BROWN_LEWIS_01593018

by Lewis Brown

Submission date: 10-Sep-2019 02:17PM (UTC+0100)

Submission ID: 110633453

File name: BROWN_LEWIS_01593018.pdf (1.56M)

Word count: 21134

Character count: 102988

A MACHINE LEARNING APPROACH TO CARDINALITY CONSTRAINED PORTFOLIO OPTIMISATION

by

Lewis Brown (CID: 01593018)

Department of Mathematics
Imperial College London
London SW7 2AZ
United Kingdom

Thesis submitted as part of the requirements for the award of the
MSc in Mathematics and Finance, Imperial College London, 2018-2019

IMPERIAL COLLEGE LONDON

THESIS DECLARATION FORM

(for candidates whose degree will be awarded by **Imperial College**)

Declaration: I hereby confirm that the work contained in this thesis is my own work unless other wises stated.

Name LEWIS BROWN **CID** 01593018

Title of Thesis A MACHINE LEARNING APPROACH TO CARDINALITY CONSTRAINED PORTFOLIO OPTIMISATON

Month and year of submission for examination SEPTEMBER 2019

Date 10/09/2019

Signature of Candidate Lewis Brown

Acknowledgements

I would like to thank my internal supervisor Dr. Thomas Cass for his continued support and advice on presenting my work and structuring this thesis. His feedback was indispensable in building my understanding of academic writing and his guidance was helpful in keeping this work as concise as it could be. Further, his suggestions forced me to consider the problems this paper addresses in detail, which was helpful in both addressing them and communicating why it was necessary to do so in this thesis.

I would also like to express my gratitude to my external supervisor Yazid Sharaiha of Norges Bank Investment Management for his advice and guidance throughout the implementation of this project and the subsequent writing of this thesis. Yazid's guidance was essential in keeping the timing of the project on track and the insights he provided me width a wealth of knowledge on the subject and it's applications. He helped immensely in teaching me what was relevant to a financial organisation and ensuring that these areas were the ultimate focus of my work.

Finally, I would like to add a thank you to all of the staff at Norges Bank that I encountered on my visits to their office. All of them were professional, approachable and willing to help me in any way they could.

Abstract

There is a wealth of literature concerning the problem of quantitative portfolio management. This is in part due to the many constraints that asset managers may impose. The purpose of this paper is to consider the application of a realistic cardinality constraint on a portfolio of assets, and present systematic machine-learning based approach to solving the allocation problem under the cardinally constraint.

The algorithm developed in this paper is a custom genetic algorithm which will be employed to perform feature selection from a set of long-short equities strategies to create an improved approach to stock selection. These strategies are build using well-known and easily automated technical indicators. The algorithm presented will autonomously assess the set of strategies and efficiently search for the optimal choice of strategies in terms of their risk-weighted rewards.

A series of experiments illustrate the robustness and usefulness of this approach. A train-test split in the data shows that - in certain market conditions - the approach is effective. From the results presented, it is clear to see that this novel approach has merit and can become a useful tool in passive portfolio management.

Keywords: Cardinality-Constrained Portfolio Optimisation, Machine Learning, Genetic Algorithm

Contents

1 Introduction

1.1 Overview

One of the most well-studied problems in quantitative finance is portfolio allocation. Stating the problem is simple - given a collection of assets, what is the optimal proportion of capital that should be invested in each asset? Answering this question is an important task for all firms in asset management and solving the problem has obvious benefits for these firms and their investors. Given the importance and scope of the task, there are myriad approaches to solving the portfolio allocation problem. They range in their level of mathematical rigour and in the realism of assumptions made.

This paper aims to illustrate how statistical arbitrage strategies can be used to solve this problem when it is formulated with additional constraints relevant to portfolio management. After presenting these strategies mathematically, the objective is to show how they can be automated and discuss areas in which they can be improved. Then this paper will introduce a machine learning approach to feature selection for such an automated strategy that is a combination of the simpler statistical arbitrage strategies. The class of statistical arbitrage based trading strategies is not new - in fact they gained popularity in the 1980s. With modern computing power and greater data availability, automating these processes is very plausible and potentially beneficial. To the knowledge of the author at the date of writing, the literature contains no other attempts to systematically combine the strategies that will be introduced using machine learning.

