortino ratios, and return per unit TIM when $0.5\mathbb{E}[D]$ is used, while the CTVHR model delivers its greatest Sharpe and Sortino ratios, and return per unit TIM when the default value of $\mathbb{E}[D]$ is used.

Table 4.22 reports the pair statistics for TVHR portfolios of the top 20 unrestricted pairs

	$0.1\mathbb{E}[D]$		$0.5\mathbb{E}[D]$		$\mathbb{E}[D]$	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Group 1 proportion	0.0238	0.0279	0.0127	0.0097	0.0167	0.4047
Group 2 proportion	0.2943	0.2163	0.0338	0.0018	0.0061	0.0034
Profitable proportion	0.1980	0.1558	0.2132	0.4583	0.3125	0.5682
Total return	-0.0627	-0.3041	-0.0524	-0.0304	-0.0287	-0.0126
Number of trades	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TIM	0.7159	0.6896	0.4353	0.2843	0.2596	0.0351
Industry-matched	0.1636	0.1317	0.2200	0.1250	0.1500	0.1136
Distance zero-crossings	37.2409	32.5025	27.8639	30.5833	27.6875	34.7045
Cointegration zero-crossings	39.1280	74.4306	27.2177	72.5000	24.6250	71.9545
BMD	0.6790	2.8560	0.9292	1.1135	0.7140	4.9396
Group 3 proportion	0.6819	0.7558	0.9535	0.9884	0.9772	0.5919
Profitable proportion	0.8622	0.7683	0.7191	0.6340	0.6441	0.5912
Total return	0.0730	0.1903	0.0380	0.0870	0.0204	0.0272
Number of trades	2.4888	2.5595	3.4725	5.8471	4.9036	2.7742
TIM	0.5391	0.5208	0.3413	0.2321	0.2103	0.0498
Industry-matched	0.1238	0.1359	0.1330	0.1342	0.1362	0.1447
Distance zero-crossings	42.2743	35.0511	41.1786	34.4137	40.5088	30.6968
Cointegration zero-crossings	44.9415	76.0510	43.7432	75.7525	42.8816	69.8373
BMD	0.5444	2.1385	0.5707	2.2598	0.5775	1.6626

Table 4.22: Pair statistics for top 20 unrestricted pairs, employed capital and execution delay, July 1962–June 2018. Sensitivity of returns to the choice of reversion parameter, $\mathbb{E}[D]$, is assessed across three levels—the default value of $\mathbb{E}[D]$, in addition to $0.1\mathbb{E}[D]$ and $0.5\mathbb{E}[D]$.

with $\alpha = f(\{0.1\mathbb{E}[D], 0.5\mathbb{E}[D], \mathbb{E}[D]\})$. Group 1 proportions for both models are lowest when $0.5\mathbb{E}[D]$ is used. In particular, the substantial reduction in Group 1 proportions for the CTVHR model from 40.47% under $\mathbb{E}[D]$ to 0.97% under $0.5\mathbb{E}[D]$ illustrates the significant effect that a slower reversion parameter has on the proportion of pairs that fail to open a position during the trading period. A slower reversion parameter means that the TVHR model is less likely to update its estimate of the hedge ratio before a position can be opened, resulting in a greater proportion of pairs that open at least one position during the trading period.

Group 2 proportions monotonically increase for the DTVHR model, from 0.61% to 29.43%, as formation period zero-crossings approach $0.1\mathbb{E}[D]$. Group 2 proportions for the CTVHR model, on the other hand, deliver their lowest figure of 0.18% when $0.5\mathbb{E}[D]$ is used. The Group 2 proportion for both models is greater than 20% when $0.1\mathbb{E}[D]$ is used. The profitable proportion and total return monotonically decline for both models as zero-crossings approach $0.1\mathbb{E}[D]$, while TIM monotonically increases. This result is unsurprising given the TVHR model's tendency to more closely resemble conventional static models as the number of formation period zero-crossings decline. In general, the less volatile a pair is during the formation period, as indicated by the number of formation period zero-crossings, the more likely such a pair will be recorded among non-convergent Group 2 pairs. This tendency is attributable to the slower reversion rate of the TVHR model estimated from the low number of formation period zero-crossings. By artificially slowing down the reversion rate parameter toward the value calculated in reference to $0.1\mathbb{E}[D]$ zero-crossings, a greater proportion of relatively unstable pairs find themselves included among the Group 2 proportion. The number of distance zero-crossings, for example, are closer among Group 2 and Group 3 pairs when $0.1\mathbb{E}[D]$ is used than when $\mathbb{E}[D]$ is used.

Group 3 proportions are highest for the DTVHR model at 97.72% when $\mathbb{E}[D]$ is used, and highest for the CTVHR model at 98.84% when $0.5\mathbb{E}[D]$ is used. The profitable proportion and total return for both models monotonically increases as $0.1\mathbb{E}[D]$ is approached, while number of trades monotonically declines for the DTVHR model but fluctuates for the CTVHR model. The fluctuation in number of trades for the CTVHR model illustrates the balancing act of slowing down the reversion rate sufficiently to allow more trades to be placed, but not enough to cause trade lengths to extend substantially and consequently preclude further trading opportunities. As with Group 2 pairs, TIM increases for both models as $0.1\mathbb{E}[D]$ is approached. For all model specifications with the exception of the CTVHR model under $0.5\mathbb{E}[D]$, BMD is greater for Group 2 pairs than Group 3 pairs.

Table 4.23 reports the individual trade statistics for TVHR portfolios of the top 20

	$0.1\mathbb{E}[D]$		$0.5\mathbb{E}[D]$		$\mathbb{E}[D]$	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
All trades						
Profitable proportion	0.7225	0.6850	0.6901	0.6004	0.6136	0.5671
Mean return	0.0157	0.0357	0.0103	0.0148	0.0041	0.0096
Standard deviation	0.0702	0.3279	0.0452	0.1553	0.0323	0.1098
Sharpe ratio	0.5883	0.3093	1.0116	0.6806	0.8659	0.9320
Mean profit	0.0457	0.1617	0.0289	0.0851	0.0197	0.0573
Mean loss	-0.0625	-0.2383	-0.0312	-0.0908	-0.0207	-0.0528
Mean long return	0.0253	0.0354	0.0110	0.0092	0.0047	0.0048
Mean short return	-0.0096	0.0004	-0.0007	0.0056	-0.0006	0.0049
Mean trade length	36.3395	31.2930	12.7425	4.9391	5.4060	2.2257
Median trade length	23.0000	19.0000	9.0000	4.0000	4.0000	2.0000
Convergent trades	0.6380	0.6964	0.8843	0.9538	0.9480	0.9786
Profitable proportion	0.9612	0.8578	0.7310	0.6075	0.6241	0.5677
Mean return	0.0472	0.1302	0.0140	0.0165	0.0048	0.0098
Standard deviation	0.0420	0.2021	0.0423	0.1555	0.0317	0.1105
Sharpe ratio	3.6379	2.1507	1.4987	0.7575	1.0340	0.9426
Mean profit	0.0497	0.1675	0.0295	0.0858	0.0199	0.0576
Mean loss	-0.0136	-0.0951	-0.0282	-0.0907	-0.0203	-0.0529
Mean long return	0.0342	0.0736	0.0124	0.0097	0.0050	0.0048
Mean short return	0.0130	0.0566	0.0016	0.0068	-0.0002	0.0050
Mean trade length	24.0682	22.6137	12.2661	4.9507	5.3762	2.2300
Median trade length	16.0000	15.0000	8.0000	4.0000	4.0000	2.0000
Non-convergent trades	0.3620	0.3036	0.1157	0.0462	0.0520	0.0214
Profitable proportion	0.3019	0.2887	0.3773	0.4545	0.4216	0.5376
Mean return	-0.0399	-0.1809	-0.0180	-0.0206	-0.0084	0.0013
Standard deviation	0.0752	0.4393	0.0555	0.1461	0.0401	0.0697
Sharpe ratio	-1.1076	-0.9137	-1.2752	-1.0335	-1.3685	0.2096
Mean profit	0.0234	0.1221	0.0201	0.0655	0.0147	0.0435
Mean loss	-0.0673	-0.3039	-0.0411	-0.0924	-0.0253	-0.0477
Mean long return	0.0094	-0.0523	0.0005	-0.0018	0.0002	0.0010
Mean short return	-0.0494	-0.1286	-0.0185	-0.0188	-0.0086	0.0003
Mean trade length	57.9640	51.1998	16.3846	4.6997	5.9484	2.0323
Median trade length	55.0000	42.0000	9.0000	3.0000	4.0000	2.0000

Table 4.23: Trade statistics for top 20 unrestricted pairs, employed capital and execution delay, July 1962–June 2018. Sensitivity of returns to the choice of reversion parameter, $\mathbb{E}[D]$, is assessed across three levels—the default value of $\mathbb{E}[D]$, in addition to $0.1\mathbb{E}[D]$ and $0.5\mathbb{E}[D]$.

unrestricted pairs with $\alpha = f(\{0.1\mathbb{E}[D], 0.5\mathbb{E}[D], \mathbb{E}[D]\})$. Profitable proportion, mean return, standard deviation, mean profit, mean loss, mean long return, and mean and median trade length all monotonically increase for both TVHR models as zero-crossings approach $0.1\mathbb{E}[D]$. The DTVHR model delivers its greatest profitable proportion of 72.25% and its greatest mean return of 1.57% when $0.1\mathbb{E}[D]$ is used. This return is comparable to the mean return of the static distance approach, whose unpublished mean return of 1.44% per trade slightly underperforms the DTVHR model. Similarly, the unpublished mean return of the static cointegration model of 3.23% per trade over the entire sample period slightly underperforms the CTVHR model return of 3.57% per trade. The Sharpe ratio is greatest for the DTVHR model when $0.5\mathbb{E}[D]$ is used, while the CTVHR model delivers its greatest Sharpe ratio under the default $\mathbb{E}[D]$ value.

Profitable proportion and mean return monotonically increase for both model specifications of convergent trades, and monotonically decline for non-convergent trades in alignment with the tendency to more closely resemble conventional static models as $0.1\mathbb{E}[D]$ is approached. The proportion of convergent trades is greatest for both models when $\mathbb{E}[D]$ is used and lowest when $0.1\mathbb{E}[D]$ is used. As the reversion rate slows, trade lengths increase as the spread is less likely to artificially converge due to the time-varying estimate of the hedge ratio. Non-convergence is therefore a function of both the greater average trade length leaving less time for the spread to converge before the end of the trading period, and also the tendency for the spread of non-convergent trades to continue to diverge in an affirmation of arbitrage risk. The lower proportion of convergent trades when $0.1\mathbb{E}[D]$ is used is a reflection of these two competing influences.

If the mean trade length of convergent and non-convergent trades were roughly equivalent, a higher proportion of non-convergent trades would be attributable to the limitations placed on trade frequency by greater trade lengths. The significant disparity between mean trade lengths of convergent and non-convergent trades when $0.1\mathbb{E}[D]$ is used, however, indicates that a substantial proportion of non-convergence is attributable to continued spread divergence as spread equilibrium shifts to a new level faster than the TVHR model can accommodate. Roughly equivalent mean trade lengths between convergent and non-convergent trades may therefore be considered the threshold beyond which further slowing of the reversion rate invites a disproportionate increase in arbitrage risk.

The sensitivity analysis conducted in Section 4.1.4 indicates a greater level of sensitivity of the TVHR model to the choice of divergence parameter, z_d, than to the choice of reversion rate parameter, $\alpha = f(\mathbb{E}[D])$. While the mean portfolio return for DTVHR and CTVHR pairs improves when $z_d = 1$ is used, as reported in Table 4.18, mean return per trade declines, as reported in Table 4.20, and vice versa when $z_d = 3$ is used. Lower

	D		DTVHR		C		CTVHR	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
Intercept	0.0002	14.7915***	0.0001	15.3837***	0.0006	1.8614*	0.0002	4.3143***
R_M	-0.0251	-14.5978***	-0.0047	-4.7839***	-0.2114	-6.2764***	-0.0024	-0.4952
SMB	-0.0092	-3.0286***	-0.0044	-2.5357**	-0.1336	-2.2447**	0.0080	0.9249
HML	-0.0188	-4.9149***	0.0023	1.0802	0.0311	0.4149	0.0129	1.1803
MOM	-0.0300	-12.7255***	0.0097	7.2635***	-0.1067	-2.3116**	0.0054	0.8017
STR	0.0295	12.5384***	0.0205	15.4159***	0.0981	2.1293**	0.0177	2.6259***
LTR	0.0157	3.8109***	-0.0010	-0.4277	-0.0405	-0.5014	-0.0060	-0.5119
R^2	0.0299		0.0203		0.0037		0.0007	

Table 4.24: Regression analysis of daily excess returns, July 1962–June 2018. The influence of the market equity premium (R_M), Fama-French factors Small Minus Big (SMB) and High Minus Low (HML), in addition to momentum (MOM), short-term reversion (STR) and long-term reversion (LTR) factors is assessed. Factor loadings are either insignificant or significant at the 90% confidence level (*), 95% confidence level (**), or 99% confidence level (***).

values of $\mathbb{E}[D]$, by contrast, improve both mean portfolio return, as reported in Table 4.21, and mean return per trade, as reported in Table 4.23. The greatest mean portfolio return for both DTVHR and CTVHR pairs under variations of the reversion parameter, which is achieved when $0.5\mathbb{E}[D]$ is used, delivers a mean monthly return of 0.57% for DTVHR pairs, and 1.45% for CTVHR pairs. These returns outperform the unpublished static distance and cointergration model mean returns of 0.50% and 1.13%, respectively, while retaining the high convergence rates offered by the TVHR model.

Risk characteristics of the conventional static distance and cointegration models, in addition to their TVHR counterparts, are investigated in Table 4.24 across the entire sample period of July 1962–June 2018. The Fama-French factors Small Minus Big and High Minus Low (Fama and French, 1992; Fama and French, 1993) are augmented by momentum (Jegadeesh and Titman, 1993; Carhart, 1997), and both short-term and long-term reversion (Fama and French, 2000) factors. The regression analysis reports the risk-adjusted daily returns of the statistical arbitrage models relative to the S&P 500 market equity premium.

All model variants produce statistically significant daily excess returns at the 99% confidence level, with the exception of the static C model whose returns are significant at the 90% confidence level. The excess returns of the static C model are greatest in magnitude, followed jointly by the D and CTVHR models, with the DTVHR model delivering the smallest excess return. All models are negatively correlated with the market equity premium, indicating that statistical arbitrage performs well when the market performs poorly, with all models except the CTVHR model statistically significant at the

	D		DTVHR		C		CTVHR	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
Intercept	-1.95E-05	-0.2972	2.40E-05	0.5778	3.00E-04	0.3393	9.52E-05	0.6420
R_M	-1.27E-02	-6.8856***	-1.81E-03	-1.5474	-2.13E-01	-8.5956***	-2.20E-03	0.7021
VIX	9.37E-06	3.6113***	6.94E-06	4.2255***	2.01E-05	0.5746	2.37E-06	0.7698
TT	-2.14E-08	-2.1930**	-1.98E-08	-3.2095***	-4.89E-08	-0.3726	1.80E-08	0.5542
R^2	0.0101		0.0047		0.0106		0.0001	

Table 4.25: Regression analysis of daily excess returns, July 1962–June 2018. The influence of the market equity premium (R_M), CBOE Volatility Index (VIX) and linear time trend (TT) factors is assessed. Factor loadings are either insignificant or significant at the 90% confidence level (*), 95% confidence level (**), or 99% confidence level (***).

99% level. The D, DTVHR and C models have a negative, statistically significant loading on the SMB factor, while the CTVHR model has a positive, statistically insignificant loading. Only the static D model has a negative, statistically significant loading on the HML factor, all other model loadings being positive and insignificant.

Significant negative loadings on the momentum factor for the static D and C models indicate negative correlation between momentum and conventional statistical arbitrage. By contrast, the TVHR models have a small positive loading on momentum, though only the DTVHR model is significant. All models have positive, statistically significant loadings on the short-term reversion factor, while only the static D model has a positive, statistically significant loading on long-term reversion. The relatively short duration of statistical arbitrage opportunities, especially for the DTVHR, C and CTVHR models, explains why none of the latter models are significantly correlated with long-term reversion. The longer duration of static D model trades is sufficiently long to straddle the interface between short- and long-term reversion, on the other hand. Despite the significance of a number of factors, the coefficient of determination for each model's factor regression is relatively low, with the greatest proportion of return variance explained by the model for static D pair returns at only 2.99%.

To investigate the influence of market volatility, and the magnitude of declining profitability, Table 4.25 reports an alternative factor regression that incorporates VIX and linear time trend factors in addition to the equity market premium. All models with the exception of the static D model deliver positive intercept coefficients, though none of the coefficients are statistically significant. The only positive, statistically significant factor loading for the static D model is on the VIX factor. Both the D and DTVHR models have negative, statistically significant loadings on the linear time trend, indicating a statistically verifiable decline in profitability over the duration of the entire study period. The only other significant factor loading for the DTVHR model is a positive coefficient on the VIX factor. All models have a positive loading on the VIX factor, while only the CTVHR model has a positive albeit statistically insignificant loading on the linear time trend, indicating that it becomes more profitable toward the end of the sample period. Only the static D and C models report a statistically significant factor loading on the market premium. Despite these results being relatively consistent with the reported empirical performance of the various models, the coefficient of determination for all regressions in Table 4.25 is very low.

4.1.5 Transaction Costs

The results reported in Section 4.1 have been calculated without explicit consideration of transaction costs, in accordance with the methodology developed by Gatev, Goetzmann, and Rouwenhorst (2006) and replicated by Do and Faff (2010). As discussed by Gatev, Goetzmann, and Rouwenhorst (2006), the one-day execution delay imposed in their framework is expected to reduce the excess returns of pairs by the round-trip transaction cost, on average. In the extreme case, supposing the price of the short positions are ask prices and the long positions are bid prices at the time of trade signal generation, delaying the execution by one day will reduce the excess returns of the pair by half the sum of the spreads for the long and short positions. If the opposite occurs at the close of the position, again in the extreme case, excess returns are again reduced by half the sum of the individual security bid-ask spreads. The execution delay therefore offers a very conservative estimate of statistical arbitrage profitability, as the reduction in profitability includes the full round-trip transaction cost in addition to the opportunity cost of rapid convergence during the first day. Estimates of real transaction costs calculated by Gatev, Goetzmann, and Rouwenhorst (2006) and Rad, Low, and Faff (2016) confirm the conservatism of this estimation procedure.

4.1.6 Summary of Proposed TVHR Model

Section 4.1 reports the statistical and economic performance of the standard distance approach, a novel cointegration approach developed in this thesis, and their adaptive TVHR extensions intended to address the first research question posited in this thesis, specifically: *is the assumption of static arbitrage relationships responsible for the declining profitability of statistical arbitrage?* Following the testing procedure used by Gatev, Goetzmann, and Rouwenhorst (2006), overlapping portfolios of the top 20 distance and cointegration pairs are formed and traded in the six-month period following their formation. Portfolios are re-estimated on a monthly basis, with six uniformly-weighted portfolios held at any given time.

Section 4.1.1 investigates the performance of the four statistical arbitrage variants over the initial study period of July 1962–June 2009, considering the impact of committed and employed capital allocation schemes, delayed trade initiation, and industry restrictions on security pairings. The proposed TVHR model accepts pairs selected under distance and cointegration specifications but employs a time-varying estimate of the hedge ratio to generate trade signals. When positions are opened immediately following a trade signal, the TVHR model delivers comparable portfolio returns to those of the conventional static model for distance pairs, but diminished portfolio returns for cointegration pairs relative to the static model. However, the TVHR model delivers significantly greater risk-adjusted returns and more favourable return distributions than its static analogue. In particular, the efficiency of the TVHR model is evident in its brief trade duration, entering positions to extract small amounts of profit before closing the positions shortly after. The TVHR model consequently spends much less time in the market than the conventional static model, allowing for the efficient allocation of surplus capital by the arbitrageur.

Individual pair and trade statistics under the different model specifications reveal the limited per-trade profitability of the TVHR model. Rather, the TVHR model realises its relatively high risk-adjusted returns by executing low-profitability trades more frequently in the case of distance pairs, and limiting the duration of trades and potential for loss in the case of cointegration pairs. The cointegration approach developed for this thesis realises greater portfolio, pair, and trade returns than the conventional distance approach at the expense of return volatility and, consequently, risk-adjusted returns. The distance approach therefore represents a more consistent albeit less profitable strategy for exploiting arbitrage relationships than the cointegration approach, with TVHR model extensions improving the risk-adjusted returns of each.

The diminished per-trade performance of the TVHR model is particularly susceptible to the impact of imposing a one-day delay before executing trades. The trade delay, imposed to temper a potential upward return bias caused by bid-ask bounce, as well as to simulate typical transaction costs, reduces the profitability of TVHR pairs by a proportionally greater degree than their static counterparts. Despite this reduced profitability in absolute terms, the TVHR portfolios continue to outperform their static counterparts in terms of risk-adjusted returns.

Restricting pair constituents to one of four major industries, namely Industrials, Transportation, Utilities and Financials, broadly improves the absolute returns of distance TVHR portfolios at the expense of risk-adjusted returns, and improves both absolute and risk-adjusted returns of cointegration TVHR portfolios. There is some decline in risk-adjusted returns for some combinations of TVHR portfolios, though this is likely due to the lack of diversification caused by pair restriction.

The objective of the TVHR model is to reduce the proportion of non-convergent trades which, according to Do and Faff (2010), constitute the manifestation of arbitrage risk and are consequently the driver of declining statistical arbitrage profitability. Given the near-universal convergence of TVHR pairs, a sub-period analysis is presented in Section 4.1.2 to determine whether TVHR pairs realise the same declining profitability as conventional static pairs.