Strategies that employ statistical arbitrage rely on the following assumption: in the capital markets, some stocks will be under-bought and others over-bought in relation to some benchmark. If these occurrences - called pricing anomalies - can be identified, they can be exploited and leveraged to create profitable investment opportunities for asset managers or any other capital market participants. There are many techniques for detecting anomalies. For example, Morgan Stanley's pairs strategy which became public knowledge in the 1980s is regarded as a pioneer in this field. This is a simple technique that aims to find pairs of correlated assets. The arbitrageur - believing one asset is over-valued and the other under-valued in relation to the fair price for that asset - buys the under-valued asset and shorts the over-valued asset. This is known as buying the spread between the two assets, and the investor holds the position until the spread narrows and they can make a profit.

The main challenge in implementing such a strategy is finding the correlated stocks in the stock market. With pairs, the number of potential stock pairs can be huge. A generalisation is a long-short strategy where stocks are ranked in accordance with a predefined ranking system, before the trader establishes short positions in the highly ranked stocks and long positions in the lower ranked stocks. Clearly the ranking system should infer "how over-bought" an asset is in relation to the market for this strategy to be effective. The definition of over-bought is rather vague though - historically traders rely on informal, in-house observations and to an extent their own instincts on choosing which stocks are in fact over-bought in relation to the market average.

Technical analysis is a field within finance that attempts to formalise the notion of over-bought and under-bought assets. It encompasses a wide, diverse range of mathematical formulae that aim to identify patterns in an asset's behaviour. These formulae can be used trading strategies, as this paper will show, but can complicate automation as the interpretation of their output may be subjective. This problem motivates the work presented in this paper. In building a strategy that uses multiple technical indicators, a reliable heuristic method for optimal feature selection from a set of technical indicators would automate the process of determining which indicators are useful and which are not. Such a method has potential to provide an asset manager with a profitable trading strategy. This paper will present an approach to solving the cardinality-constrained portfolio allocation problem using genetic programming. Further, this paper will describe the features of the presented algorithm in detail and demonstrate that this approach to building a mixed-strategy is both feasible and beneficial.

1.2 Cardinality Constrained Portfolio Optimisation

The formulation of portfolio selection as an optimisation problem is down to the renowned work of Harry Markowitz in the 1950's [14]. This approach is not free of criticism, but it has been used since then until the present day. Observed the following notation. Let N be the total number of assets available to an investor. Then for assets $i, j \in \{1, 2, ..., N\}$ such that $i \neq j$, μ_i is the associated expected return of asset i, σ_{ij} is the covariance between assets i and j, and w_i is the proportion of the portfolio devoted to a position in asset i (or equivalently, it is the portfolio weight for asset i). In addition, let $R^* \in \mathbb{R}$ be an investors desired expected return. The classical problem is formulated in equations (1.1a)-(1.1d).

$$\underset{w \in \mathbb{R}^N}{\text{minimize}} \quad \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \tag{1.1a}$$

$$\text{subject to} \quad \sum_{i=1}^{N} w_i \mu_i = R^* \tag{1.1b}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{1.1c}$$

$$0 \leq w_i \leq 1,\, i = 1, 2, ..., N. \tag{1.1d}$$

This is the unconstrained portfolio problem (UCPP). Equation (1.1a) is the minimisation of the objective function which is the variance of the portfolio. The optimal weights are a vector

$w \in \mathbb{R}^N$ that minimise the variance. Solving this problem is not challenging - it is a simple quadratic program and there are computationally efficient algorithms which can be used to find the optimal solution.

The constraints are given by equations (1.1b)-(1.1d). The first of them, (1.1b), insists that the portfolio attains a desired level of return. (1.1c) forces the weights to sum to 1 and (1.1d) bounds the weights to lie in the interval [0,1].