The sub-period analysis reveals that static distance and cointegration pairs, in addition to TVHR distance pairs, exhibit declining profitability between subsequent sub-periods. The declining profitability of distance pairs is attributable largely to longer trade lengths and diminished returns per trade, a symptom of declining spread reversion rates. By contrast, cointegration TVHR pairs regain some of their profitability in the final sub-period after first losing half their profitability in the second sub-period. Both static and TVHR pairs maintain or improve their risk-adjusted returns between subsequent sub-periods, though only the TVHR model improves its absolute portfolio returns.

Declining distance pair profitability continues in the first out-of-sample analysis of the conventional distance approach since its investigation by Do and Faff (2010). Section 4.1.3 reports the economic and statistical performance of the four model variants across July 2009–June 2018. The static distance approach delivers a mean portfolio return of 0.00% over the period while its TVHR analogue delivers a small positive return. Though the TVHR model's declining profitability is not as pronounced as that of the static distance model, it nevertheless affirms the trend. By contrast, both the static and TVHR cointegration models continue to become more profitable relative to previous sub-periods, with the TVHR model delivering its greatest absolute and risk-adjusted returns out of any sub-period.

The assertion that declining statistical arbitrage profitability is attributable to spread non-convergence is not validated by the empirical data presented in this thesis. The TVHR model specifically addresses the issue of non-convergence, ensuring that a significant majority of trades converge during the trading period, yet distance TVHR pairs are just as susceptible to the trend of declining profitability as their conventional static counterparts, indicating that non-convergence is not the cause. Nor is increasing market efficiency indicated as the cause of declining profitability, due to the increasing profitability of cointegration pairs and their TVHR extension in particular. Rather, declining profitability appears to be due to the failure of the distance approach in selecting tradable pairs.

The distance approach selects paired securities whose SSD is minimised over a formation period. This robust yet simple selection method seeks to minimise the stochastic drift between two securities and, by proxy, maximise the probability of their co-evolution. An unintended consequence of the procedure is that spread volatility is minimised. Spread volatility is the profitable component of a statistical arbitrage relationship, and its diminution over the years and across industries indicates a declining level of market-wide volatility. The cointegration approach outperforms the distance approach in the most recent sub-period because it specifically seeks to maximise spread volatility at the expense of a high probability of security co-evolution within the pair. The short duration of TVHR trades, however, largely negate the necessity for stable cointegrating relationships which could otherwise lead to protracted divergence and loss-making spread non-convergence.

Sensitivity of the TVHR model parameters is assessed in Section 4.1.4. The model is specified by its divergence parameter—the point at which a normalised spread is considered to be statistically mis-priced—and its reversion rate parameter, which specifies the speed with which the time-varying hedge ratio is re-estimated. Varying the divergence parameter presents the arbitrageur with a trade-off between portfolio profitability on the one hand and trade profitability on the other; the greater the degree of portfolio profitability, the lower the per-trade profitability, and vice versa. Varying the reversion rate parameter, by contrast, indirectly influences the average trade duration. Variation of the reversion rate allows the arbitrageur to more closely replicate the dynamics of static statistical arbitrage models by slowing the reversion rate down. The resultant portfolio inherits the high per-trade profitability of the static models while retaining the high risk-adjusted returns and guaranteed convergence, given sufficient time, of the TVHR model. The reversion rate parameter therefore represents a more attractive avenue for tuning model performance.

4.2 Regime Switching Model Extension

Section 4.2 considers the use of a logistic regression model, in an extension of the proposed TVHR model, which seeks to determine whether a trading signal is likely to profitably exploit a statistical arbitrage opportunity. In doing so, the logistic regression extension seeks to classify the trade as belonging to either a profitable or unprofitable regime. The logistic regression model considers only two variables, namely the level of pair-specific volatility, and market-wide volatility, both at the time of trade initiation. This logistic regression model with volatility-based features is intended to address the second research question posited in this thesis, specifically: *are statistical arbitrage returns dependent on the prevailing volatility regime?* Using observed trade outcomes during an in-sample

	DTVHR		CTVHR	
	Coefficient	t-Statistic	Coefficient	t-Statistic
Intercept	0.4353	12.0802***	0.2303	3.9164***
Pair Volatility	0.0764	5.3309***	0.0432	1.7798*
VIX	-0.0020	-1.2240	-0.0024	-0.9424
Minimum $Pr(R \geq 0)$	0.5139		0.5049	
Maximum $Pr(R \geq 0)$	0.7005		0.5984	
AUROC	0.5087		0.5062	

Table 4.26: Logistic regression statistics for binary classification of trade outcomes based on pair-specific volatility at the time of trade initiation (Pair Volatility), and the level of the CBOE Volatility Index (VIX) at the time of trade initiation. Factor loadings are either insignificant or significant at the 90% confidence level (*), 95% confidence level (**), or 99% confidence level (***). Minimum and maximum model-forecast probabilities of trade success are reported in addition to Area Under Receiver Operating Characteristics (AUROC) statistics.

period spanning January 1990–July 2009, the beginning of which aligns with the inception of the VIX, the model returns the probability that a specific trade will be profitable; that is, profitable trades are assigned a class label of 1, and unprofitable trades are assigned a class label of 0. Table 4.26 reports the model statistics for the in-sample period.

The intercepts for both DTVHR and CTVHR models are positive and statistically significant at the 99% confidence level. The coefficient for pair volatility is again positive and statistically significant for both models, though at the 99% confidence level for DTVHR pairs, but only at the 90% confidence level for CTVHR pairs. The magnitude of the pair-specific volatility coefficient is greater for the DTVHR model than for the CTVHR model, indicating that pair volatility plays a greater role in determining trade profitability for the DTVHR model. The coefficients for the VIX variable are negative for both DTVHR and CTVHR models, suggesting that market-wide volatility has a detrimental impact on trade profitability. Neither coefficient for VIX is statistically significant, however, so it is not possible to draw conclusions about the influence of market volatility on trade profitability. Rather, it appears that pair-specific volatility has the greatest impact on trade profitability for both models.

The minimum probabilities predicted for both DTVHR and CTVHR models are above 50%. If a classification threshold of 50% were used, every trade for both models would be predicted to be profitable, with all trades being pursued without intervention from the logistic regression model. The maximum probabilities of trade profitability, on the other hand, range from 70.05% for the DTVHR model, down to only 59.84% for the CTVHR model. An appropriate classification threshold between the minimum and maximum

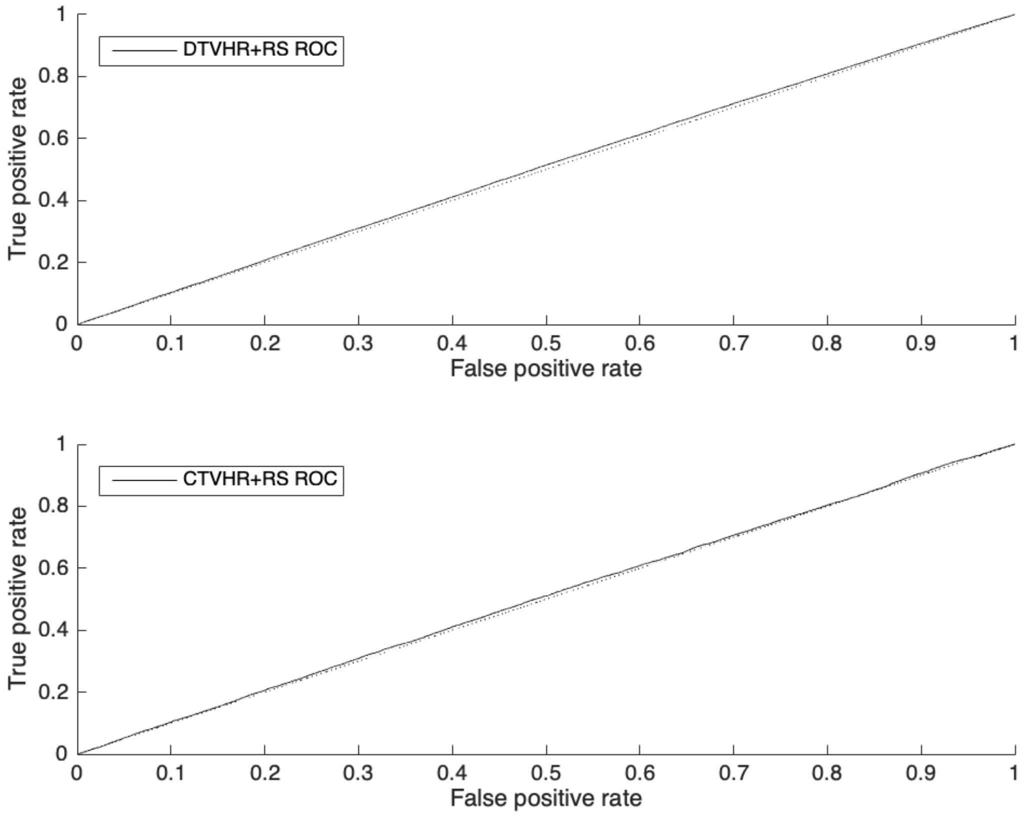

Figure 4.19: In-sample Receiver Operating Characteristics curve for regime switching model of trade outcomes, January 1990–July 2009.

predicted probabilities must therefore be chosen to ensure trades do not proceed without oversight from the regime switching model. On average, the logistic regression model predicts greater probability of trade profitability for the DTVHR model than the CTVHR model, though all trades for both models are assigned probabilities greater than 50%, indicating the relatively poor statistical performance of the logistic regression model. This is further demonstrated by the low Area Under Receiver Operating Characteristics (AUROC) for both models, measuring the tradeoff between sensitivity and specificity for binary classification models—the closer to 1.0 the better, and the closer to 0.5 the poorer. AUROC measures the area beneath the Receiver Operating Characteristics (ROC) curve, itself quantifying the comparative rate of true and false positives of a binary classifier as its classification threshold varies, delivering a useful statistic for the purpose of model comparison. Figure 4.19 displays the ROC curve for both DTVHR and CTVHR regime switching models.

Table 4.27 reports the binary classification statistics for both DTVHR and CTVHR regime switching models. In addition to reporting statistics under a 50% classification threshold, statistics are also reported for classification thresholds which are optimised according to Youden's Index, which is equal to the sum of model sensitivity and specificity minus one. Youden's Index therefore represents the probability of an informed decision, which is maximised relative to the model's true positive and true negative rates. The analysis conducted in Section 4.2 adopts the classification threshold determined by the optimisation of Youden's Index.

The optimal classification thresholds for both models are approximately in the middle of their respective predicted probability ranges, with the optimal DTVHR threshold realised at a value of 0.59 and the optimal CTVHR threshold realised at a value of 0.55. The proportion exceeding this threshold declines to 65.72% for the DTVHR model and 30.64% for the CTVHR model relative to a classification threshold of 50%, under which 100% of both models' trades are predicted to be profitable. The proportion of positive and negative outcomes remains unchanged between classification thresholds, though the DTVHR model has a higher proportion of profitable trades than the CTVHR model. Under the optimal classification threshold, both models experience a reduction in the proportion of true positives and an increase in the proportion of true negatives.

False positives are more detrimental to the arbitrageur than false negatives, as false positives commit the arbitrageur to an unprofitable trade, tying up capital and contributing to portfolio losses. False negatives, on the other hand, only cost the arbitrageur a profitable opportunity, but neither commit capital nor contribute to portfolio losses. False positives for the DTVHR model decline from 40.27% under a 50% classification threshold to 25.69% under Youden's optimal threshold. False positives decline even further for the CTVHR model, from 45.52% under a 50% classification threshold to 13.38% under Youden's optimal threshold. An undesirable consequence of using the optimised classification threshold is an increase in the proportion of false negatives, from 0.00% for both models under a 50% classification threshold to 19.70% for the DTVHR model and 37.22% under the CTVHR model. Sensitivity and specificity, both used in the calculation of Youden's Index, respectively decline and increase between the 50% and optimal classification thresholds. The optimal Youden's Index values of 0.0322 for the DTVHR model and 0.0227 for the CTVHR model reflect the relatively poor predictability of the logistic regression model, with a Youden's Index value of 1 representing perfect classification.

The arbitrageur can use the predictions made by the logistic regression model in one of two ways: decline to place a trade when its predicted probability of profitability is below the classification threshold, or invert the trading signal and place the opposite trade. If $\iota \in \{-1, 0, 1\}$ is the original trading signal generated by the TVHR model, θ_J is the classification threshold, and $Pr\left(R \geq 0\right)$ is the model-predicted probability of trade

	50% classification threshold		Youden's Index optimised	
	DTVHR	CTVHR	DTVHR	CTVHR
Threshold	0.5000	0.5000	0.5900	0.5500
Proportion exceeding threshold	1.0000	1.0000	0.6572	0.3064
P	0.5973	0.5448	0.5973	0.5448
N	0.4027	0.4552	0.4027	0.4552
TP	0.5973	0.5448	0.4002	0.1725
TN	0.0000	0.0000	0.1458	0.3214
FP (Type I Error)	0.4027	0.4552	0.2569	0.1338
FN (Type II Error)	0.0000	0.0000	0.1970	0.3722
Sensitivity	1.0000	1.0000	0.6701	0.3167
Specificity	0.0000	0.0000	0.3621	0.7060
Youden's Index	0.0000	0.0000	0.0322	0.0227

Table 4.27: Binary classification statistics of trade outcomes under 50% classification threshold and Youden's Index optimised threshold. Reported statistics include the proportion of positive (P), negative (N), true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes. Sensitivity (TP/P), specificity (TN/N), and Youden's Index (TP/P + TN/N - 1) are additionally reported.

profitability, then

$$\iota^* = \begin{cases} \iota, & Pr(R \geq 0) \geq \theta_J \\ 0, & Pr(R \geq 0) < \theta_J \end{cases}$$

where ι^* is the augmented trading rule inferred from the forecast probability of trade profitability. This specification is consistent with the former way the arbitrageur may use the prediction of the logistic regression model, in which trades that are not predicted to be profitable are ignored. For clarity, trades using the regime switching model in this trade-negating way are given the suffix RSN. The alternative use of the regime switching model,

$$\iota^* = \begin{cases} \iota, & Pr(R \geq 0) \geq \theta_J \\ -\iota, & Pr(R \geq 0) < \theta_J \end{cases}$$

in which trades that are not predicted to be profitable are inverted, are given the suffix RSI. If standard statistical arbitrage exploits mean-reverting phenomena, then the inversion of the trading rule can be considered an attempt to exploit mean-averting or momentum dynamics. Table 4.28 reports the in-sample excess return statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal, with and without the regime switching model extensions.

Relative to the standard TVHR model, the RSN regime switching model improves the mean monthly return of the CTVHR model from 0.37% to 0.51%, but slightly diminishes

Model variant	TVHR		TVHR+RSN		TVHR+RSI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Mean	0.0022	0.0037	0.0020	0.0051	0.0016	0.0001
t-Statistic	4.9911	2.7775	4.9687	2.4629	4.4256	0.1326
Median	0.0020	0.0011	0.0015	0.0000	0.0011	0.0000
Standard deviation	0.0051	0.0215	0.0049	0.0284	0.0047	0.0181
Skewness	1.3300	1.0278	1.7209	4.2537	1.6462	0.0316
Kurtosis	8.6759	11.6635	10.2678	42.3901	9.9546	14.6886
Minimum	-0.0098	-0.0978	-0.0099	-0.1431	-0.0111	-0.1081
Maximum	0.0304	0.1113	0.0289	0.2704	0.0274	0.1009
Observations < 0	0.3034	0.3120	0.3162	0.2479	0.3547	0.4402
Lower semi-deviation	0.0021	0.0120	0.0019	0.0115	0.0020	0.0127
Upper semi-deviation	0.0052	0.0182	0.0049	0.0264	0.0045	0.0129
Sharpe ratio	1.4637	0.5930	1.3776	0.6197	1.1482	0.0250
Sortino ratio	3.5435	1.0654	3.5101	1.5247	2.7280	0.0357
TIM	0.2266	0.0279	0.1828	0.0091	0.2255	0.0279
Return/TIM	0.0096	0.1315	0.0107	0.5569	0.0069	0.0047

Table 4.28: Excess return statistics for portfolios of top 20 unrestricted pairs, execution delay, January 1990–June 2009. The first two columns report the statistics of the standard model with no regime switching extension, the middle two columns report the statistics of the regime switching extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the regime switching extension which inverts the position of trades that are not predicted to be profitable.

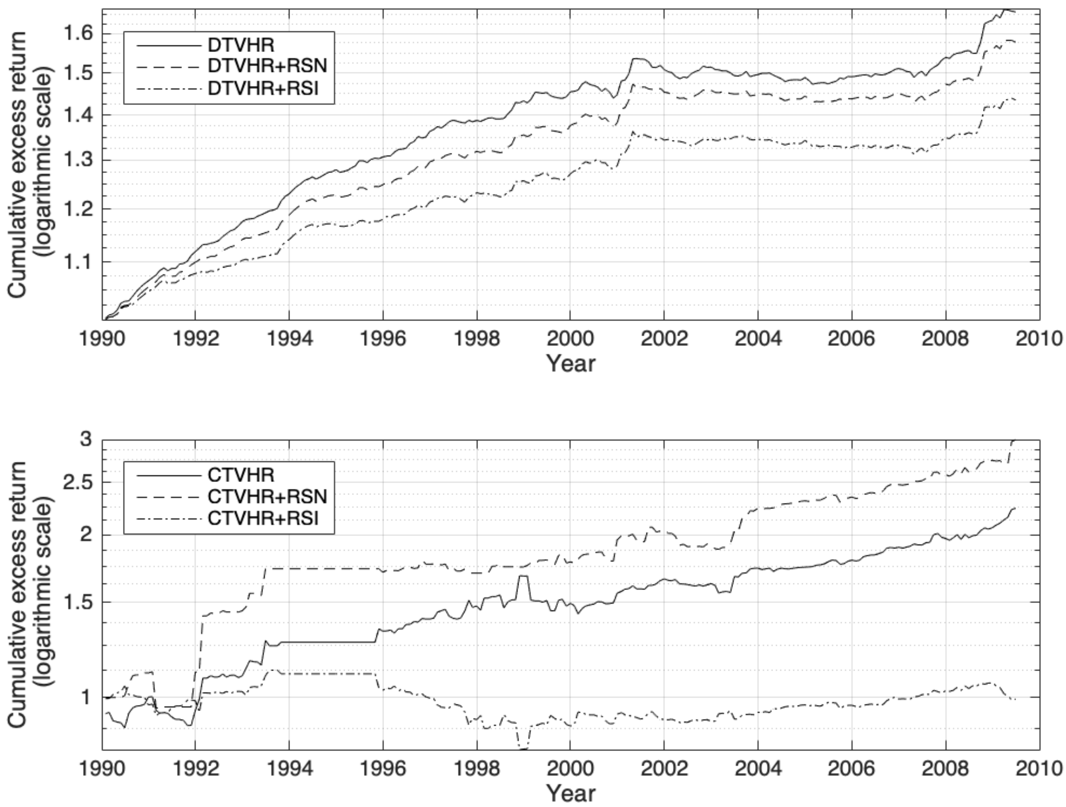

Figure 4.20: Cumulative excess return of top 20 unrestricted pairs with regime switching-inferred negation or inversion of trade position, January 1990–June 2009.

the mean monthly return of the DVTHR model. The RSI variant of the regime switching model, by contrast, diminishes the mean monthly return of both the DTVHR and CTVHR models, with the CTVHR mean monthly return declining to 0.01%. Figure 4.20 displays the cumulative excess portfolio returns for DTVHR and CTVHR models with and without regime switching extensions.

The median monthly return of the DTVHR model declines from 0.20% to 0.15% under the RSN variant, and further declines to 0.11% under the RSI variant. The median monthly return of the CTVHR model declines from 0.11% to 0.00% under both regime switching variants. Both regime switching variants diminish the standard deviation of returns for the DTVHR model while increasing the skewness, kurtosis, and proportion of negative observations. The CTVHR model reports a greater standard deviation under the RSN variant due to the limited number of trades executed, and a lower standard deviation for the RSI variant relative to the standard model. The RSN variant increases the CTVHR model's skewness, kurtosis, and magnitude of minimum and maximum returns while diminishing the proportion of negative returns. The RSI variant diminishes the standard deviation, skewness, and maximum monthly return while increasing the kurtosis, magnitude of minimum return, and proportion of negative returns.

Only the CTVHR model reports improved Sharpe and Sortino ratios under the RSN regime switching variant while all other models report a decline. In particular, the Sortino ratio declines from 1.07 under the CTVHR model to 0.04 under the CTVHR+RSI model. The RSN variant reduces the TIM for both the DTVHR and CTVHR models, while the RSI variant has little impact on TIM. Consequently, the RSN variant delivers the highest return per unit TIM of 1.07% for the DTVHR model and 55.69% for the CTVHR model. By contrast, the lowest return per unit TIM for both models is reported under the RSI variant.

The in-sample economic performance of the regime switching model reveals the detrimental impact of inverting trade signals under the RSI variant. With the exception of standard deviation, skewness, kurtosis, lower semi-deviation and TIM, all portfolio statistics for the DTVHR model are adversely impacted by the deleterious effects of trade signal inversion. The CTVHR model experiences an even more visible decline in performance, with only standard deviation and kurtosis improving under the RSI variant. The RSN variant improves the mean portfolio return of the CTVHR model and the return per unit TIM for both the DTVHR and CTVHR models, but does so at the expense of median monthly return.