This framework is renowned within quantitative finance and is well used, but there are limitations to it's applicability. Within asset management it may be necessary to take more complex constraints into account. For example, the portfolio may need to achieve a level of exposure to certain stocks or industrial sectors, or may have to adhere to limitations on transaction costs. Another possible constraint is on the number of assets available that are allowed to enter the portfolio. This is called a carnality constraint.

Define

$$\Pi := \big\{ i \in \{1, 2, \ldots, N\} \text{ such that asset } i \text{ is held by the investor } \big\}.$$

Clearly, in general $\Pi \subseteq \{1, 2..., N\}$ and this becomes an equality in the UCPP. Now define z_i with the mapping

$$\{1, 2, \ldots, N\} \to \{0, 1\} \quad \text{where} \quad i \mapsto z_i = \begin{cases} 1, & \text{if } i \in \Pi \\ 0, & \text{otherwise,} \end{cases}$$

then z_i indicates whether or not the asset i is held in the portfolio. Further, let ϵ_i be the lower bound on the weight of asset i if it is held in the portfolio, and δ_i be the upper bound on the weight for asset i if it is held. Finally, let $K \in \mathbb{Z}$ such that $K < N$.

The Cardinality Constrained Portfolio Allocation Problem (CCPP) is presented in equations (1.2a) - (1.2f).

$$\underset{w \in \mathbb{R}^N}{\text{minimize}} \quad \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \tag{1.2a}$$

$$\text{subject to} \quad \sum_{i=1}^{N} w_i \mu_i = R^* \tag{1.2b}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{1.2c}$$

$$\sum_{i=1}^{N} z_i = K \tag{1.2d}$$

$$\epsilon_i z_i \le w_i \le \delta_i z_i, \; i = 1, 2, ..., N \tag{1.2e}$$

$$0 \le w_i \le 1, i = 1, 2, ..., N. \tag{1.2f}$$

The objective (1.2a) is the previously seen variance minimisation, and constraints (1.2b), (1.2c) and (1.2f) are as they were in the UCPP. In addition, (1.2d) is the cardinality constraint - this ensures card(Π) $< N$. The constraint (1.2e) applies bounds on the portfolio weights of any assets in the portfolio. The effect is that, should asset i be held, the weight must lie in the range $[\epsilon_i, \delta_i]$.

In both UCPP and CCPP, the objective function need not be the variance. The Sharpe ratio - a popular risk-adjusted return metric introduced by Sharpe in [20] - may be used in it's place. Maximising the Sharpe ratio as the objective function is detailed further in [6]. With this alteration, the return constraints (1.1b) in UCPP and (1.2b) in CCPP become redundant and may be ignored. The objective becomes

$$\underset{w \in \mathbb{R}^N}{\text{maximise}} \quad \frac{\sum_{i=1}^{N} w_i \mu_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}}. \tag{1.3}$$

1.3 Finding Feasible Solutions with Long-Short Strategies

A feasible solution to CCPP is any set of weights w_i and stock indicators, z_i that satisfy the constraints (1.2b)-(1.2f). This section presents a long-short strategy as means of finding a feasible set of stocks and weights. A long-short strategy is a trading paradigm that lies within the field of statistical arbitrage. It calculates a real number referred to as a signal for all assets in some collection and uses these signals to create a portfolio. To be consistent with the notation of UCPP and CCPP, the collection of all available stocks is of size N. Before further detailing this stock selection, consider the weights. It suffices to consider equal portfolio weights if the goal is not to find the optimal solution as this is guaranteed to satisfy constraint (1.2c). So only equally weighted solutions are considered which means that

$$\epsilon_i = \delta_i \quad \forall i \in \{1, 2, ..., N\},$$

and the weights are given by

$$\frac{1}{\sum_{i=1}^{N} z_i} = \frac{1}{K},$$

by the definition of z_i. Further, in the formulation of CCPP the constraint (1.2f) reduces to