The out-of-sample economic performance of the regime switching model is reported in Table 4.29. As with the in-sample period, the only improvement contributed by the regime switching model is reported by the CTVHR model under the RSN variant. Mean monthly return improves from 0.56% to 0.66%, though median return declines from 0.26% to 0.23%. Mean monthly return under all other model variants declines under consideration of the regime switching model. The greatest decline in performance is delivered by the CTVHR+RSI model, with mean return declining from 0.56% to 0.20%, and median monthly return declining from 0.26% to −0.04%. Figure 4.21 displays the cumulative excess portfolio returns for DTVHR and CTVHR models with and without regime switching extensions during the out-of-sample period.

Return standard deviation improves under the RSN variant for the DTVHR model, and both the DTVHR and CTVHR models under the RSI variant. Skewness and kurtosis improve for the CTVHR+RSI model but decline for all other models. Minimum and maximum return for regime switching variants underperform their standard TVHR analogues with the exception of the CTVHR+RSN maximum return. Proportion of negative monthly returns declines for the DTVHR model under both regime switching variants, but increases for the CTVHR model under both variants.

Model variant	TVHR		TVHR+RSN		TVHR+RSI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Mean	0.0010	0.0056	0.0008	0.0066	0.0006	0.0020
t-Statistic	2.3458	3.4205	2.2032	2.9922	1.6590	1.0841
Median	0.0012	0.0026	0.0012	0.0023	0.0008	-0.0004
Standard deviation	0.0036	0.0200	0.0034	0.0252	0.0033	0.0188
Skewness	0.4847	4.7455	0.0953	4.1810	0.0419	5.7358
Kurtosis	4.5150	35.8846	3.9744	25.8963	3.5620	48.3148
Minimum	-0.0073	-0.0335	-0.0079	-0.0406	-0.0075	-0.0352
Maximum	0.0155	0.1610	0.0121	0.1755	0.0100	0.1603
Observations < 0	0.3889	0.3241	0.3704	0.3426	0.3611	0.5278
Lower semi-deviation	0.0019	0.0055	0.0020	0.0067	0.0020	0.0065
Upper semi-deviation	0.0032	0.0199	0.0029	0.0250	0.0026	0.0177
Sharpe ratio	0.9877	0.9766	0.8480	0.9146	0.6201	0.3613
Sortino ratio	1.8833	3.5398	1.4476	3.4133	0.9993	1.0504
TIM	0.2634	0.0682	0.2211	0.0248	0.2634	0.0682
Return/TIM	0.0039	0.0826	0.0038	0.2685	0.0022	0.0287

Table 4.29: Excess return statistics for portfolios of top 20 unrestricted pairs, execution delay, July 2009–June 2018. The first two columns report the statistics of the standard model with no regime switching extension, the middle two columns report the statistics of the regime switching extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the regime switching extension which inverts the position of trades that are not predicted to be profitable.

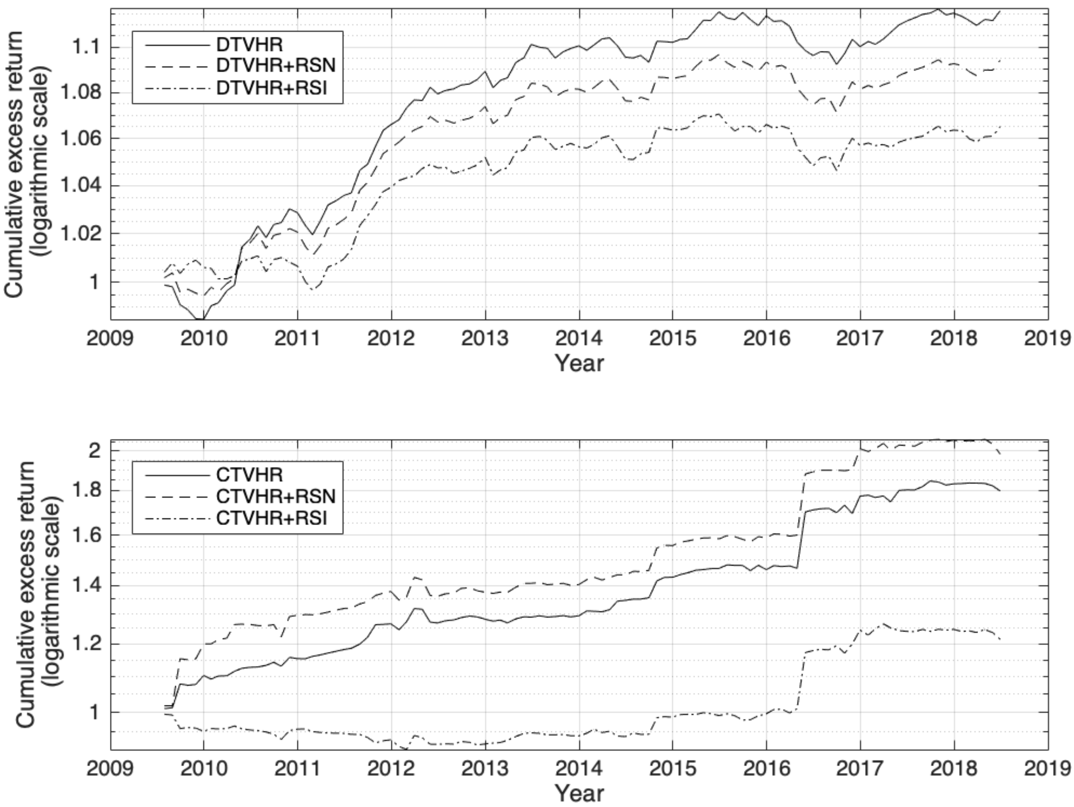

Figure 4.21: Cumulative excess return of top 20 unrestricted pairs with regime switching-inferred negation or inversion of trade position, July 2009–June 2018.

Model variant	TVHR		TVHR+RSN		TVHR+RSI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Group 1 proportion	0.0123	0.0442	0.0570	0.3841	0.0123	0.0442
Group 2 proportion	0.0187	0.0015	0.0211	0.0064	0.0167	0.0015
Profitable proportion	0.2368	0.6667	0.2326	0.5385	0.2353	0.3333
Total return	-0.0282	0.0018	-0.0249	-0.0073	-0.0283	-0.0018
Number of trades	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TIM	0.2932	0.0162	0.2620	0.0141	0.3123	0.0162
Industry-matched	0.2105	0.0000	0.1860	0.0000	0.1765	0.0000
Distance zero-crossings	21.9211	21.3333	24.5116	14.3846	21.4412	21.3333
Cointegration zero-crossings	17.4211	69.3333	20.4651	65.1538	16.7941	69.3333
BMD	0.8072	697.3834	0.8207	3.3083	0.7517	697.3834
Group 3 proportion	0.9691	0.9543	0.9219	0.6095	0.9710	0.9543
Profitable proportion	0.5976	0.5872	0.6052	0.6052	0.5604	0.5142
Total return	0.0072	0.0332	0.0057	0.0386	0.0037	0.0159
Number of trades	3.8794	3.5270	3.0970	1.9952	4.2210	3.5440
TIM	0.2656	0.0717	0.2346	0.0412	0.2653	0.0717
Industry-matched	0.2767	0.1462	0.2824	0.1507	0.2772	0.1462
Distance zero-crossings	31.1272	25.4169	30.8871	25.7486	31.1168	25.4169
Cointegration zero-crossings	33.2894	66.2934	32.9382	65.7260	33.2681	66.2934
BMD	1.3142	4.6105	1.2584	5.2716	1.3331	4.6105

Table 4.30: Pair statistics for top 20 unrestricted pairs, execution delay, July 2009–June 2018. The first two columns report the statistics of the standard model with no regime switching extension, the middle two columns report the statistics of the regime switching extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the regime switching extension which inverts the position of trades that are not predicted to be profitable.

Sharpe and Sortino ratio decline for both models under both regime switching variants with the greatest decline observed under the CTVHR+RSI model, with Sharpe ratio declining from 0.98 to 0.36, and Sortino ratio declining from 3.54 to 1.05. Despite the poor performance of the regime switching model relative to other statistics, TIM declines under the CTVHR+RSN model from 6.82% to 2.48%, with return per unit TIM increasing from 8.26% to 26.85%. The same improvement in return per unit TIM was not observed for the DTVHR model, unlike during the in-sample period, with return per unit TIM declining slightly from 0.39% to 0.38% under the DTVHR+RSN model. This slight reduction is due to the decline in mean monthly return.

Table 4.30 reports the pair statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal during the out-of-sample period. Group 1 proportions increase substantially for both models under the RSN variant

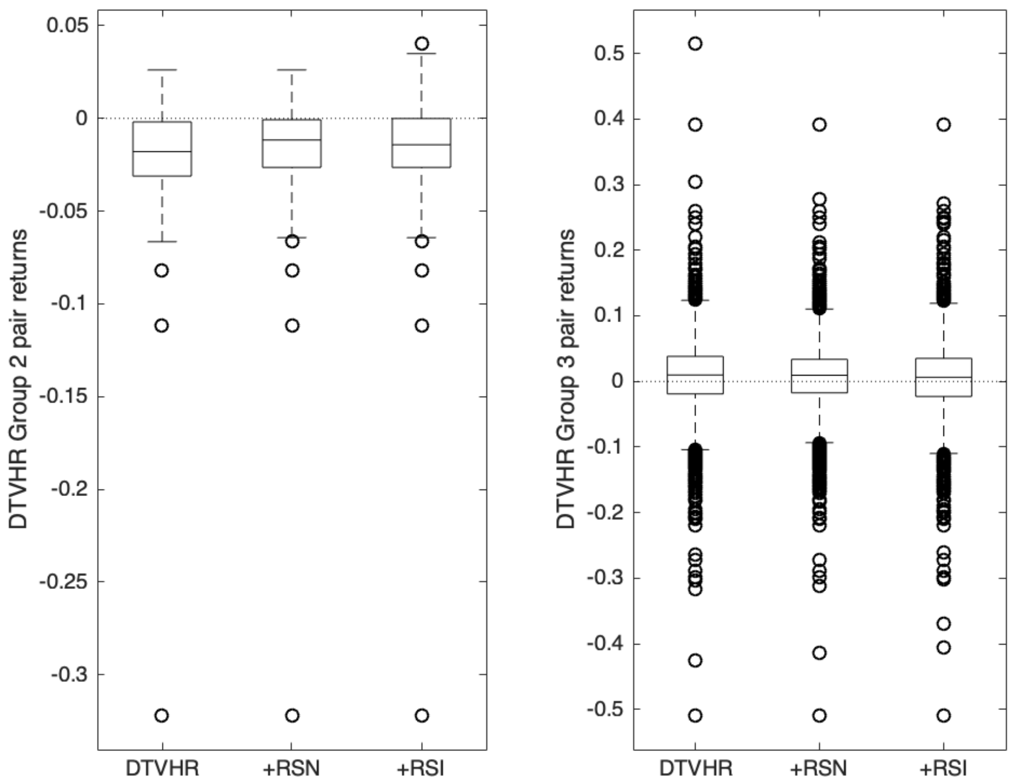

Figure 4.22: Distribution of DTVHR pair excess return of top 20 unrestricted pairs with regime switching-inferred negation or inversion of trade position, July 2009–June 2018.

but remain identical under the RSI variant. The DTVHR Group 1 proportion increases from 1.23% to 5.70%, and the CTVHR Group 1 proportion increases from 4.42% to 38.41% under the RSN variant. The substantial increase in non-trading Group 1 proportions explains the reduction in TIM and consequent increase in return per unit TIM observed by both models and the CTVHR model in particular. Figure 4.22 displays the distribution of Group 2 and Group 3 pair returns for the DTVHR model, and Figure 4.23 displays the distribution of Group 2 and Group 3 pair returns for the CTVHR model under both regime switching variants.

Group 2 proportions increase for both models under the RSN variant, and decline slightly for the DTVHR model under the RSI variant. There is a slight discrepancy in the proportion of Group 2 pairs between TVHR and TVHR+SLI variants, despite the latter trading the exact same pairs as the TVHR model albeit with variations in the position taken. This discrepancy is attributable to the embargo of portfolio return data that overlaps in-sample and out-of-sample periods so as to eliminate forward-looking bias. Profitable proportion of Group 2 pairs for both models declines for both regime

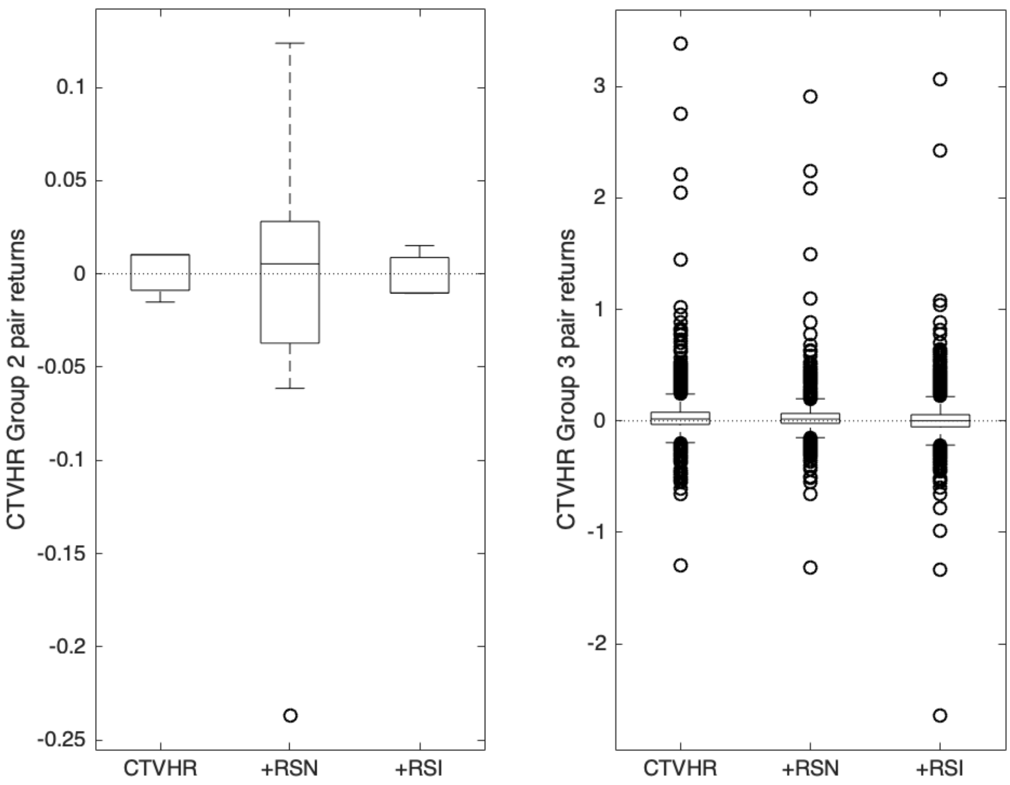

Figure 4.23: Distribution of CTVHR pair excess return of top 20 unrestricted pairs with regime switching-inferred negation or inversion of trade position, July 2009–June 2018.

switching variants, with the CTVHR model profitable proportion declining from 66.67% to 33.33% under the RSI variant. Total return for Group 2 pairs improves slightly for the DTVHR+RSN model but deteriorates for all other model specifications. TIM decreases for both models under the RSN variant but increases for the DTVHR model under the RSI variant. There is a slight reduction in the proportion of industry-matched Group 2 DTVHR pairs under both regime switching variants, while the number of distance and cointegration zero-crossings remain relatively consistent between the standard TVHR model and its RSI extension. The RSN extension, by contrast, reports an increase in the number of distance zero-crossings for the DTVHR model and a decline in the number of cointegration zero-crossings for the CTVHR model. BMD fluctuates for both models under the different regime switching variants, though the small sample sizes of Group 2 pairs makes it difficult to draw conclusions from these findings.

Group 3 proportions are more heavily influenced by the increase in Group 1 proportions than Group 2 proportions, with the CTVHR Group 3 proportion declining from 95.43% to 60.95% under the RSN variant thanks to its substantial Group 1 proportion of

38.41%. Profitable proportion for both models increases slightly under the RSN variant and declines under the RSI variant. Total return of Group 3 pairs for the DTVHR model declines under both regime switching variants, and increases slightly for CTVHR pairs under the RSN variant. Total return of the CTVHR+RSI model is less than half that of the CTVHR model. Number of trades and TIM are almost halved under the CTVHR+RSN model, contributing to the model's relatively high return per unit TIM. Unlike the results reported in Section 4.1, and with the exception of the CTVHR and CTVHR+RSI models, BMD is higher for Group 3 pairs than Group 2 pairs. The TIM of Group 2 CTVHR+RSN pairs is relatively short, while TIM of CTVHR+RSN Group 3 pairs is more than double that of Group 2 pairs. This indicates that the cumulative BMD for Group 3 pairs is greater than that of Group 2 pairs because Group 2 pairs do not have sufficient time to converge at the end of the trading period. By contrast, the lower BMD of DTVHR Group 2 pairs relative to Group 3 pairs is attributable to the slower reversion rate estimated for Group 2 pairs. Because the reversion rate is estimated from the formation period zero-crossings, the proportionally lower number of distance zero-crossings for Group 2 pairs leads to a slower reversion rate and, consequently, lower BMD than Group 3 pairs.

Table 4.31 reports the individual trade statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal during the out-of-sample period. Profitable proportion of DTVHR trades increases marginally for the RSN variant and declines for the RSI variant. Similarly, profitable proportion for the CTVHR model increases from 55.35% to 58.39% under the RSN variant and declines to 51.01% under the RSI variant. Mean return halves for both models under the RSI variant, but doubles for CTVHR trades under the RSN variant. This increase in mean return is accompanied by an increase in return standard deviation for the CTVHR+RSN model with all other models delivering relatively consistent standard deviations. Sharpe ratio increases by ∼60% for the CTVHR model under the RSN variant, driven by the doubling of mean return which offsets the increase in standard deviation. Consistent with their halved mean returns, both models under the RSI variant deliver Sharpe ratios that are approximately half their standard TVHR analogues. Mean profit improves slightly for both models under the RSN variant but remains relatively stable under the RSI variant. Mean loss is approximately 1% greater in magnitude under the CTVHR+RSN model with other models remaining relatively consistent. The improvement in mean return under the CTVHR+RSN model is driven by a substantial increase in mean short return despite mean long return declining by half. Conversely, the decline in mean return for the CTVHR+RSI model is driven by a substantial reduction in both mean long return and mean short return of approximately 0.22% each. Figure 4.24 displays the distribution of convergent and non-convergent trade returns for the DTVHR model, and Figure 4.25

Model variant	TVHR		TVHR+RSN		TVHR+RSI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
All trades						
Profitable proportion	0.6143	0.5535	0.6164	0.5839	0.5644	0.5101
Mean return	0.0019	0.0086	0.0018	0.0173	0.0008	0.0043
Standard deviation	0.0271	0.0976	0.0282	0.1242	0.0261	0.0933
Sharpe ratio	0.3738	0.8773	0.3204	1.3862	0.1734	0.4561
Mean profit	0.0161	0.0496	0.0170	0.0669	0.0158	0.0495
Mean loss	-0.0207	-0.0422	-0.0227	-0.0522	-0.0186	-0.0429
Mean long return	0.0065	0.0019	0.0070	0.0010	0.0055	-0.0002
Mean short return	-0.0046	0.0067	-0.0053	0.0163	-0.0047	0.0045
Mean trade length	8.7649	2.5347	9.6494	2.5535	8.0593	2.5224
Median trade length	6.0000	2.0000	7.0000	2.0000	6.0000	2.0000
Convergent trades	0.9152	0.9776	0.9098	0.9765	0.9220	0.9777
Profitable proportion	0.6362	0.5546	0.6423	0.5874	0.5801	0.5110
Mean return	0.0030	0.0090	0.0031	0.0180	0.0017	0.0044
Standard deviation	0.0268	0.0984	0.0279	0.1253	0.0258	0.0940
Sharpe ratio	0.6099	0.9072	0.5669	1.4287	0.3687	0.4663
Mean profit	0.0165	0.0500	0.0174	0.0675	0.0161	0.0499
Mean loss	-0.0205	-0.0422	-0.0227	-0.0524	-0.0183	-0.0431
Mean long return	0.0073	0.0021	0.0080	0.0013	0.0062	-0.0002
Mean short return	-0.0043	0.0069	-0.0050	0.0168	-0.0045	0.0046
Mean trade length	8.6719	2.5423	9.5498	2.5573	7.9439	2.5291
Median trade length	6.0000	2.0000	7.0000	2.0000	6.0000	2.0000
Non-convergent trades	0.0848	0.0224	0.0902	0.0235	0.0780	0.0223
Profitable proportion	0.3775	0.5064	0.3549	0.4426	0.3790	0.4744
Mean return	-0.0105	-0.0079	-0.0114	-0.0119	-0.0096	-0.0017
Standard deviation	0.0272	0.0512	0.0282	0.0555	0.0268	0.0515
Sharpe ratio	-1.9562	-1.6489	-1.9766	-2.2017	-1.8594	-0.3451
Mean profit	0.0092	0.0263	0.0095	0.0289	0.0095	0.0344
Mean loss	-0.0224	-0.0430	-0.0230	-0.0443	-0.0213	-0.0342
Mean long return	-0.0026	-0.0042	-0.0031	-0.0086	-0.0027	-0.0009
Mean short return	-0.0079	-0.0036	-0.0084	-0.0033	-0.0070	-0.0008
Mean trade length	9.7686	2.2051	10.6543	2.3934	9.4247	2.2308
Median trade length	6.0000	2.0000	7.0000	2.0000	6.0000	2.0000

Table 4.31: Trade statistics for top 20 unrestricted pairs, execution delay, July 2009–June 2018. The first two columns report the statistics of the standard model with no regime switching extension, the middle two columns report the statistics of the regime switching extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the regime switching extension which inverts the position of trades that are not predicted to be profitable.