$$0 \leq w_i \leq \frac{1}{K} z_i, \quad i \in \{1, 2, ..., N, \}.$$

It is straightforward to show that the solution found by applying a long-short strategy is in fact a feasible solution to CCPP. As such strategies are commonly evaluated by their Sharpe ratio, consider the formulation of CCPP that uses (1.3) as the objective function. Clearly the portfolio weights sum to 1,

$$\sum_{i \in \Pi} w_i = \sum_{i \in \Pi} \frac{1}{K} = 1.$$

By simply choosing K stocks at the signal generation stage, the portfolio is forced to obey the cardinality constraint, and it is obvious that weights will lie in the range $[0, 1]$.

The success of a long short strategy completely dependent on the quality of buy/sell signals on which it is based. As this it is a statistical arbitrage strategy, the signal should infer whether or not an asset is over-valued or undervalued by the market. In other words, the signal should identify pricing anomalies. If it does this well, the portfolio ought be perform well with respect to the Sharpe ratio. This approach may also be able to exploit over-valued stocks if short selling is allowed. To do so, at signal generation assets are ordered in descending order of how overbought they are. This is measured by the signal. Then long short strategy to allocate stocks so that the investor opens short positions stocks corresponding the highest ranking and opens long positions in stocks with the lowest rankings. In doing so, they assume that stock prices will ultimately converge to the fair price implied by market.

1.4 Trading Signals from Technical Analysis

Technical Analysis is a field within quantitative investing that aims to measure the historical performance of a stock using mathematical formulae and use this to predict some aspect of the stocks future behaviour. This statement violates the Efficient Market's Hypothesis of [7], which states that no insight into the future performance of a financial assets can be obtained from knowledge of it's previous performance. Nonetheless, empirical evidence justifies the use of technical analysis in finance.

Potential problems with technical analysis include the fact that some of the indicators used may be useless, or perhaps only suitable for certain market conditions. There is not axiomatic mathematical criteria that defines an indicator - many of them were born from the financial sector. A diverse collection of indicators are known in the public domain, and there are likely many more in-house indicators used by financial firms. To use the indicators effectively is usually a skill that is developed by experience. Indicators may have many possible interpretations or interpretations that are subjective.

Often it happens that a trader relies not on a single indicator, but rather on a set of indicators. Some indicator may reveal information regarding trends in the data, for example the Average Directional Indicator suggests the direction of a trend. On the other hand, an indicator like the Bollinger Bands relates to volatility. Both of these signals may be relevant to the trading decision.

To incorporate technical analysis into the signal generation phase of a long-short strategy, the signals will be generated by technical indicators that use historical price-volume data. The data used in called the window. The simplest approach is to used a single indicator to generate the signal. This is a single-signal long-short strategy. A natural progression is to combine the signals generated by several indicators using a weighted sum. With the weighted sum, certain indicators are allowed greater influence. Constructing a portfolio in this manner is a mixed-signal long-short strategy.

1.5 Genetic Programming As An Optimisation Heuristic

Suppose that in constructing a mixed trading strategy, there are a total of M indicators that may be included in a mixed signal strategy, and let M' be the number of strategies that may be used in the mixed strategy. The number of permitted indicators is constrained for two reasons. If all M are used it is likely that the strategy will contain strategies that recommend stocks based on a similar rationale to other strategies. If multiple strategies recommend stocks based on similar mechanisms, then it is then there is nothing to be gained by including both of them - one will suffice. There is also the possibility that some trading strategies may simply not work, perhaps because of the data or they may just be unreliable.

This is a cardinality constraint on the set of permitted indicators used in the mixed-signal strategy. Should the cardinality constraint be too low, or if only a single indicator is used, there is no logic behind building a mixed strategy as there is likely little gain in terms of the information utilised in the trading decision. If it is too large, information overlap is likely, as is correlation between similar indicators. Let $M' < M$. This gives a potentially huge number of possible combinations of indicators. An optimisation heuristic is useful to perform feature selection for the mixed strategy.