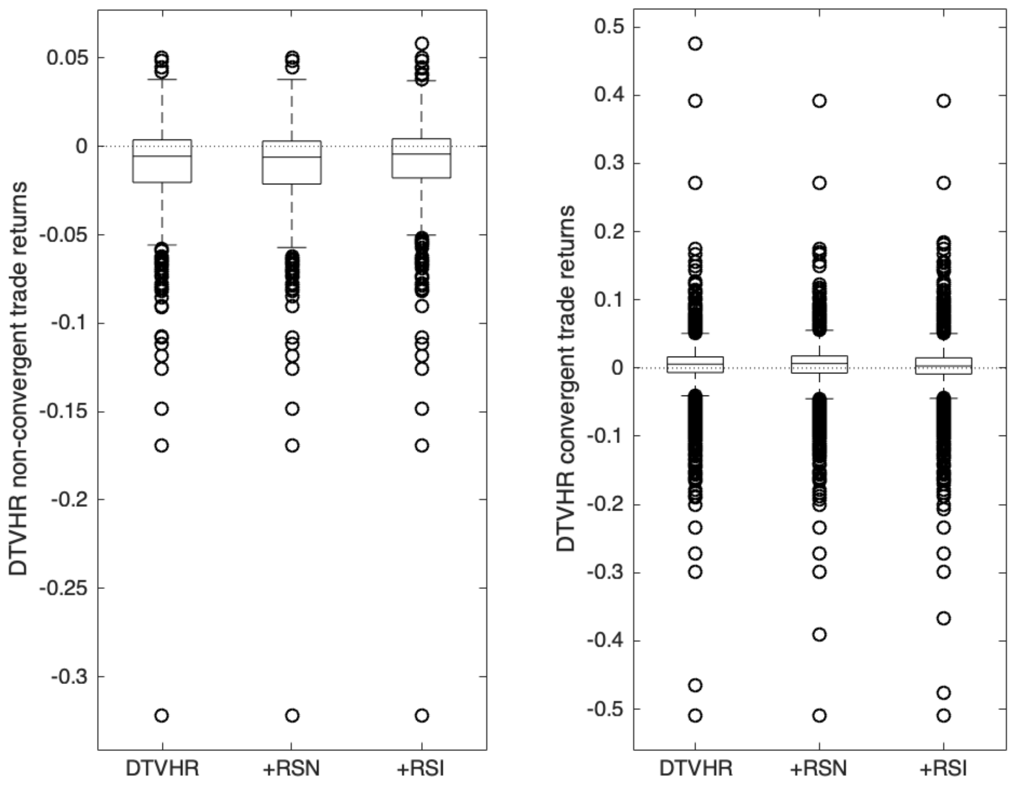

Figure 4.24: Distribution of DTVHR trade excess return of top 20 unrestricted pairs with regime switching-inferred negation or inversion of trade position, July 2009–June 2018.

displays the distribution of convergent and non-convergent trade returns for the CTVHR model under both regime switching variants.

The proportion of convergent trades declines for both models under the RSN variant but increases marginally under the RSI variant. Conversely, the profitable proportion increases under the RSN variant but declines under the RSI variant. The high proportion of convergent trades ensures convergent trade statistics closely resemble those for all trades, with mean return for the CTVHR+RSN model delivering the greatest improvement, and mean return for the CTVHR+RSI model declining the most. Among non-convergent trades, profitable proportion declines for both models under the RSN variant but for only the CTVHR model under the RSI variant. The magnitude of mean return increases for both models under the RSN variant and declines for both models under the RSI variant. The most significant decline in magnitude of mean return is delivered by the CTVHR+RSI model, declining from -0.79% to -0.17%. Standard deviations and Sharpe ratios are relatively consistent for all models with the exception of the CTVHR+RSI model, whose decline in mean return causes a substantial decline in the magnitude of its

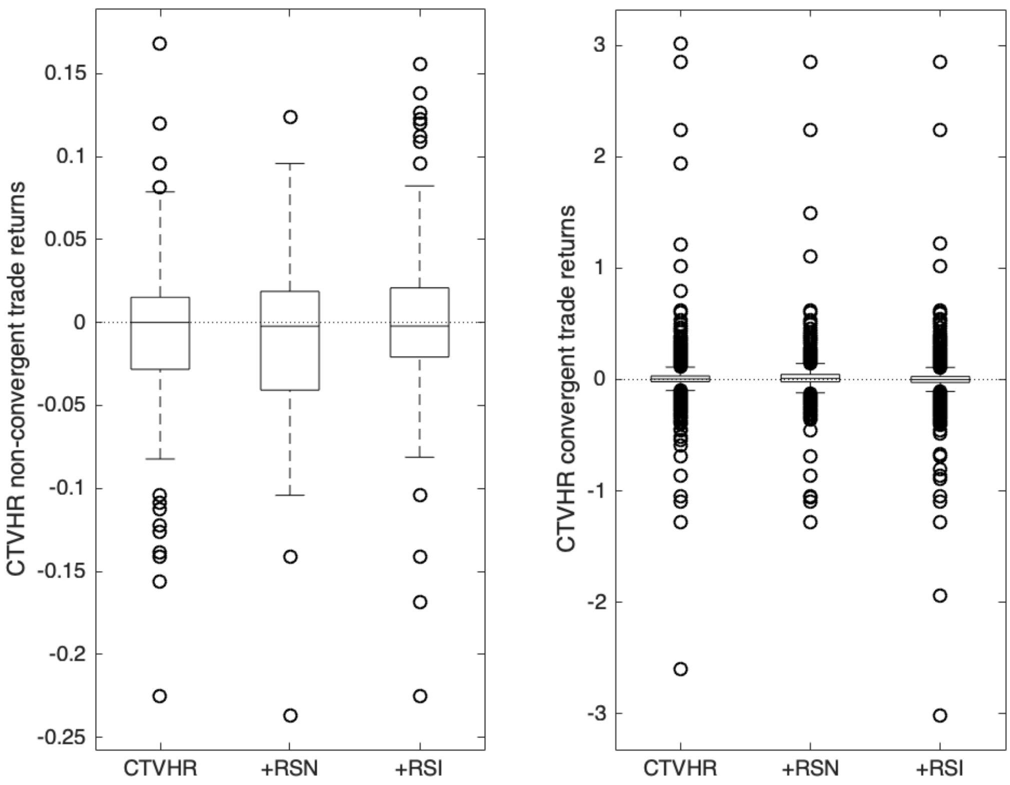

Figure 4.25: Distribution of CTVHR trade excess return of top 20 unrestricted pairs with regime switching-inferred negation or inversion of trade position, July 2009–June 2018.

Sharpe ratio. Mean profit increases in magnitude and mean loss declines in magnitude for the CTVHR+RSI model with all other models remaining relatively consistent.

4.2.1 Summary of Regime Switching Model Extension

Section 2.3 posed the research question, *are statistical arbitrage returns dependent on the prevailing volatility regime?* The objective of the regime switching model extension is to reduce the proportion and magnitude of losses in the case of the RSN variant, and to improve aggregate portfolio returns in the case of the RSI variant by exploiting momentum regimes, both objectives informed by the prevailing level of volatility at the time of trade initiation. The results reported in Section 4.2 indicate an inability of the model extension to deliver its objectives under either of its variants in terms of aggregate portfolio returns.

In-sample statistics report an improvement in mean return for the CTVHR+RSN model, but a decline for all other model variants. The CTVHR+RSN model fails to deliver a commensurate improvement in portfolio median return, however, instead delivering a median monthly return of 0.00%. Both the DTVHR and CTVHR model under the RSN variant achieved a greater level of in-sample efficiency, as indicated by the return per unit TIM, though only the CTVHR+RSN model was able to realise increased efficiency in the out-of-sample period. That improved efficiency was attributable to a substantially greater proportion of Group 1 pairs, causing an increased proportion of capital to be allocated to traded pairs under the employed capital allocation scheme. Overall, the regime switching logistic regression model investigated in Section 4.2 failed to deliver excess statistical or economic performance at the portfolio level over the sample period July 2009–June 2018.

There was a substantial improvement in per-trade performance of CTVHR pairs under the SLN variant during the out-of-sample period that was not replicated in other model variants. The improved mean return and Sharpe ratio of individual CTVHR+SLN trades came at the expense of substantially fewer trades being opened due to the negation of signals whose return was forecast to be negative. While this improved per-trade performance was not reflected at the portfolio level out-of-sample, it does indicate that volatility regimes at least partially influence returns of the CTVHR model, though not when used to consider mean-averting behaviour under the SLI variant.

Pairs selected under the cointegration specification are naturally more susceptible to volatility than their distance specification counterparts, due to their selection being contingent on the level of formation period volatility. Despite the statistically insignificant coefficient estimated for the VIX factor reported in Table 4.26, the factor relating idiosyncratic pair volatility to trade profitability was found to be statistically significant with a positive coefficient—the greater the level of pair volatility at the time of trade initiation, the greater the probability of profitability. Though it is difficult to conclusively answer the research question given these findings, there is evidence to conclude that statistical arbitrage returns are dependent on the prevailing volatility regime, though only under the cointegration specification.

4.3 Statistical Learning Model Extension

Section 4.3 considers a statistical learning extension of the TVHR model that builds on the regime switching extension discussed in Section 4.2. This section therefore seeks to augment the regime prediction framework with a more powerful statistical learning model and additional explanatory variables, addressing the third and final research question

	DTVHR	CTVHR
Input nodes	6	6
Hidden nodes	100	100
Minimum $Pr(R \geq 0)$	0.0163	0.0000
Maximum $Pr(R \geq 0)$	1.0000	1.0000
AUROC	0.5184	0.5388

Table 4.32: Extreme Learning Machine statistics for binary classification of trade outcomes based on pair-specific volatility at time of trade initiation (Pair Volatility), the level of the CBOE Volatility Index (VIX) at the time of trade initiation, logistic regression forecast of trade outcome, an indicator variable flagging pairs whose constituents come from the same industry, pair position in the top 20, and the first-order autocorrelation of the pair spread during the formation period. Minimum and maximum model-forecast probabilities of trade success are reported in addition to Area Under Receiver Operating Characteristics (AUROC) statistics.

posited in this thesis: *are statistical learning models better equipped than conventional models to capture and detect latent market regimes?* The ELM model used in this section was selected for its simplicity, global optimality, analytical solution and non-linear function mapping capability. The model's independent variables include those used in the regime switching model, namely pair-specific volatility and market-wide volatility, as well as the regime switching forecast of profitability, in addition to an indicator variable flagging pairs whose constituents come from the same industry, pair position in the top 20, and the first-order autocorrelation of pair spread during the formation period.

Table 4.32 reports model statistics for the ELM model of trade outcomes. As with the logistic regression model used in Section 4.2, the ELM model produces a forecast of the probability of trade profitability as its dependent variable. Profitable trades are given a class label of 1 while unprofitable trades are given a class label of 0, with the model estimated on an in-sample period spanning January 1990–July 2009. The statistical learning models estimated for both DTVHR and CTVHR trades use the six aforementioned independent variables and 100 hidden nodes in the hidden layer of the network. Given the computational efficiency of the ELM model's analytical solution, the high number of hidden nodes was arbitrarily selected to realise the full potential of the model and avoid overfitting concerns. Both models realise a maximum in-sample probability of profitability of 100%, while only the CTVHR ELM model realises a minimum probability of profitability of 0%. The DTVHR ELM model, by contrast, realises a minimum probability of profitability of 1.63%. Both statistical learning models achieve a higher AUROC than their regime switching counterparts, with the DTVHR model increasing from 0.51 to 0.52, and the CTVHR model increasing from 0.51 to 0.54. Figure

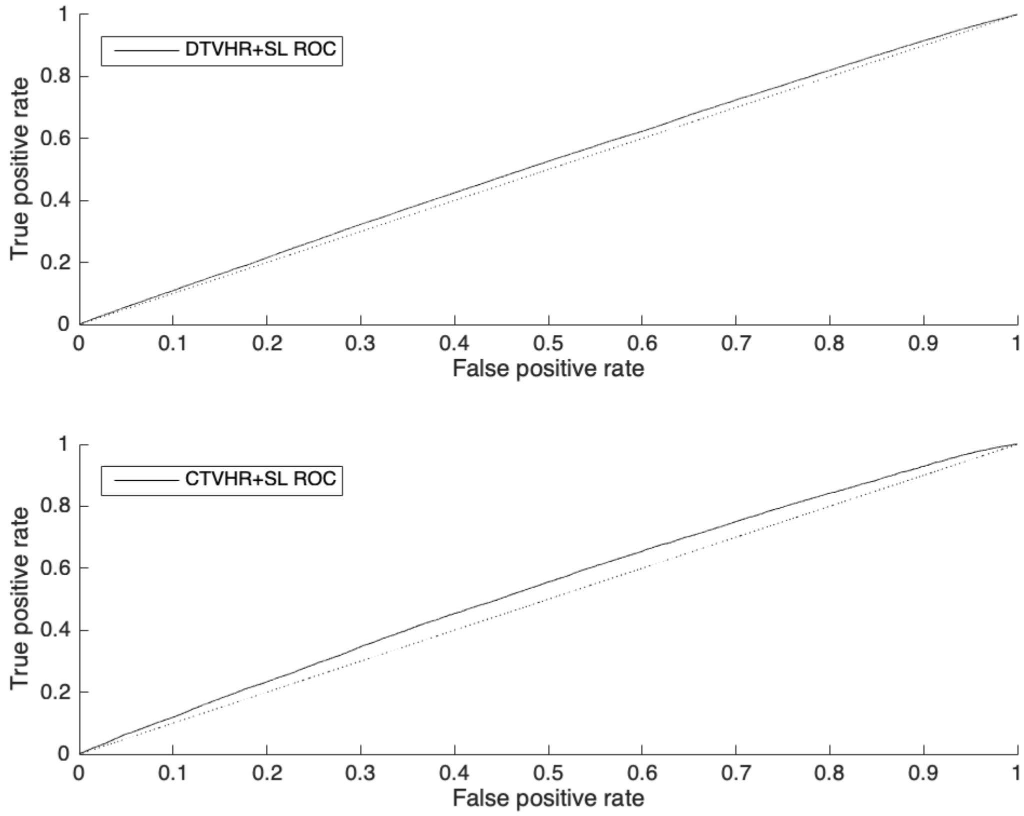

Figure 4.26: In-sample Receiver Operating Characteristics curve for statistical learning model of trade outcomes, January 1990–July2009.

4.26 displays the ROC curve for both DTVHR and CTVHR statistical learning models.

Table 4.33 reports the binary classification statistics for both DTVHR and CTVHR statistical learning models. Under a 50% classification threshold, 98.88% of all DTVHR trades would have a forecast of trade profitability, while only 78.62% of CTVHR trades would be forecast as being profitable. Under a Youden's Index optimised classification threshold of 0.59 for the DTVHR model and 0.54 under the CTVHR model, however, the predicted profitable proportion declines to 59.07% for the DTVHR model and 52.26% for the CTVHR model. This represents a substantial increase in forecast profitable trades for CTVHR pairs relative to the regime switching model, which predicts only 30.64% of trades being profitable. The statistical learning model delivers a slightly lower proportion of true positives and a slightly higher proportion of true negatives for DTVHR pairs, and a higher proportion of true positives and lower proportion of true negatives for CTVHR pairs relative to the regime switching model. False positives are slightly higher for DTVHR pairs and lower for CTVHR pairs relative to the regime switching model, while false negatives are slightly higher for DTVHR pairs but substantially lower for CTVHR pairs. Both DTVHR and CTVHR pairs report a higher Youden's Index under the statistical

	50% classification threshold		Youden's Index optimised	
	DTVHR	CTVHR	DTVHR	CTVHR
Threshold	0.5000	0.5000	0.5900	0.5400
Proportion exceeding threshold	0.9888	0.7862	0.5907	0.5226
P	0.5973	0.5448	0.5973	0.5448
N	0.4027	0.4552	0.4027	0.4552
TP	0.5888	0.4673	0.3786	0.3059
TN	0.0113	0.1082	0.1733	0.2516
FP (Type I Error)	0.3914	0.3470	0.2295	0.2036
FN (Type II Error)	0.0084	0.0774	0.2186	0.2388
Sensitivity	0.9859	0.8579	0.6339	0.5616
Specificity	0.0281	0.2377	0.4302	0.5528
Youden's Index	0.0140	0.0956	0.0641	0.1144

Table 4.33: Binary classification statistics of trade outcomes under 50% classification threshold and Youden's Index optimised threshold. Reported statistics include the proportion of positive (P), negative (N), true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes. Sensitivity (TP/P), specificity (TN/N), and Youden's Index (TP/P + TN/N - 1) are additionally reported.

learning model than their regime switching counterparts.

Table 4.28 reports the in-sample excess return statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal, with and without the statistical learning model extensions. As with the regime switching model extensions, the variant that negates trades that are predicted to be unprofitable uses the suffix SLN, while the variant that inverts trades that are predicted to be unprofitable uses the suffix SLI. For the first time in this thesis, across all tested model specifications and data sub-periods, negative mean monthly returns are realised for both the DTVHR and CTVHR model under the SLI variant. The greatest reduction in profitability is realised by the CTVHR model, declining from 0.37% to −0.16% under the SLI variant. DTVHR mean monthly returns decline more modestly from 0.22% to −0.08%. As with the regime switching model, only the CTVHR+SLN model delivers an improvement in mean monthly return, increasing from 0.37% to 0.66%. Despite this improvement in mean monthly return, the t-statistic of returns declined by half and the median monthly return declined to 0.00%—the same median monthly return as the CTVHR+SLI model which delivers a negative monthly mean return. Figure 4.27 displays the cumulative excess portfolio returns for DTVHR and CTVHR models with and without statistical learning extensions.

The only statistics that improve under the SLI variant are return standard deviation for both models, and TIM for the DTVHR model. By contrast, the SLN variant offers

Model variant	TVHR		TVHR+SLN		TVHR+SLI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Mean	0.0022	0.0037	0.0015	0.0066	-0.0008	-0.0016
t-Statistic	4.9911	2.7775	3.8941	1.3887	-2.5356	-1.7773
Median	0.0020	0.0011	0.0009	0.0000	-0.0011	0.0000
Standard deviation	0.0051	0.0215	0.0054	0.0729	0.0040	0.0172
Skewness	1.3300	1.0278	2.1496	13.7573	0.7970	-1.3662
Kurtosis	8.6759	11.6635	16.1206	204.2894	5.7266	14.9953
Minimum	-0.0098	-0.0978	-0.0190	-0.1442	-0.0111	-0.1113
Maximum	0.0304	0.1113	0.0380	1.0815	0.0187	0.0849
Observations < 0	0.3034	0.3120	0.3162	0.2265	0.6239	0.4402
Lower semi-deviation	0.0021	0.0120	0.0024	0.0138	0.0031	0.0142
Upper semi-deviation	0.0052	0.0182	0.0051	0.0717	0.0027	0.0096
Sharpe ratio	1.4637	0.5930	0.9618	0.3119	-0.6849	-0.3238
Sortino ratio	3.5435	1.0654	2.1646	1.6426	-0.8973	-0.3902
TIM	0.2266	0.0279	0.0481	0.0062	0.2255	0.0279
Return/TIM	0.0096	0.1315	0.0314	1.0627	-0.0035	-0.0575

Table 4.34: Excess return statistics for portfolios of top 20 unrestricted pairs, execution delay, January 1990–June 2009. The first two columns report the statistics of the standard model with no statistical learning extension, the middle two columns report the statistics of the statistical learning extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the statistical learning extension which inverts the position of trades that are not predicted to be profitable.

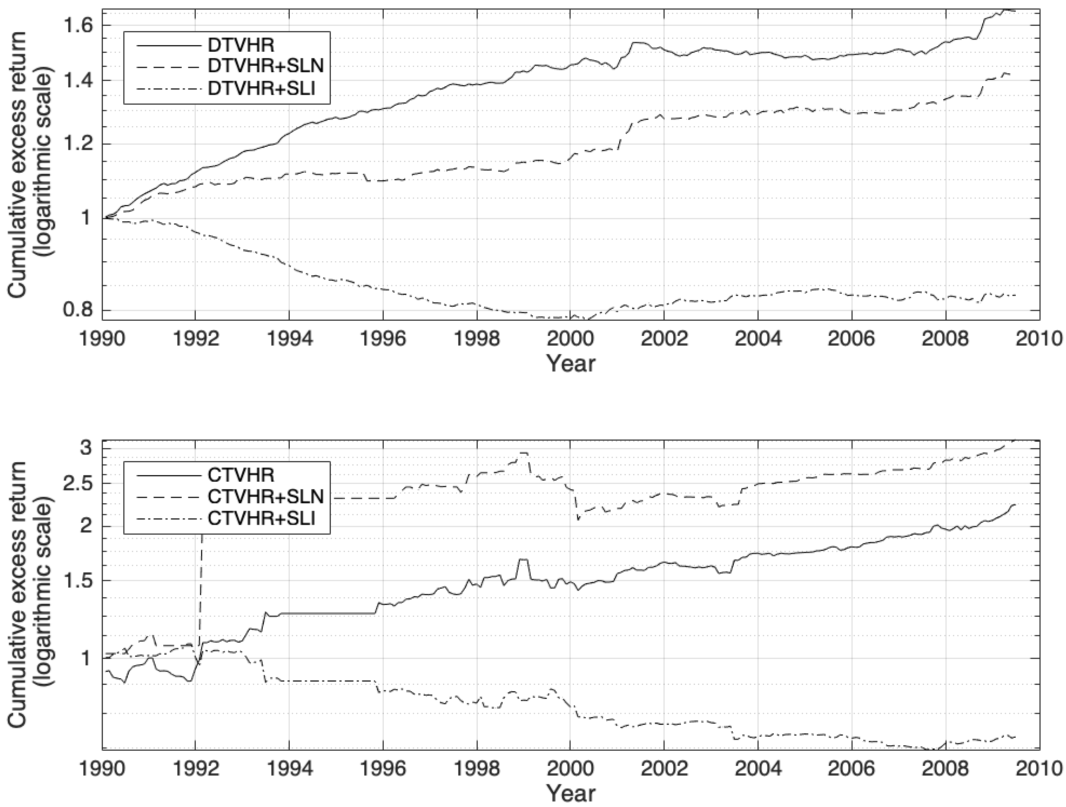

Figure 4.27: Cumulative excess return of top 20 unrestricted pairs with statistical learning-inferred negation or inversion of trade position, January 1990–June 2009.

improvements in skewness, maximum monthly return, TIM and return per unit TIM for both models, and proportion of negative monthly observations, upper semi-deviation, and Sortino ratio for the CTVHR model. In particular, the return per unit TIM for the CTVHR+SLN model of 106.27% is approximately double that of the analogous regime switching model figure of 55.69%, reflecting both an increase in monthly mean return and a decline in TIM. Also worth noting is the significant increase by one order of magnitude in maximum return under the CTVHR+SLN model, increasing from 11.13% to 108.15%. This significant increase is attributable to a large allocation under the employed capital allocation scheme due to a large proportion of non-trading pairs during the month of its observation.

Table 4.35 reports the out-of-sample excess return statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal, with and without the statistical learning model extensions. Both models under the SLI variant deliver negative mean monthly returns, with the DTVHR model declining from 0.10% to −0.01%, and the CTVHR model declining from 0.56% to −0.26%. Both models

Model variant	TVHR		TVHR+SLN		TVHR+SLI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Mean	0.0010	0.0056	0.0010	0.0029	-0.0001	-0.0026
t-Statistic	2.3458	3.4205	2.5282	2.4378	-0.3497	-2.9876
Median	0.0012	0.0026	0.0003	0.0015	-0.0006	-0.0017
Standard deviation	0.0036	0.0200	0.0042	0.0139	0.0032	0.0105
Skewness	0.4847	4.7455	1.0157	4.1882	0.7931	-1.7701
Kurtosis	4.5150	35.8846	5.7679	34.6050	4.6580	7.6928
Minimum	-0.0073	-0.0335	-0.0104	-0.0387	-0.0081	-0.0434
Maximum	0.0155	0.1610	0.0184	0.1089	0.0126	0.0192
Observations < 0	0.3889	0.3241	0.4259	0.3611	0.5648	0.6111
Lower semi-deviation	0.0019	0.0055	0.0020	0.0056	0.0021	0.0098
Upper semi-deviation	0.0032	0.0199	0.0038	0.0130	0.0024	0.0044
Sharpe ratio	0.9877	0.9766	0.8260	0.7271	-0.1450	-0.8517
Sortino ratio	1.8833	3.5398	1.7215	1.8193	-0.2200	-0.9083
TIM	0.2634	0.0682	0.0464	0.0097	0.2634	0.0682
Return/TIM	0.0039	0.0826	0.0218	0.3013	-0.0005	-0.0378

Table 4.35: Excess return statistics for portfolios of top 20 unrestricted pairs, execution delay, July 2009–June 2018. The first two columns report the statistics of the standard model with no statistical learning extension, the middle two columns report the statistics of the statistical learning extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the statistical learning extension which inverts the position of trades that are not predicted to be profitable.

also report negative median monthly returns of −0.06% and −0.17%, respectively. With the exception of skewness and kurtosis under the DTVHR+SLI model, all other statistics underperform their standard DTVHR counterparts. Figure 4.28 displays the cumulative excess portfolio returns for DTVHR and CTVHR models with and without statistical learning extensions.

In contrast to the regime switching model investigated in Section 4.2, the out-of-sample performance of the CTVHR+SLN model underperforms its CTVHR counterpart in terms of mean monthly return, declining from 0.56% to 0.29%. Similarly, the DTVHR+SLN model contrasts the out-of-sample performance of the regime switching model by delivering an equitable mean monthly return with its DTVHR counterpart. The DTVHR+SLN model also delivers a higher t-statistic than the DTVHR model, though it does so at the expense of delivering a lower median monthly return, declining from 0.12% to 0.03%.

Standard deviation of returns favours the CTVHR+SLI model over the CTVHR model, while skewness and kurtosis favour the DTVHR+SLI model over the DTVHR model. The DTVHR+SLN model delivers a higher maximum monthly return, 1.84%, than its DVTHR counterpart maximum monthly return of 1.55%, though both SLI models deliver minimum monthly returns that are greater in magnitude than their standard TVHR analogues. SLI

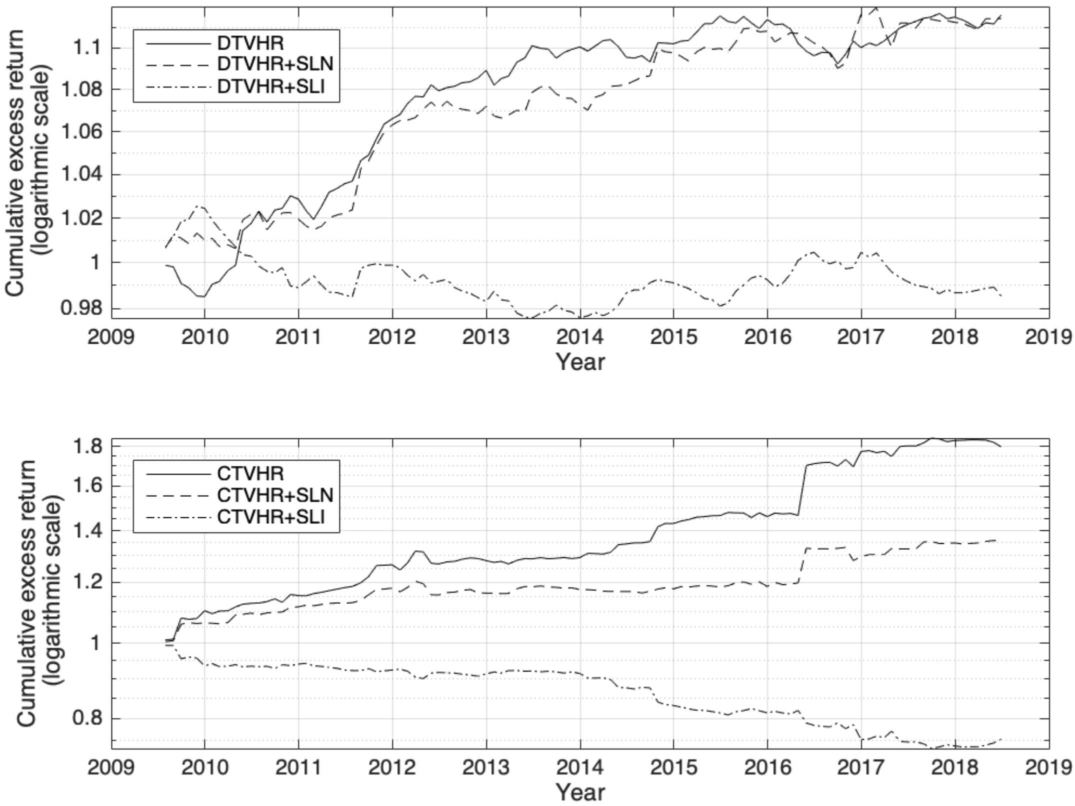

Figure 4.28: Cumulative excess return of top 20 unrestricted pairs with statistical learning-inferred negation or inversion of trade position, July 2009–June 2018.

Model variant	TVHR		TVHR+SLN		TVHR+SLI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
Group 1 proportion	0.0123	0.0442	0.6714	0.6827	0.0123	0.0442
Group 2 proportion	0.0187	0.0015	0.0290	0.0054	0.0133	0.0015
Profitable proportion	0.2368	0.6667	0.2881	0.3636	0.5556	0.3333
Total return	-0.0282	0.0018	-0.0159	-0.0073	0.0038	-0.0018
Number of trades	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TIM	0.2932	0.0162	0.1369	0.0123	0.2264	0.0162
Industry-matched	0.2105	0.0000	0.3559	0.0000	0.1481	0.0000
Distance zero-crossings	21.9211	21.3333	26.4237	12.4545	21.6667	21.3333
Cointegration zero-crossings	17.4211	69.3333	26.3220	64.0000	18.0370	69.3333
BMD	0.8072	697.3834	0.9827	1.2782	0.7937	697.3834
Group 3 proportion	0.9691	0.9543	0.2996	0.3119	0.9745	0.9543
Profitable proportion	0.5976	0.5872	0.6607	0.6016	0.4461	0.4601
Total return	0.0072	0.0332	0.0099	0.0256	-0.0014	-0.0147
Number of trades	3.8794	3.5270	1.6705	1.5417	3.9869	3.5013
TIM	0.2656	0.0717	0.1389	0.0310	0.2667	0.0717
Industry-matched	0.2767	0.1462	0.3082	0.1953	0.2772	0.1462
Distance zero-crossings	31.1272	25.4169	31.2443	24.7795	31.0796	25.4169
Cointegration zero-crossings	33.2894	66.2934	32.5016	66.0520	33.1930	66.2934
BMD	1.3142	4.6105	1.3087	4.7388	1.3119	4.6105

Table 4.36: Pair statistics for top 20 unrestricted pairs, execution delay, July 2009–June 2018. The first two columns report the statistics of the standard model with no statistical learning extension, the middle two columns report the statistics of the statistical learning extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the statistical learning extension which inverts the position of trades that are not predicted to be profitable.

variant Sharpe and Sortino ratios are inferior to those of the standard TVHR model, while TIM and return per unit TIM favour the SLN models. The reduction in DTVHR+SLN TIM from 26.34% to 4.64% leads to an increase in return per unit TIM from 0.39% to 2.18%. Similarly, the reduction in TIM for the CTVHR+SLN model to one-seventh its original value, despite a reduction in mean monthly return, leads to an increase in return per unit TIM from 8.26% to 30.13%. These figures illustrate the improved efficiency of the SLN variants at the expense of absolute and risk-adjusted returns.

Table 4.36 reports the pair statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal during the out-of-sample period. Group 1 proportions under the SLI variant are identical to those of the standard TVHR models, while the proportions under the SLN variant are one order of magnitude higher. The majority of pairs under the SLN variant fail to open a single position during

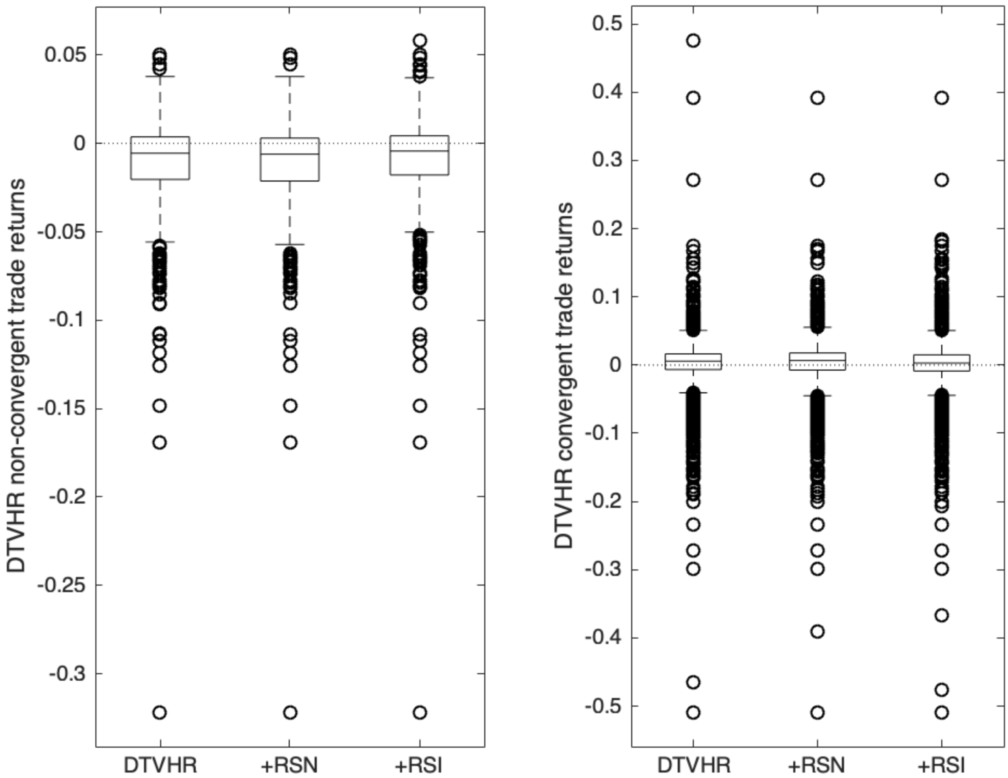

Figure 4.29: Distribution of DTVHR pair excess return of top 20 unrestricted pairs with statistical learning-inferred negation or inversion of trade position, July 2009–June 2018.

the trading period, with 67.14% of DTVHR+SLN pairs and 68.27% of CTVHR+SLN pairs belonging to Group 1.

Figure 4.29 displays the distribution of Group 2 and Group 3 pair returns for the DTVHR model, and Figure 4.30 displays the distribution of Group 2 and Group 3 pair returns for the CTVHR model under both statistical learning variants. The proportion of Group 2 DTVHR+SLI pairs is slightly lower than its DTVHR counterpart at 1.33%, while both SLN models report higher Group 2 proportions than the standard TVHR models at 2.90% for the DTVHR+SLN model, and 0.54% for the CTVHR+SLN model. The profitable proportion of CTVHR pairs declines for both the SLN and SLI variants to approximately half their original proportion, while the profitable proportion of Group 2 DTVHR pairs increases under both the SLN and SLI variants. In particular, the DTVHR+SLI model delivers a profitable Group 2 proportion of 55.56% and a positive total return of 0.38%. The only other model specification that delivers a positive total return for Group 2 pairs is that of the standard CTVHR model, with a profitable proportion of 66.67% and a total return of 0.18%. Both statistical learning variants

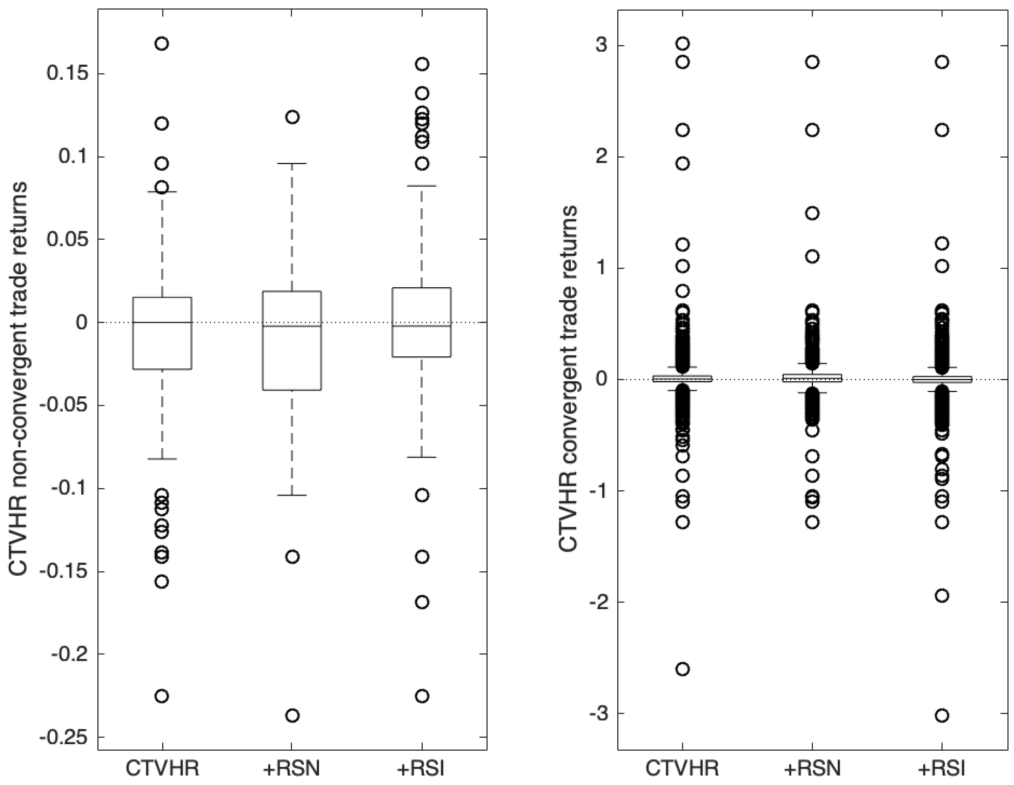

Figure 4.30: Distribution of CTVHR pair excess return of top 20 unrestricted pairs with statistical learning-inferred negation or inversion of trade position, July 2009–June 2018.

offer an improvement in total return for the DTVHR model, but a decline in total return for the CTVHR model.

Group 3 proportions reflect the significant number of Group 2 pairs under the SLN variant, with only 29.96% of DTVHR+SLN pairs and 31.19% of CTVHR+SLN pairs belonging to Group 3. The SLN variant increases the profitable proportion for both models, while the SLI variant decreases the profitable proportion for both models. Total return of Group 3 pairs is negative for both models under the SLI variant, with CTVHR+SLI Group 3 pairs delivering a total return of −1.47%. Number of trades remains relatively consistent for the SLI variant, but are reduced by approximately one-half under the SLN variant, despite the SLN variant delivering total returns commensurate with the standard TVHR models. This increased efficiency indicates a higher return per trade under the SLN variant, a result further illustrated by the halved TIM for both SLN models. As with the regime switching model, industry-matching is relatively consistent across model variants, and BMD reflects a combination of the higher TIM of Group 3 pairs relative to Group 2 pairs, and the faster reversion rate estimated from formation period zero-crossings.

Model variant	TVHR		TVHR+SLN		TVHR+SLI	
	DTVHR	CTVHR	DTVHR	CTVHR	DTVHR	CTVHR
All trades						
Profitable proportion	0.6143	0.5535	0.6381	0.5895	0.4314	0.4741
Mean return	0.0019	0.0086	0.0047	0.0149	-0.0005	-0.0038
Standard deviation	0.0271	0.0976	0.0314	0.0949	0.0268	0.0973
Sharpe ratio	0.3738	0.8773	0.7165	1.5774	-0.1085	-0.3888
Mean profit	0.0161	0.0496	0.0216	0.0612	0.0200	0.0446
Mean loss	-0.0207	-0.0422	-0.0252	-0.0517	-0.0161	-0.0475
Mean long return	0.0065	0.0019	0.0124	0.0051	0.0051	-0.0043
Mean short return	-0.0046	0.0067	-0.0078	0.0098	-0.0056	0.0005
Mean trade length	8.7649	2.5347	10.7767	2.4886	8.4960	2.5527
Median trade length	6.0000	2.0000	8.0000	2.0000	6.0000	2.0000
Convergent trades	0.9152	0.9776	0.8924	0.9802	0.9178	0.9775
Profitable proportion	0.6362	0.5546	0.6701	0.5933	0.4194	0.4732
Mean return	0.0030	0.0090	0.0068	0.0156	-0.0008	-0.0040
Standard deviation	0.0268	0.0984	0.0307	0.0955	0.0268	0.0981
Sharpe ratio	0.6099	0.9072	1.0778	1.6446	-0.1646	-0.3995
Mean profit	0.0165	0.0500	0.0224	0.0618	0.0204	0.0448
Mean loss	-0.0205	-0.0422	-0.0249	-0.0517	-0.0162	-0.0478
Mean long return	0.0073	0.0021	0.0143	0.0054	0.0052	-0.0044
Mean short return	-0.0043	0.0069	-0.0076	0.0102	-0.0061	0.0005
Mean trade length	8.6719	2.5423	10.5516	2.5025	8.4201	2.5601
Median trade length	6.0000	2.0000	8.0000	2.0000	6.0000	2.0000
Non-convergent trades	0.0848	0.0224	0.1076	0.0198	0.0822	0.0225
Profitable proportion	0.3775	0.5064	0.3729	0.4000	0.5662	0.5128
Mean return	-0.0105	-0.0079	-0.0128	-0.0229	0.0025	0.0024
Standard deviation	0.0272	0.0512	0.0324	0.0525	0.0264	0.0517
Sharpe ratio	-1.9562	-1.6489	-1.7643	-5.1608	0.4944	0.4911
Mean profit	0.0092	0.0263	0.0103	0.0225	0.0166	0.0360
Mean loss	-0.0224	-0.0430	-0.0266	-0.0532	-0.0159	-0.0330
Mean long return	-0.0026	-0.0042	-0.0036	-0.0106	0.0035	0.0010
Mean short return	-0.0079	-0.0036	-0.0092	-0.0123	-0.0010	0.0014
Mean trade length	9.7686	2.2051	12.6441	1.8000	9.3440	2.2308
Median trade length	6.0000	2.0000	9.0000	1.0000	6.0000	2.0000

Table 4.37: Trade statistics for top 20 unrestricted pairs, execution delay, July 2009–June 2018. The first two columns report the statistics of the standard model with no statistical learning extension, the middle two columns report the statistics of the statistical learning extension which ignores trades that are not predicted to be profitable, and the final two columns report the statistics of the statistical learning extension which inverts the position of trades that are not predicted to be profitable.

Table 4.37 reports the individual trade statistics for portfolios of the top 20 unrestricted pairs with one-day execution delay upon generation of a trading signal during the out-of-sample period. The profitable proportion of trades increases under the SLN variant from 61.43% to 63.81% for DTVHR pairs, and from 55.35% to 58.95% for CTVHR pairs. Both models under the SLI variant, however, deliver a profitable proportion below 50% with negative mean returns. The SLN variant more than doubles the mean return of DTVHR pairs and almost doubles the mean return of CTVHR pairs. These higher mean returns combined with comparable return standard deviations lead to higher Sharpe ratios, with the DTVHR+SLN Sharpe ratio increasing from 0.37 to 0.72, and the CTVHR+SLN Sharpe ratio increasing from 0.88 to 1.58. Mean profit and mean loss both increase in magnitude for both models under the SLN variant, while only the mean short return of DTVHR+SLN trades reports a deterioration in performance. Mean and median trade length increase for the DTVHR+SLN model but remain relatively consistent for all other model specifications. Figure 4.31 displays the distribution of convergent and non-convergent trade returns for the DTVHR model, and Figure 4.32 displays the distribution of convergent and non-convergent trade returns for the CTVHR model under both statistical learning variants.

The proportion of convergent trades declines slightly under the DTVHR+SLN model from 91.52% to 89.24%, but remains relatively stable and above 90% for all other model specifications. Due to the high proportion of convergent trades, the trade statistics of convergent trades are similar to those of all trades. As with all trades, the profitable proportion, mean return and Sharpe ratio of the SLN variant outperform those of the standard TVHR model, while the same statistics of the SLI variant underperform the standard TVHR model.

Proportions of non-convergent trades remain below 10% for all model specifications with the exception of the DTVHR+SLN model, whose non-convergent trade proportion is 10.76%. The profitable proportion under the SLN variant declines slightly for DTVHR pairs from 37.75% to 37.29%, and more substantially for CTVHR pairs from 50.64% to 40.00%. The profitable proportion of trades under the SLI variant, by contrast, increases to 56.62% for DTVHR pairs and 51.28% for CTVHR pairs. Only the SLI variants deliver positive mean returns for non-convergent trades, in accordance with its objective of generating positive returns from mean-averting spreads. The greater magnitude of losses under the DTVHR+SLN model is offset by an increased standard deviation of returns and mean trade length to deliver a slightly improved Sharpe ratio of −1.76 relative to the standard DTVHR model's Sharpe ratio of −1.96. The substantially greater magnitude of loss in combination with a reduced mean trade length inflates the magnitude of the CTVHR+SLN model's Sharpe ratio from −1.65 to −5.16. Mean long return and mean short return increase in magnitude for both model specifications under the SLN variant,

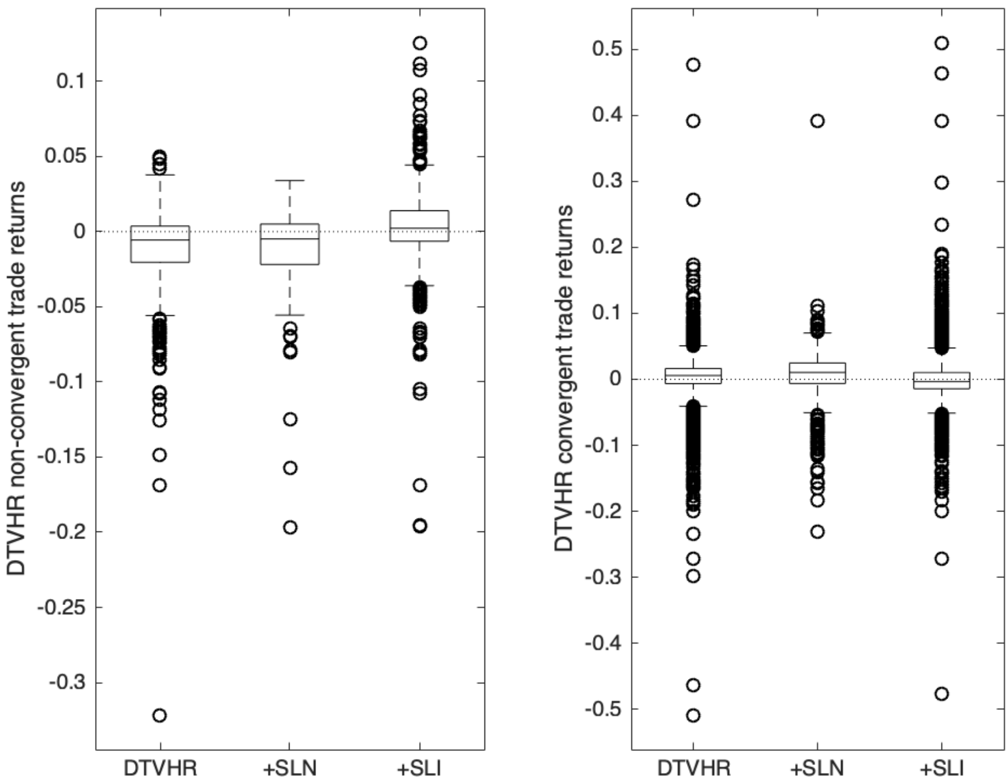

Figure 4.31: Distribution of DTVHR trade excess return of top 20 unrestricted pairs with statistical learning-inferred negation or inversion of trade position, July 2009–June 2018.

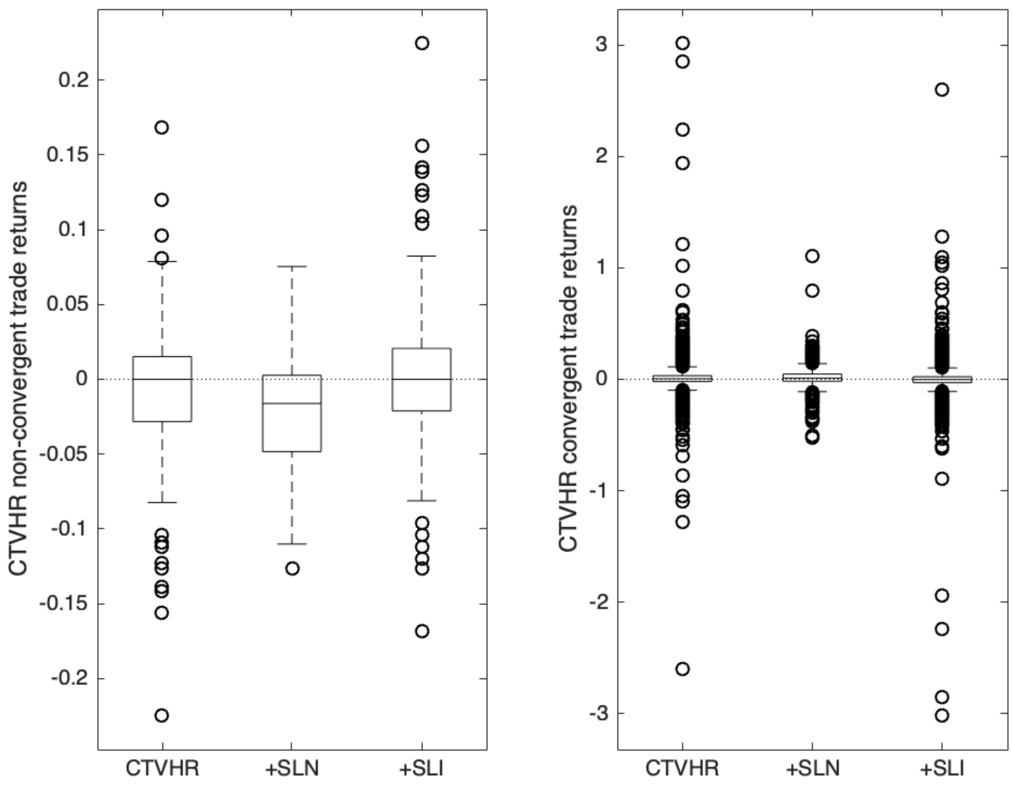

Figure 4.32: Distribution of CTVHR trade excess return of top 20 unrestricted pairs with statistical learning-inferred negation or inversion of trade position, July 2009–June 2018.

while only mean loss under the CTVHR+SLN model delivers a substantial increase over that of the standard CTVHR model.

4.3.1 Summary of Statistical Learning Model Extension

Section 2.3 posed the research question, *are statistical learning models better equipped than conventional models to capture and detect latent market regimes?* The regime switching model extension investigated in Section 4.2 substantially improved the per-trade returns of the CTVHR model under the RSN regime switching variant, offering some evidence that statistical arbitrage returns under the cointegration specification are dependent on the prevailing volatility regime. High and low volatility regimes are a recognised market phenomenon that indirectly influence security returns, but it is possible that there are other, subtler regimes that influence the latent dynamics of the financial markets. Such regimes may not be detectable by conventional data or models, so Section 4.3 addresses the research question by incorporating exogenous variables and the universal mapping capability of ELMs.

As with the regime switching model extension, the objective of the statistical learning extension is to forecast the probability that a trade will be profitable, either negating trades that are predicted to be unprofitable, or inverting the trade signal to exploit mean-averting behaviour as the spread of a pair continues to diverge. The CTVHR+SLN model outperformed the standard CTVHR model in terms of portfolio mean monthly return in-sample, but failed to maintain that outperformance out-of-sample. The DTVHR+SLN model failed to outperform the standard DTVHR model in- and out-of-sample in terms of mean monthly return. Both model specifications under the SLN variant delivered their portfolio returns more efficiently than the standard TVHR model in terms of return per unit TIM, however. Portfolios formed under the SLI variant underperformed the standard TVHR model for both pair specifications and both in- and out-of-sample, illustrating the difficulty in exploiting mean-averting behaviour simply by inverting the trading rules for trades that are predicted to be unprofitable.

In contrast to the regime switching model investigated in Section 4.2 which was only able to deliver improved trade performance for CTVHR pairs, the statistical learning model under the SLN variant delivered greater per-trade performance for both DTVHR and CTVHR pairs out-of-sample. Profitable proportion, mean return and Sharpe ratio all improved under the SLN variant relative to the standard TVHR model. Additionally, while the SLI variant underperformed the standard TVHR model in terms of portfolio mean and risk-adjusted returns, it was the only model specification that delivered positive mean returns for non-convergent trades. Though the statistical learning model was unable to translate its forecasting ability into additional portfolio returns, its improvement of per-trade performance under the SLN variant, and delivery of profitable non-convergent trades under the SLI variant, offer some evidence that statistical learning models are better equipped than conventional models to capture and detect latent market regimes, especially when considering their comparative statistical performance reported in Tables 4.32 and 4.33.

5 Conclusion and Future Work

This thesis has explored the current state of research into statistical arbitrage phenomena within the broader context of capital market anomalies, and in particular its limitations and empirical failure. The literature reviewed in Chapter 2 spans the initial pairs trading investigations under the distance approach developed by Gatev, Goetzmann, and Rouwenhorst (2006) and scrutinised by Do and Faff (2010), refinement of the pairs trading algorithm under the cointegration approach proposed by Meucci (2009) among other researchers, a more rigorous treatment of the phenomenon under the time series approach of Elliott, Van Der Hoek, and Malcolm (2005), and the abstraction of the definition of statistical arbitrage under numerous alternative approaches.

Gatev, Goetzmann, and Rouwenhorst (2006) find evidence of the declining profitability of statistical arbitrage in their updated study, further confirmed by Do and Faff (2010) who attribute the decline to the growing proportion of pairs that fail to converge following trade initiation. Rad, Low, and Faff (2016) find that the non-convergence of statistical arbitrage opportunities is not unique to the distance approach, demonstrating its incidence among cointegration and alternative approaches in addition to the distance approach.

Uninformed demand shocks are identified by Andrade, Di Pietro, and Seasholes (2005) as the driver of statistical arbitrage opportunities. Uninformed trading takes place on one of the constituent securities, most commonly the rising overvalued security, facilitating spread divergence which is subsequently corrected by informed traders. Idiosyncratic shocks, on the other hand, change factor loadings on market risk premia and cause persistent divergence and the failure of arbitrage relationships. Firm-specific earnings announcements and general media coverage are found to degrade statistical arbitrage performance, while market-wide volatility improves performance due to the increased presence of uninformed trading.

Irrespective of the underlying cause, spread non-convergence is attributable to continued spread divergence following the identification of a statistical arbitrage opportunity, resulting in either a temporary albeit prolonged deviation from normal spread dynamics, or the establishment of a new equilibrium level between the constituent securities. In both cases, the static estimate of the spread equilibrium level no longer represents the appropriate reference point with which to consider spread convergence, leading to trade losses and the opportunity cost of other arbitrage opportunities that cannot be exploited while capital is deployed on the non-convergent opportunity.

Section 2.3 summarises the state of statistical arbitrage research and poses three research questions, the first of which being motivated by the empirical failure of arbitrage opportunities due to spread non-convergence:

> *Is the assumption of static arbitrage relationships responsible for the declining profitability of statistical arbitrage?*

The proposed TVHR model, with its time-varying estimation of the equilibrium level in the arbitrage relationship, addresses non-convergence by forcing the stochastic spread of statistical arbitrage opportunities to converge within a brief timeframe informed by the formation period dynamics. The brevity of the timeframe is supported by the literature; a number of investigations find that early convergence of pairs trading opportunities contributes most to the profitability of statistical arbitrage, with Jacobs and Weber (2015) finding that the majority of pair excess returns are generated by those pairs that converge within a month from the date of trade initiation. Similarly, Huck (2015) instructs that trade should not be delayed once a mis-pricing has been identified, and Engelberg, Gao, and Jagannathan (2009) find that by condensing the trading period to just 10 days, pairs generate monthly excess returns that exceed those of the conventional six-month trading period by 1.05%. This evidence, along with the increasing prevalence of time-varying estimation procedures such as those proposed by Montana, Triantafyllopoulos, and Tsagaris (2009), Dunis, Laws, and Evans (2006) and Stübinger and Bredthauer (2017), empirically motivates the TVHR model.

Chapter 5 evaluates the economic utility of the proposed model alongside the conventional distance approach of Gatev, Goetzmann, and Rouwenhorst (2006), and a cointegration approach developed in this thesis and motivated by the prescriptions of Krauss (2017). The empirical evaluation follows the backtesting procedure of Gatev, Goetzmann, and Rouwenhorst (2006) and Do and Faff (2010), extending the sample period by nine years for the first investigation of the distance approach since its most recent publication. The initial study period investigated in Section 4.1.1, spanning July 1962–June 2009, demonstrated the favourable risk-adjusted performance of the TVHR model and comparable portfolio returns relative to the conventional distance approach upon immediate execution of the trade. Delaying the trade by one day has a deleterious effect on both distance and cointegration pairs, with the effect especially pronounced on trades placed under the TVHR model. While returns are depressed under the imposition of the trade delay, the TVHR model continues to deliver competitive portfolio returns for both distance and cointegration pairs, outperforming static variants' portfolio and per-trade returns in risk-adjusted terms.

Of special relevance to the research question is the vanishingly small proportion of pairs under the TVHR model that place a single non-convergent trade during the trading period. Constituting just 0.37% of distance pairs and 0.38% of cointegration pairs, the non-convergent proportions under the TVHR model are dwarfed by the static variant proportions of 29.42% and 35.03%, respectively. Given sufficient time, statistical arbitrage opportunities exploited under the TVHR model are assured by construction to converge due to the time-varying estimation of the equilibrium level of the arbitrage relationship, ensuring that divergence is only temporary in contrast to the possibility of permanent divergence under the conventional static approaches.

Partitioning the initial sample period into sub-periods spanning July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009, the declining performance of statistical arbitrage is evident. The initial sub-period delivers the greatest mean monthly portfolio return for both distance and cointegration pairs under static and TVHR execution models. Each successive sub-period delivers progressively poorer mean portfolio returns and Sharpe and Sortino ratios for distance pairs under both execution models. Cointegration pairs traded under the TVHR model, on the other hand, deliver an intermediate portfolio return and their highest Sharpe and Sortino ratios in the final sub-period. Static cointegration pairs also improve their risk-adjusted returns between successive sub-periods despite delivering declining mean returns. Individual trade statistics favour the TVHR model for both pair selection methods in terms of Sharpe ratio, despite them being unable to match their static counterparts' mean trade returns. In all sub-periods, the TVHR model continues to deliver small proportions of non-convergent pairs, with the greatest distance and cointegration proportions of 0.88% and 0.44% significantly smaller than the lowest static variant proportions of 23.53% and 34.59%, respectively. Both static and TVHR model variants observe their greatest proportion of non-convergent pairs in the final sub-period.

Extending the analysis beyond the sample period used by Do and Faff (2010) to consider model performance across July 2009–June 2018 establishes the failure of the distance approach under the static estimation procedure. Mean monthly portfolio return during the sample period declines to 0.00%. By contrast, static cointegration pairs deliver a mean monthly portfolio return of 0.67% and an improved Sortino ratio relative to the preceding sub-period. The TVHR model delivers a mean monthly return for distance pairs that exceeds that of the static model for the first time in the analysis, returning 0.10% with Sharpe and Sortino ratios several orders of magnitude greater than those of its static counterpart. Cointegration pairs under the TVHR model achieve their greatest mean monthly return and Sortino ratio of any sub-period, returning 0.56% with a Sortino ratio of 3.54. The TVHR model continues to deliver small proportions of non-convergent pairs during the period, with trades that generate superior mean returns in the case of distance pairs, and superior Sharpe ratios for both distance and cointegration pairs.

While the proposed TVHR model attenuated the declining profitability of statistical arbitrage in the final period, it did not prevent it in the case of distance pairs, though the decline across all sub-periods was less pronounced than for pairs traded under the static model. The TVHR model, however, was able to deliver greater risk-adjusted returns under both pair selection procedures than its static counterpart, and near-null proportions of non-convergent pairs. The BMD statistic, which quantifies the average deviation of the time-varying estimate of the hedge ratio from its initial formation period estimate, consistently reported higher deviations for non-convergent pairs than for convergent pairs, validating the hypothesis that pair spreads are not stationary. Given these findings, and in view of the first research question, there is evidence to conclude that the assumption of a static arbitrage relationship is at least partially responsible for the declining profitability of distance pairs.

The first research question does not distinguish between the pair selection procedure used to identify statistical arbitrage opportunities. Evidence presented by Rad, Low, and Faff (2016) demonstrates the incidence of pair non-convergence among cointegration and copula-based alternative approaches. The cointegration approach developed in this thesis, by contrast, delivers some of its greatest returns in the extended study period, halting and even reversing the trend of declining profitability. The TVHR model extends that profitability further, particularly in risk-adjusted terms. Determination of the magnitude of increasing market efficiency is beyond the scope of this thesis, though the continued profitability of the cointegration approach indicates that the declining profitability of statistical arbitrage is largely a concern for pairs identified under the distance approach.

General characteristics of the TVHR model include high risk-adjusted performance and high capital efficiency, driven primarily by the brief trade lengths observed under the model, allowing it to deliver high returns per unit TIM. As discussed by Jacobs and Weber (2015), Huck (2015) and Engelberg, Gao, and Jagannathan (2009), the majority of pairs' profitability is observed in the first days following the identification of a trade opportunity—a feature of statistical arbitrage that the TVHR model capitalises on. While the imposition of the one-day trade execution delay depresses the TVHR model's profitability, it maintains its superior risk-adjusted portfolio and per-trade returns relative to the static model.

Restricting pairs to those that are selected from the same industry improves TVHR portfolio mean monthly return at the expense of risk-adjusted performance for distance pairs. Cointegration pairs see an improvement in both portfolio mean and risk-adjusted returns for all but Financial industry pairings. The inability of distance pairs to deliver better risk-adjusted returns under pair restriction is due to the loss of portfolio diversification caused by restriction—the industry-restricted portfolio is susceptible to idiosyncratic return volatility within each respective industry. Only pairs formed among Industrials, Transportation, Utilities and Financials stocks were considered in this thesis in accordance with Do and Faff (2010).

The TVHR model requires the specification of five parameters, namely the divergence z-score that initiates a trade, z_d, the convergence z-score that closes a trade, z_c, and the EMA decay parameters that determine the updating speed of the process mean, bias, and volatility level estimates, $\alpha_0, \alpha_1, \alpha_2$, respectively. For simplicity, this thesis used a single value for each of the EMA decay parameters, α, estimated from the number of formation period zero-crossings, while values of $z_d = 2$ and $z_c = 0$ were employed for consistency with the surveyed literature. Section 4.1.4 explored the sensitivity of the portfolio performance to variations in these parameter choices.

Decreasing the divergence z-score to $z_d = 1$ while leaving the convergence z-score unaltered has the effect of increasing portfolio performance at the expense of per-trade performance. The narrower significance bands initiate trades sooner and more frequently than under the standard value of $z_d = 2$, but the exploitation of less significant spread divergence causes diminished trade performance. Conversely, increasing the divergence z-score to $z_d = 3$ improves per-trade performance at the expense of portfolio performance, with portfolio mean monthly returns of only 0.11% and 0.07% for distance and cointegration pairs, respectively. While the arbitrageur may find the portfolio performance of lower divergence z-scores enticing, transaction costs inhibit the profitable exploitation of arbitrage opportunities that are initiated under an excessively low divergence z-score. This represents one of the practical limits to arbitrage, where the arbitrageur must balance portfolio performance with the ability of trades to deliver returns in excess of transaction costs, implicitly considered in this thesis under the imposition of the one-day execution delay.

Given the very short duration of trades placed under the TVHR model, Section 4.1.4 explores the impact of varying the EMA decay parameter, α, to slow the update speed and consequently extend the duration of trade holding periods. This is accomplished by reducing the number of formation period zero-crossings to one-half and one-tenth, respectively, relative to the true number of zero-crossings observed for each pair. In contrast to the divergence z-score, slowing the reversion rate of the TVHR model improves both portfolio mean return and trade mean return for both distance and cointegration specifications. Slowing the reversion rate also causes the TVHR model to more closely resemble the portfolio and trade characteristics of the conventional static arbitrage model while retaining more favourable risk-adjusted returns and trade non-convergence rates. The EMA decay parameter therefore presents a more attractive avenue for model optimisation than the divergence parameter.

Excess returns generated by the TVHR model are found to be positive and statistically significant, with significant negative factor loadings on the market equity premium and Fama-French SMB factor for distance pairs, and significant positive factor loadings on momentum for distance pairs, and short-term reversion for both distance and cointegration pairs. TVHR model returns are less correlated with traditional risk premia than their static counterparts, with the notable exception of short-term reversion. Only distance pair returns under the TVHR model are found to be positively correlated with market volatility, evidenced by a positive and statistically significant loading on VIX level. Distance pair returns are also found to be susceptible to declining profitability across the entire sample period, July 1962–June 2018, delivering a negative statistically significant loading on a linear time trend. Cointegration pairs traded under the TVHR model do not share this negative correlation with the linear time trend.

The TVHR model delivers its objective of neutralising pair non-convergence. The result is a model that is flexible and adaptable, able to be optimised by the arbitrageur and evolve with changing market dynamics. Despite its inability to halt the declining profitability of distance pairs, it represents a simple and robust way to observe the time-varying equilibrium relationship between co-moving securities, allowing the arbitrageur to exploit statistical arbitrage opportunities efficiently.

The second research question posed in this thesis concerns the evidence of regimes in statistical arbitrage returns, and their inconsistent application in the literature:

Are statistical arbitrage returns dependent on the prevailing volatility regime?

A substantial body of evidence, explored in Chapter 2 and Section 2.2.1 in particular, emphasises the importance of volatility regimes on the outcome of statistical arbitrage opportunities. Do and Faff (2010) and Liu, Chang, and Geman (2017) find that times of market turmoil contribute to improved absolute and risk-adjusted returns for the distance approach. Huck (2015) is unable to determine a relationship between statistical arbitrage returns and a volatility-based regime switching framework, though the author reasons that trade execution delay induced by the regime filter serves to decouple volatility and return dynamics. Caldeira and Moura (2013) are unable to find evidence of outperformance during periods of market turmoil, while Chen, Chen, and Chen (2014) and Bee and Gatti (2015) explicitly model statistical arbitrage spreads in the presence of different volatility regimes, finding greater performance than modelling paradigms that consider only a single regime.

Section 4.2 extends the TVHR model with a regime switching framework that classifies the outcome of a trade as either profitable or unprofitable, aligning the objective function of the model with that of the arbitrageur. The logistic regression model employed in the regime switching extension offers a simple procedure for classifying trades based on the prevailing volatility regime, expressed in terms of the VIX level and the magnitude of pair-specific volatility at the time of trade initiation. Such a regime switching approach captures the essence of more sophisticated Markov switching models, most of which in the statistical arbitrage domain are applied to the modelling of high and low volatility regimes, while maintaining model simplicity and allowing comparison with the more advanced statistical learning extension that additionally considers exogenous variables.

The regime switching model extension considered both the negation of trades whose model-forecast probability of profitability was too low, and the inversion of such trades in accordance with the exploration of mean-averting momentum regimes in statistical arbitrage. Regime switching models estimated for distance and cointegration pairs both demonstrate a significant positive factor loading on pair-specific volatility, but neither pair specification has a significant loading on VIX level. This is likely due to the brief duration of TVHR trades, whose success or failure is more dependent on idiosyncratic volatility than market-wide volatility at the time of trade initiation. Both regime switching models deliver poor statistical performance, with all trades during the in-sample period predicted to be profitable under a classification threshold of 50%, and AUROC scores barely surpassing the minimum possible value of 0.50.

In-sample portfolio returns over the period January 1990–June 2009 demonstrate the ability of the regime switching model to improve the performance of cointegration pairs whose unprofitable trades are negated, though that improved performance does not extend to the inversion of cointegration trades, nor to either variant of distance pairs. The out-of-sample period, July 2009–June 2018, further demonstrates the inability of the regime switching model to improve portfolio returns for all but the cointegration TVHR model under the trade negation variant, whose improved portfolio returns are attributable to the more efficient use of capital among profitable trades—the substantial reduction in TIM and consequent improvement in return per unit TIM are evidence of the model's improved efficiency.

The improved efficiency of cointegration pair trades under the regime switching model is reflected in the per-trade performance of the negation variant. Relative to the standard TVHR model with no regime filter, mean trade return more than doubles from 0.86% to 1.73%, and Sharpe ratio increases from 0.88 to 1.39. The improvement is driven by a doubling of convergent trade mean return, itself primarily driven by a substantial increase in mean short return. The proportion of convergent trades does not differ substantially from that observed under the standard TVHR model, though this is not due to a failure of the regime switching model but rather an inherent property of the TVHR model—convergence is assured given sufficient time, so the regime switching model identifies unprofitable trades caused by forces unrelated to spread non-convergence. Given these findings, and in view of the second research question, there is evidence to conclude that statistical arbitrage returns are dependent on the prevailing volatility regime, though only when considering cointegration pairs. There was no observed improvement in the performance of distance pairs under either regime switching variant, in- or out-of-sample.

The exploration of time-varying arbitrage relationships, such as those investigated by Montana, Triantafyllopoulos, and Tsagaris (2009), Triantafyllopoulos and Montana (2011), Burgess (2000), Dunis, Laws, and Evans (2006) and Stübinger and Bredthauer (2017) further justify the consideration of regimes, though ones that are not explicitly identifiable as known econometric regimes. Additionally, the exploration of mean-averting or momentum regimes by Hogan, Jarrow, Teo, and Warachka (2004) and Krauss and Stübinger (2017) give further evidence of unconventional regimes. Sections 2.1.4 and 2.2.2 explore alternative approaches to the identification and exploitation of statistical arbitrage, some of which incorporate statistical learning models. Huck (2010), Montana and Parrella (2009), Dunis, Laws, and Evans (2008), Dunis, Laws, Middleton, and Karathanasopoulos (2015) and Nóbrega and Oliveira (2013), for example, all apply some variation of ANN, SVR or ELM model to the identification of temporal mis-pricings between securities. The ubiquitous presence and successful application of statistical learning models to statistical arbitrage motivates the third and final research question:

> *Are statistical learning models better equipped than conventional models to capture and detect latent market regimes?*

Both the second and third research questions are concerned with the presence and identification of regimes, though while the former considers conventional volatility regimes identified by a simple modelling approach, the latter expands the definition of regimes to consider those that cannot be easily classified. As such, both the second and third research questions seek to quantify the economic and statistical benefit offered by identifying regimes that are favourable for the exploitation of statistical arbitrage opportunities, differing only in their approach to the identification of those regimes.

Section 2.2.2 builds on the modelling approach investigated in Section 2.2.1, drawing on the universal function mapping capability of ELMs to estimate the probability of trade profitability in the presence of additional input variables. Specifically, in addition to the volatility variables employed by the regime switching model, the statistical learning model incorporates the probability forecast generated by the logistic regression model, a dummy variable indicating whether the constituent securities in the pair are from the same industry, the pair's position in the top 20, and the TVHR model's mean-reversion rate parameter. The combination of the ELM model and these additional exogenous variables improves the statistical performance relative to the regime switching model. Model-forecast probabilities of trade profitability vary from 1.63% to 100.00% for distance pairs, and from 0.00% to 100.00% for cointegration pairs. Additionally, while AUROC and Youden's Index remain low for both distance and cointegration pairs, both are improved relative to the figures observed under the regime switching model.

As with the regime switching model, the statistical learning model only improves cointegration portfolio mean returns in-sample, and only for the variant that negates unprofitable trades. In contrast to the regime switching model, however, no statistical learning variant delivers improved portfolio returns in the out-of-sample period, though both distance and cointegration pairs realise substantially greater efficiency in terms of TIM and return per unit TIM when unprofitable trades are negated. The significant increase in the proportion of Group 1 pairs—those that do not place a single transaction during the trading period—facilitates a greater proportion of capital being allocated to the remaining pairs. The resultant portfolios are less diverse than would otherwise be observed, leading to lower risk-adjusted returns than the standard TVHR portfolios.

The improved per-trade performance under the statistical learning model extends to both distance and cointegration pairs whose unprofitable trades are negated, in contrast to the regime switching extension which only delivered an improvement to cointegration pairs. The profitable proportion, mean return and Sharpe ratio of trades all improve for both distance and cointegration pairs under the statistical learning model. While mean trade returns are higher for cointegration pairs under the regime switching model, profitable proportion and Sharpe ratio are higher under the statistical learning model. With reference to the third research question, there is evidence to conclude that statistical learning models are better equipped than conventional models to capture and detect latent market regimes. Their ability to incorporate unconventional, unstandardised exogenous variables, in addition to their universal function mapping capability, allow them to more accurately model the underlying drivers of statistical arbitrage profitability.

This thesis has produced a number of contributions to statistical arbitrage theory in the course of answering the research questions it posited, specifically:

- Demonstration of the sub-optimality of the distance approach in the selection of candidate pairs, greatly restricting the number of opportunities considered.

- Reformulation of the spread variable as an OU process to unify the distance, cointegration and time series approaches.
- Demonstration of the equivalence of the unified spread formulation with its conventional analogues.
- Extension of the unified spread formulation to accommodate a time-varying mean in the OU process, leading to the proposed TVHR model.
- Development of a mathematically tractable discrete TVHR model and accompanying parameter estimation procedure.
- Proposal of a theoretically-sound procedure for selecting pairs under the cointegration approach.
- Empirical evaluation of the proposed TVHR model alongside its static distance and cointegration analogues, following the testing procedure of Gatev, Goetzmann, and Rouwenhorst (2006) and extending the sample period beyond that considered by Do and Faff (2010).
- Confirmation of the continuing trend of declining profitability for pairs selected under the distance approach, and the robustness of both the proposed cointegration approach and TVHR model in opposing that trend.
- Exploration of the relationship between volatility and statistical arbitrage returns, finding evidence of dependence on the prevailing volatility regime for the cointegration approach.
- Exploration of the ability of statistical learning to more accurately forecast statistical arbitrage profitability, finding evidence of superior performance relative to a simple regime switching model.

These contributions offer evidence to conclude that statistical arbitrage constitutes an effective means of generating excess returns that are largely uncorrelated with the market, and the work in this thesis refines the theory and implementation of the phenomenon's exploitation.

A number of limitations constrained the analysis conducted in this thesis. For simplicity and continuity, the simple evaluation procedure developed by Gatev, Goetzmann, and Rouwenhorst (2006) and extended by Do and Faff (2010) was replicated, allowing for the comparative evaluation of the various approaches. Following Gatev, Goetzmann, and Rouwenhorst (2006), transaction costs were not imposed on any trades, instead relying on the implicit cost imposed by delaying execution for one day. This implicit cost offers a conservative estimate of model performance, depressing trade returns to a greater degree than the round-trip transaction costs estimated by Rad, Low, and Faff (2016). No short-sale constraints were considered, and a risk-free rate of zero was assumed due to the self-financing nature of statistical arbitrage. Future work should incorporate more realistic transaction cost modelling, eliminating the execution delay in favour of a modelling approach that considers the impact of short-sale costs, bid-ask spreads, slippage, and commissions payable to brokers.

The TVHR model proposed in this thesis requires the specification of five parameters, though for practical reasons only one was estimated in the analysis—that is, the EMA reversion rate parameter which was estimated from the number of formation period zero-crossings. The rigid evaluation procedure proposed by Gatev, Goetzmann, and Rouwenhorst (2006) and replicated in this thesis considers a relatively brief formation period of only 12 months, over which daily observations of the pair spread are calculated. A consequence of using such a brief formation period is the over-estimation of formation period zero-crossings, leading to the estimation of an EMA reversion rate that updates the time-varying hedge ratio too quickly. As discussed in Section 4.1.4, the reversion rate can be deflated so that the TVHR model will more closely resemble a static approach. This flexibility allows the TVHR model to be tuned to optimise performance, though optimisation is itself a non-trivial domain.

More fundamentally, the TVHR model does not have its own identification procedure, instead relying on distance and cointegration approaches to select appropriate pairs. A truly unified model for statistical arbitrage would require its own identification procedure, though the flexibility offered by the TVHR model's five tuning parameters would make such an identification procedure unnecessary. Instead, given the abundance of high-performance computing assets and the prevalence of big data applications, future work should consider the optimisation of the TVHR model to find the most suitable pair candidates and parameter combinations among the universe of investable securities. The researcher would, for example, specify a grid of potential parameter values, testing each set of parameters on every combination of securities and selecting the parameter-security combinations that optimise some chosen function of strategy performance, such as the Sharpe ratio. Though the chosen parameter-security combinations would almost certainly suffer from data-mining bias given the multiple testing approach, such bias could be tempered or eliminated by optimising on synthetic financial time series whose dynamics reflect those of the series under consideration. Lopez De Prado (2018) proposes the use of an AR(1) model to estimate the data generating process of the underlying securities, generating 100,000 or more synthetic time series on which the strategy can be evaluated. The researcher would select only the parameter-security combinations whose mean Sharpe ratio, for example, across all 100,000 backtests on synthetic data meet some acceptance threshold. In so doing, the TVHR model extends the conventional search among candidate securities to include a search among parameter combinations, controlling for the effect of data-mining bias by evaluating the phase space of combinations on synthetic financial time series.

The regime switching model extension sought to augment the TVHR model with a filter that either negated or inverted trades based on the probability of trade profitability, informed by the prevailing volatility regime. The VIX level was not found to contribute significantly to the model forecast of trade profitability, though this is likely due to the brief holding periods of TVHR trades. Future research should consider the influence of market volatility on statistical arbitrage opportunities that emerge over longer periods of time. The incorporation of a realistic transaction cost model would further isolate the influence of market volatility, allowing the function approximation to be optimised without the one-day execution delay obscuring the objective.

Future research may also explore the use of either a regime switching or statistical learning model in a capital allocation scheme. The significant increase in the proportion of pairs that did not place a trade during the trading period was responsible for lower portfolio returns under both model extensions, so the forecast probability of trade profitability could instead be used to weight an allocation of capital to each statistical arbitrage opportunity. Doing so will ensure that all trades are placed, but that trades with a low probability of profitability are given an appropriately small allocation of capital. While model extensions that negated trades showed the greatest per-trade performance, it is likely that inverted trades could also generate excess returns if correctly identified by the model. Rather than having just two classes, namely profitable and unprofitable, the model could be extended to consider three classes: profitable under a standard trade, unprofitable, and profitable under an inverted trade. Such a model would capture the essence of the emerging literature that finds evidence of statistical arbitrage opportunities under both mean-reverting and mean-averting regimes.

Given the superior performance of the statistical learning model in identifying profitable trades, future research should also expand the investigation to consider other statistical learning models and other exogenous data. Equity market premium, risk-free rate, market sentiment and other exogenous data could all be considered by ANNs, SVMs, random forest or even deep learning models. Additionally, it would be worth exploring the comparative performance of classification and regression models, where a regression model might instead forecast the expected return of a trade for the consideration of the arbitrageur.

This thesis has found evidence to conclude that time-varying equilibrium relationships more closely resemble the empirical features of statistical arbitrage, that cointegration approach pairs are dependent of the prevailing volatility regime, and that statistical learning models are better equipped than conventional models to capture and detect latent market regimes. The questions motivating the research that produced this evidence are concerned with the declining profitability of statistical arbitrage, and whether it can be mitigated through appropriate measures. While these conclusions are unable to fully address the question of declining profitability, the above considerations will allow future research to explore the phenomenon more thoroughly and arrive at a more definitive conclusion.

Bibliography

Abdou, H., J. Pointon, and A. El-Masry (2008). "Neural Nets versus Conventional Techniques in Credit Scoring in Egyptian Banking". In: *Expert Systems with Applications* 35.3, pp. 1275–1292. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2007.08.030.

Alexander, C. and A. Dimitriu (2005). "Indexing and Statistical Arbitrage". In: *The Journal of Portfolio Management* 31.2, pp. 50–63. ISSN: 0095-4918. DOI: https://doi.org/10.3905/jpm.2005.470578.

Andrade, S., V. Di Pietro, and M. Seasholes (2005). *Understanding the Profitability of Pairs Trading*. Tech. rep. U.C. Berkeley, Northwestern University. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.1045&rep=rep1&type=pdf.

Ang, A. and G. Bekaert (2002a). "International Asset Allocation with Regime Shifts". In: *The Review of Financial Studies* 15.4, pp. 1137–1187. ISSN: 0893-9454. DOI: https://doi.org/10.1093/rfs/15.4.1137.

– (2002b). "Regime Switches in Interest Rates". In: *Journal of Business & Economic Statistics* 20.2, pp. 163–182. ISSN: 07350015. DOI: https://doi.org/10.3386/w6508.

Ang, A. and A. Timmermann (2012). "Regime Changes and Financial Markets". In: *Annual Review of Financial Economics* 4.1, pp. 313–337. ISSN: 1941-1375. DOI: https://doi.org/10.2139/ssrn.1919497.

Avellaneda, M. and J. Lee (2010). "Statistical Arbitrage in the US Equities Market". In: *Quantitative Finance* 10.7, pp. 761–782. ISSN: 14697688. DOI: https://doi.org/10.2139/ssrn.1153505.

Barro, R. (2006). "Rare Disasters and Asset Markets in the Twentieth Century". In: *The Quarterly Journal of Economics* 121.3, pp. 823–866. ISSN: 0033-5533. DOI: https://doi.org/10.1162/qjec.121.3.82.

Bartlett, P. (1997). "For Valid Generalization, the Size of the Weights Is More Important Than the Size of the Network". In: *Advances in Neural Information Processing Systems* 9, pp. 134–140. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.57.2302&rep=rep1&type=pdf.

Bee, M. and G. Gatti (2015). *An Improved Pairs Trading Strategy Based on Switching Regime Volatility*. DEM Discussion Papers 13. Department of Economics and Management. URL: https://ssrn.com/abstract=2636212.

Bertram, W. (2010). "Analytic Solutions for Optimal Statistical Arbitrage Trading". In: *Physica A: Statistical Mechanics and its Applications* 389.11, pp. 2234–2243. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2010.01.045.

Bianchi, R., M. Drew, and Y. Zhu (2009). "Pairs Trading Profits in Commodity Futures Markets". In: *Proceedings of the Asian Finance Association 2009 International Conference*. University of Queensland, pp. 1–26. URL: https://core.ac.uk/download/pdf/143879548.pdf.

Bock, M. and R. Mestel (2009). "A Regime-Switching Relative Value Arbitrage Rule". In: *Operations Research Proceedings 2008*. Ed. by B. Fleischmann, K. Borgwardt, R. Klein, and A. Tuma. Berlin: Springer, pp. 9–14. ISBN: 978-3-642-00141-3. DOI: https://doi.org/10.1007/978-3-642-00142-0_2.

Bogomolov, T. (2013). "Pairs Trading Based on Statistical Variability of the Spread Process". In: *Quantitative Finance* 13.9, pp. 1411–1430. ISSN: 1469-7696. DOI: https://doi.org/10.1080/14697688.2012.748934.

Bowen, D. and M. Hutchinson (2016). "Pairs Trading in the UK Equity Market: Risk and Return". In: *The European Journal of Finance* 22.14, pp. 1363–1387. ISSN: 14664364. DOI: https://doi.org/10.1080/1351847X.2014.953698.

Broussard, J. and M. Vaihekoski (2012). "Profitability of Pairs Trading Strategy in an Illiquid Market with Multiple Share Classes". In: *Journal of International Financial Markets, Institutions and Money* 22.5, pp. 1188–1201. ISSN: 1042-4431. DOI: https://doi.org/10.1016/j.intfin.2012.06.002.

Bruder, B., T. Dao, J. Richard, and T. Roncalli (2011). *Trend Filtering Methods for Momentum Strategies*. DOI: https://doi.org/10.2139/ssrn.2289097. URL: https://ssrn.com/abstract=2289097.

Burgess, A. (2000). "A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage". PhD thesis. University of London. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.6514&rep=rep1&type=pdf.

Caldeira, J. and G. Moura (2013). "Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy". In: *Revista Brasileira de Finanças* 11.1, pp. 49–80. ISSN: 1679-0731. DOI: https://doi.org/10.2139/ssrn.2196391.

Carhart, M. (1997). "On Persistence in Mutual Fund Performance". In: *The Journal of Finance* 52.1, pp. 57–82. ISSN: 00221082. DOI: https://doi.org/10.2307/2329556.

Chen, C., M. Chen, and S. Chen (2014). "Pairs Trading via Three-Regime Threshold Autoregressive GARCH Models". In: *Modeling Dependence in Econometrics*. Ed. by V. Huynh, V. Kreinovich, and S. Sriboonchitta. Cham: Springer International Publishing, pp. 127–140. ISBN: 978-3-319-03394-5. DOI: https://doi.org/10.1007/978-3-319-03395-2_8.

Chen, H., S. Chen, Z. Chen, and F. Li (2019). "Empirical Investigation of an Equity Pairs Trading Strategy". In: *Management Science* 65.1, pp. 370–389. ISSN: 0025-1909. DOI: https://doi.org/10.1287/mnsc.2017.2825.

Clegg, M. and C. Krauss (2018). "Pairs Trading with Partial Cointegration". In: *Quantitative Finance* 18.1, pp. 121–138. ISSN: 1867-6707. DOI: https://doi.org/10.1080/14697688.2017.1370122.

Cui, K. and W. Cui (2012). "Bayesian Markov Regime-Switching Models for Cointegration". In: *Applied Mathematics* 3.12, pp. 1892–1897. ISSN: 2152-7385. DOI: https://doi.org/10.4236/am.2012.312259.

Cummins, M. and A. Bucca (2012). "Quantitative Spread Trading on Crude Oil and Refined Products Markets". In: *Quantitative Finance* 12.12, pp. 1857–1875. ISSN: 1469–7688. DOI: https://doi.org/10.1080/14697688.2012.715749.

D'Aspremont, A. (2011). "Identifying Small Mean-Reverting Portfolios". In: *Quantitative Finance* 11.3, pp. 351–364. ISSN: 1469-7688. DOI: https://doi.org/10.1080/14697688.2010.481634.

Dai, Q., K. Singleton, and W. Yang (2007). "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields". In: *The Review of Financial Studies* 20.5, pp. 1669–1706. ISSN: 0893-9454. DOI: https://doi.org/10.1093/rfs/hhm021.

Do, B. and R. Faff (2010). "Does Simple Pairs Trading Still Work?" In: *Financial Analysts Journal* 66.4, pp. 83–95. ISSN: 1938-3312. DOI: https://doi.org/10.2469/faj.v66.n4.1.

Do, B., R. Faff, and K. Hamza (2006). "A New Approach to Modeling and Estimation for Pairs Trading". In: *Proceedings of the 2006 Financial Management Association European Conference*. Vol. 1. Citeseer, pp. 87–99. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.681&rep=rep1&type=pdf.

Dunis, C. and R. Ho (2005). "Cointegration Portfolios of European Equities for Index Tracking and Market Neutral Strategies". In: *Journal of Asset Management* 6.1, pp. 33–52. ISSN: ZDB-ID 2209717X. DOI: https://doi.org/10.1057/palgrave.jam.2240164.

Dunis, C., J. Laws, and B. Evans (2006). "Trading Futures Spreads: An Application of Correlation and Threshold Filters". In: *Applied Financial Economics* 16.12, pp. 903–914. ISSN: ZDB-ID 1077973-5. DOI: https://doi.org/10.1080/09603100500426432.

– (2008). "Trading Futures Spread Portfolios: Applications of Higher Order and Recurrent Networks". In: *The European Journal of Finance* 14.6, pp. 503–521. ISSN: 14664364. DOI: https://doi.org/10.1080/13518470801890834.

Dunis, C., J. Laws, P. Middleton, and A. Karathanasopoulos (2015). "Trading and Hedging the Corn/Ethanol Crush Spread Using Time-Varying Leverage and Nonlinear Models". In: *The European Journal of Finance* 21.4, pp. 352–375. ISSN: 14664364. DOI: https://doi.org/10.1080/1351847X.2013.830140.

Elliott, R., J. Van Der Hoek, and W. Malcolm (2005). "Pairs TSrading". In: *Quantitative Finance* 5.3, pp. 271–276. ISSN: 1469-7696. DOI: https://doi.org/10.1080/14697680500149370.

Emery, G. and Q. Liu (2002). "An Analysis of the Relationship Between Electricity and Natural-Gas Futures Prices". In: *Journal of Futures Markets* 22.2, pp. 95–122. ISSN: 1096-9934. DOI: https://doi.org/10.1002/fut.2209.

Engelberg, J., P. Gao, and R. Jagannathan (2009). "An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity". In: Singapore. URL: https://ssrn.com/abstract=1330689.

Engle, R. and C. Granger (1987). "Co-Integration and Error Correction: Representation, Estimation, and Testing". In: *Econometrica* 55.2, pp. 251–276. ISSN: 00129682. DOI: https://doi.org/10.2307/1913236.

Fama, E. (1965). "The Behavior of Stock-Market Prices". In: *The Journal of Business* 38.1, pp. 34–105. ISSN: 15375374. DOI: https://doi.org/10.1086/294743.

– (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2, pp. 383–417. ISSN: 15406261. DOI: https://doi.org/10.2307/2325486.

– (1984). "Forward and Spot Exchange Rates". In: *Journal of Monetary Economics* 14.3, pp. 319–338. ISSN: 0304-3932. DOI: https://doi.org/10.1016/0304-3932(84)90046-1.

– (1991). "Efficient Capital Markets: II". In: *The Journal of Finance* 46.5, pp. 1575–1617. ISSN: 15406261. DOI: https://doi.org/10.2307/2328565.

Fama, E., L. Fisher, M. Jensen, and R. Roll (1969). "The Adjustment of Stock Prices to New Information". In: *International Economic Review* 10.1, pp. 1–21. ISSN: 14682354. DOI: https://doi.org/10.2307/2525569.

Fama, E. and K. French (1992). "The Cross-Section of Expected Stock Returns". In: *The Journal of Finance* 47.2, pp. 427–465. ISSN: 00221082. DOI: https://doi.org/10.2307/2329112.

– (1993). "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics* 33.1, pp. 3–56. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(93)90023-5.

– (2000). "Forecasting Profitability and Earnings". In: *The Journal of Business* 73.2, pp. 161–175. ISSN: 15375374. DOI: https://doi.org/10.1086/209638.

Gatev, E., W. Goetzmann, and K. Rouwenhorst (2006). "Pairs Trading: Performance of a Relative-Value Arbitrage Rule". In: *The Review of Financial Studies* 19.3, pp. 797–827. ISSN: 0893-9454. DOI: https://doi.org/10.3386/w7032.

Girma, P. and A. Paulson (1999). “Risk Arbitrage Opportunities in Petroleum Futures Spreads”. In: *Journal of Futures Markets* 19.8, p. 931. ISSN: ZDB-ID 395139-X. DOI: https://doi.org/10.1002/(SICI)1096-9934(199912)19:8<931::AID-FUT5>3.0.CO;2-L.

Guidolin, M. (2011). *Markov Switching Models in Empirical Finance*. Working Papers 415. Innocenzo Gasparini Institute for Economic Research, Bocconi University. DOI: https://doi.org/10.1108/S0731-9053(2011)000027B004.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Vol. 1. 10. New York: Springer Series in Statistics. ISBN: 978-0-387-84857-0.

Hogan, S., R. Jarrow, M. Teo, and M. Warachka (2004). “Testing Market Efficiency Using Statistical Arbitrage with Applications to Momentum and Value Strategies”. In: *Journal of Financial Economics* 73.3, pp. 525–565. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2003.10.004.

Huang, G. and H. Babri (1998). “Upper Bounds on the Number of Hidden Neurons in Feedforward Networks with Arbitrary Bounded Nonlinear Activation Functions”. In: *IEEE Transactions on Neural Networks* 9.1, pp. 224–229. ISSN: 1941-0093. DOI: https://doi.org/10.1109/72.655045.

Huang, G., Q. Zhu, and C. Siew (2004). “Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks”. In: *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks*. Vol. 2, pp. 985–990. DOI: https://doi.org/10.1109/IJCNN.2004.1380068.

Huck, N. (2009). “Pairs Selection and Outranking: An Application to the S&P 100 Index”. In: *European Journal of Operational Research* 196.2, pp. 819–825. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2008.03.025.

– (2010). “Pairs Trading and Outranking: The Multi-Step-Ahead Forecasting Case”. In: *European Journal of Operational Research* 207.3, pp. 1702–1716. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2010.06.043.

– (2013). “The High Sensitivity of Pairs Trading Returns”. In: *Applied Economics Letters* 20.14, pp. 1301–1304. ISSN: 1466-4291. DOI: https://doi.org/10.1080/13504851.2013.802121.

– (2015). “Pairs Trading: Does Volatility Timing Matter?” In: *Applied Economics* 47.57, pp. 6239–6256. ISSN: 0003-6846. DOI: https://doi.org/10.1080/00036846.2015.1068923.

– (2019). “Large Data Sets and Machine Learning: Applications to Statistical Arbitrage”. In: *European Journal of Operational Research* 278.1, pp. 330–342. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2019.04.013.

Huck, N. and K. Afawubo (2015). “Pairs Trading and Selection Methods: Is Cointegration Superior?” In: *Applied Economics* 47.6, pp. 599–613. ISSN: ZDB-ID 280176-0. DOI: https://doi.org/10.1080/00036846.2014.975417.

Jacobs, H. (2015). “What Explains the Dynamics of 100 Anomalies?” In: *Journal of Banking & Finance* 57, pp. 65–85. ISSN: 0378-4266. DOI: https://doi.org/10.1016/j.jbankfin.2015.03.006.

Jacobs, H. and M. Weber (2015). “On the Determinants of Pairs Trading Profitability”. In: *Journal of Financial Markets* 23, pp. 75–97. ISSN: 1386-4181. DOI: https://doi.org/10.1016/j.finmar.2014.12.001.

Jaskowski, M. and D. Van Dijk (2015). “First-Passage-Time in Discrete Time”. In: *8th Annual Society for Financial Econometrics Conference*. Aarhus University, Aarhus.

Jegadeesh, N. and S. Titman (1993). “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency”. In: *The Journal of Finance* 48.1, pp. 65–91. ISSN: 15406261. DOI: https://doi.org/10.2307/2328882.

Johansen, S. (1988). “Statistical Analysis of Cointegration Vectors”. In: *Journal of Economic Dynamics and Control* 12.2, pp. 231–254. ISSN: 0165-1889. DOI: https://doi.org/10.1016/0165-1889(88)90041-3.

Kalaba, R. and L. Tesfatsion (1989). “Time-Varying Linear Regression via Flexible Least Squares”. In: *Computers & Mathematics with Applications* 17.8, pp. 1215–1245. ISSN: 0898-1221. DOI: https://doi.org/10.1016/0898-1221(89)90091-6.

Kara, Y., M. Boyacioglu, and Ö. Baykan (2011). “Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of the Istanbul Stock Exchange”. In: *Expert Systems with Applications* 38.5, pp. 5311–5319. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2010.10.027.

Krauss, C. (2017). “Statistical Arbitrage Pairs Trading Strategies: Review and Outlook”. In: *Journal of Economic Surveys* 31.2, pp. 513–545. ISSN: 1867-6707. DOI: https://doi.org/10.1111/joes.12153.

Krauss, C., X. Do, and N. Huck (2017). *Deep Neural Networks, Gradient-Boosted Trees, Random Forests: Statistical Arbitrage on the S&P 500*. Tech. rep. 2, pp. 689–702. DOI: https://doi.org/10.1016/j.ejor.2016.10.031.

Krauss, C. and J. Stübinger (2017). *Nonlinear Dependence Modeling with Bivariate Copulas: Statistical Arbitrage Pairs Trading on the S&P 100*. Tech. rep. 52, pp. 5352–5369. DOI: https://doi.org/10.1080/00036846.2017.1305097.

Lakonishok, J., A. Shleifer, and R. Vishny (1994). “Contrarian Investment, Extrapolation, and Risk”. In: *The Journal of Finance* 49.5, pp. 1541–1578. ISSN: 1540-6261. DOI: https://doi.org/10.3386/w4360.

Lettau, M. and S. Van Nieuwerburgh (2008). “Reconciling the Return Predictability Evidence”. In: *Review of Financial Studies* 21.4, pp. 1607–1652. ISSN: 0893-9454. DOI: https://doi.org/10.1093/rfs/hhm074.

Li, M., C. Chui, and C. Li (2014). "Is Pairs Trading Profitable on China AH-share Markets?" In: *Applied Economics Letters* 21.16, pp. 1116–1121. ISSN: 1350-4851. DOI: https://doi.org/10.1080/13504851.2014.912030.

Lintilhac, P. and A. Tourin (2017). "Model-Based Pairs Trading in the Bitcoin Markets". In: *Quantitative Finance* 17.5, pp. 703–716. ISSN: 1469-7688. DOI: https://doi.org/10.1080/14697688.2016.1231928.

Liu, B., L. Chang, and H. Geman (2017). "Intraday Pairs Trading Strategies on High Frequency Data: The Case of Oil Companies". In: *Quantitative Finance* 17.1, pp. 87–100. ISSN: 1469-7688. DOI: https://doi.org/10.1080/14697688.2016.1184304.

Liu, D. and L. Zhang (2010). "China Stock Market Regimes Prediction with Artificial Neural Network and Markov Regime Switching". In: *Proceedings of the World Congress on Engineering*. Vol. 1, pp. 378–383. URL: http://www.iaeng.org/publication/WCE2010/WCE2010_pp378-383.pdf.

Liu, S. and C. Chou (2003). "Parities and Spread Trading in Gold and Silver Markets: A Fractional Cointegration Analysis". In: *Applied Financial Economics* 13.12, pp. 899–911. ISSN: 14664305. DOI: https://doi.org/10.1080/0960310032000129626.

Lo, A. and A. MacKinlay (Jan. 1988). "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test". In: *The Review of Financial Studies* 1.1, pp. 41–66. ISSN: 0893-9454. DOI: https://doi.org/10.3386/w2168.

– (2011). *A Non-Random Walk Down Wall Street*. Princeton University Press. ISBN: 978-0691092560.

Lopez De Prado, M. (2018). *Advances in Financial Machine Learning*. 1st ed. Wiley Publishing. Chap. 13. ISBN: 978-1119482086.

Malkiel, B. (2003). "The Efficient Market Hypothesis and Its Critics". In: *The Journal of Economic Perspectives* 17.1, pp. 59–82. ISSN: 0895-3309. DOI: https://doi.org/10.1257/089533003321164958.

Meucci, A. (2009). *Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein-Uhlenbeck*. URL: https://ssrn.com/abstract=1404905.

Montana, G. and F. Parrella (2009). "Data Mining for Algorithmic Asset Management". In: *Data Mining for Business Applications*. Ed. by L. Cao, P. Yu, C. Zhang, and H. Zhang. Boston, Massachusetts: Springer, pp. 283–295. ISBN: 978-0-387-79419-8.

Montana, G., K. Triantafyllopoulos, and T. Tsagaris (2009). "Flexible Least Squares for Temporal Data Mining and Statistical Arbitrage". In: *Expert Systems with Applications* 36.2 (2), pp. 2819–2830. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2008.01.062.

Mori, M. and A. Ziobrowski (2011). "Performance of Pairs Trading Strategy in the U.S. REIT Market". In: *Real Estate Economics* 39.3, pp. 409–428. ISSN: 1540-6229. DOI: https://doi.org/10.1111/j.1540-6229.2010.00302.x.

Nath, P. (2003). *High Frequency Pairs Trading with US Treasury Securities: Risks and Rewards for Hedge Funds*. URL: https://ssrn.com/abstract=565441.

Nóbrega, J. and A. Oliveira (2013). "Improving the Statistical Arbitrage Strategy in Intraday Trading by Combining Extreme Learning Machine and Support Vector Regression with Linear Regression Models". In: *Proceedings of the 2013 IEEE 25th International Conference on Tools with Artificial Intelligence*. Washington, District of Columbia, USA: IEEE Computer Society, pp. 182–188. ISBN: 9781479929719. DOI: https://doi.org/10.1109/ICTAI.2013.36.

Panda, C. and V. Narasimhan (2007). "Forecasting Exchange Rate Better with Artificial Neural Network". In: *Journal of Policy Modeling* 29.2, pp. 227–236. ISSN: 0161-8938. DOI: https://doi.org/10.1016/j.jpolmod.2006.01.005.

Pao, H. (2008). "A Comparison of Neural Network and Multiple Regression Analysis in Modeling Capital Structure". In: *Expert Systems with Applications* 35.3, pp. 720–727. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2007.07.018.

Pástor, L. and R. Stambaugh (2001). "The Equity Premium and Structural Breaks". In: *Journal of Finance* 56.4, pp. 1207–1239. ISSN: 0022-1082. DOI: https://doi.org/10.1111/0022-1082.00365.

Perlin, M. (2007). *M of a Kind: A Multivariate Approach at Pairs Trading*. URL: https://ssrn.com/abstract=952782.

Peters, G., B. Kannan, B. Lasscock, C. Mellen, and S. Godsill (2011). "Bayesian Cointegrated Vector Autoregression Models Incorporating Alpha-Stable Noise for Inter-Day Price Movements via Approximate Bayesian Computation". In: *Bayesian Analysis* 6.4, pp. 755–792. ISSN: 1936-0975. DOI: https://doi.org/10.1214/11-BA628.

Qiu, M., Y. Song, and F. Akagi (2016). "Application of Artificial Neural Network for the Prediction of Stock Market Returns: The Case of the Japanese Stock Market". In: *Chaos, Solitons & Fractals* 85, pp. 1–7. ISSN: 0960-0779. DOI: https://doi.org/10.1016/j.chaos.2016.01.004.

Rad, H., R. Low, and R. Faff (2016). "The Profitability of Pairs Trading Strategies: Distance, Cointegration and Copula Methods". In: *Quantitative Finance* 16.10, pp. 1541–1558. ISSN: 1469-7688. DOI: https://doi.org/10.1080/14697688.2016.1164337.

Ramos-Requena, J., J. Trinidad-Segovia, and M. Sánchez-Granero (2017). "Introducing Hurst Exponent in Pair Trading". In: *Physica A: Statistical Mechanics and its Applications* 488, pp. 39–45. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2017.06.032.

Schaller, H. and S. Van Norden (1997). "Regime Switching in Stock Market Returns". In: *Applied Financial Economics* 7.2, pp. 177–191. ISSN: 14664305. DOI: https://doi.org/10.1080/096031097333745.

Sharpe, W., G. Alexander, and J. Bailey (1999). *Investments*. 6th ed. Upper Saddle River, New Jersey: Prentice Hall. ISBN: 9788120309821.

Shleifer, A. and R. Vishny (1997). “The Limits of Arbitrage”. In: *The Journal of Finance* 52.1, pp. 35–55. ISSN: 15406261. DOI: https://doi.org/10.2307/2329555.

Sims, C. and T. Zha (Mar. 2006). “Were There Regime Switches in U.S. Monetary Policy?” In: *The American Economic Review* 96.1, pp. 54–81. ISSN: 0002-8282. DOI: https://doi.org/10.1257/000282806776157678.

Stephenson, J., B. Vanstone, and T. Hahn (2017). “A Survey of Statistical Arbitrage: Commonalities and Limitations”. Under review.

– (2020). “A Unifying Model for Statistical Arbitrage: Model Assumptions and Empirical Failure”. In: *Computational Economics.* Advance online publication. Reproduced with permission from Springer. ISSN: 09277099. DOI: https://doi.org/10.1007/s10614-020-09980-6.

Stübinger, J. (2019). “Statistical Arbitrage with Optimal Causal Paths on High-Frequency Data of the S&P 500”. In: *Quantitative Finance* 19.6, pp. 921–935. ISSN: 1867-6707. DOI: https://doi.org/10.1080/14697688.2018.1537503.

Stübinger, J. and J. Bredthauer (2017). “Statistical Arbitrage Pairs Trading with High-frequency Data”. In: *International Journal of Economics and Financial Issues* 7.4. ISSN: 2146-4138.

Stübinger, J. and S. Endres (2018). “Pairs Trading with a Mean-Reverting Jump–Diffusion Model on High-Frequency Data”. In: *Quantitative Finance* 18.10, pp. 1735–1751. ISSN: 1867-6707. DOI: https://doi.org/10.1080/14697688.2017.1417624.

Stübinger, J., B. Mangold, and C. Krauss (2018). “Statistical Arbitrage with Vine Copulas”. In: *Quantitative Finance* 18.11, pp. 1831–1849. ISSN: 1867-6707. DOI: https://doi.org/10.1080/14697688.2018.1438642.

Tamura, S. and M. Tateishi (1997). “Capabilities of a Four-Layered Feedforward Neural Network: Four Layers versus Three”. In: *IEEE Transactions on Neural Networks* 8.2, pp. 251–255. ISSN: 1941-0093. DOI: https://doi.org/10.1109/72.557662.

Triantafyllopoulos, K. and G. Montana (2011). “Dynamic Modeling of Mean-Reverting Spreads for Statistical Arbitrage”. In: *Computational Management Science* 8.1, pp. 23–49. ISSN: 1619-6988. DOI: https://doi.org/10.1007/s10287-009-0105-8.

Tsai, C. and J. Wu (2008). “Using Neural Network Ensembles for Bankruptcy Prediction and Credit Scoring”. In: *Expert Systems with Applications* 34.4, pp. 2639–2649. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2007.05.019.

Tu, J. (2010). “Is Regime Switching in Stock Returns Important in Portfolio Decisions?” In: *Management Science* 56.7, pp. 1198–1215. ISSN: 0025-1909. DOI: https://doi.org/10.2139/ssrn.1028445.

Welch, I. and A. Goyal (2008). “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction”. In: *The Review of Financial Studies* 21.4, pp. 1455–1508. ISSN: 0893-9454. DOI: https://doi.org/10.1093/rfs/hhm014.

West, D., S. Dellana, and J. Qian (2005). "Neural Network Ensemble Strategies for Financial Decision Applications". In: *Computers & Operations Research* 32.10, pp. 2543–2559. ISSN: 0305-0548. DOI: https://doi.org/10.1016/j.cor.2004.03.017.

Yang, J., S. Tsai, S. Shyu, and C. Chang (2016). "Pairs Trading: The Performance of a Stochastic Spread Model with Regime Switching–Evidence from the S&P 500". In: *International Review of Economics & Finance* 43, pp. 139–150. ISSN: 1059-0560. DOI: https://doi.org/10.1016/j.iref.2015.10.036.

Ylvisaker, N. (1965). "The Expected Number of Zeros of a Stationary Gaussian Process". In: *The Annals of Mathematical Statistics* 36.3, pp. 1043–1046. ISSN: 00034851. DOI: https://doi.org/10.1215/kjm/1250524817.

Zeng, Z. and C. Lee (2014). "Pairs Trading: Optimal Thresholds and Profitability". In: *Quantitative Finance* 14.11, pp. 1881–1893. ISSN: 1469-7696. DOI: https://doi.org/10.1080/14697688.2014.917806.