Testing all combinations of the M' strategies from a larger M is computationally expensive and time-consuming. Therefore a search heuristic is suggested to parameterise the mixed strategy. This heuristic is a Genetic Algorithm (GA) which is an intelligent search algorithm. Genetic programming is an optimisation technique that draws on Charles Darwin's theory of evolution observed in bio-organisms. Given a set of feasible solutions to an optimisation problem, the iterative application of chosen functions - known as genetic operations - works to improve create improved members of that set by mimicking genealogical concepts. Members of the population are selected for mating and then genetic operations govern the creation of a child (a new member of the population which will replace the weakest current member). The solutions are encoded as chromosomes that characterise the solutions using genes. When mating occurs, the genetic operators distribute the genes from the parents to the child and then ensure diversity in the population by further altering the child's genetic composition. These operations are called crossover and mutation respectively.

If the characteristics of the solutions can be effectively encoded into genes and the mating pool is chosen in a manner that favours the fittest chromosomes and then the desirable features of these solutions - or genes - will propagate through the population. In the context of building a mixed strategy, the presence of an indicator in a weighted sum may be encoded in a binary vector of size M and it's weights are also stored. A large number of possible solutions will form the initial population and then parental selection, crossover and mutation are applied to create a child which replaces the weakest solution in the population. Repeating this cycle of creating a child solution and replacing a weak solution will achieve improved solutions to the CCPP.

2 Literature Review

2.1 Quantitative Portfolio Management

The relevant literature on quantitative portfolio management begins with Markowitz's work in the 1950s. The framework that was used to formulate the problems seen previously was introduced in Markowitz's 1952 paper [14] and was further developed throughout that decade[15]. One key feature of the Markowitz approach to portfolio management is that financial returns are assumed to be mutually independent and normally distributed - an assumption that is not necessarily true in reality but does not detract from the utilisation of the Markowitz framework. The area of quantitative finance that addresses portfolio optimisation under the assumption of normally distributed financial returns is known as Modern Portfolio Theory (MPT).

MPT has been criticised since it's inception by Markowitz, most notably by Mandelbrot and Fama. As early as the 1960's, it was argued by Mandelbrot and Fama that financial returns were better characterised by a Lévy stable distribution [8]. Similar arguments have been made that a Pareto distribution is preferable when modelling financial returns. A shared characteristic of return distributions in PMPT is that they are platykurtic or fat tailed. In the literature, many more papers have analysed portfolio management using different parametric return distributions. Works that do not assume that the returns of financial assets are Gaussian are generally called post-Modern Portfolio Theory (PMPT). Both this field and it's precursor are still relevant today, although PMPT is argued to be more realistic in reality. Stylised facts observed empirically in financial returns can be found in the work of Cont [5]. Another criticism of the Markowitz approach is the use of variance as a risk measure. While there are known problems with the variance the work of Agouram and Lakhnati [11] found that the approach was still profitable under adverse market conditions when compared with other risk measures.

The assumption of mutually independent returns between financial assets is another which is untrue in practise. It is intuitive that returns on different stocks ought to be correlated - in fact it is possible to take the an objective function that maximised the portfolio diversification. For example, the diversification ratio is such a function. Alternatively, the dependence between stocks may be modelled using copulas as is done in the work of Stulajter [22].

In versions of portfolio optimisation where the risk-adjusted return is maximised, the most common metric is the Sharpe ratio [20]. However, other metrics do exist and may also be applied. For instance, the Sortino ratio was introduced in 1994 [1]. When risk minimisation is the objective, measures of risk beyond the variance have also been used. The work of Konno and Yamazaki [12] extended the classical Markowitz framework by using the Mean Absolute Deviation as a metric of portfolio risk.

2.2 Heuristic Methods in Quantitative Finance

There are numerous reasons that heuristic methods are applied in quantitative portfolio management. First, with modern computing power, they can be implemented easily and often converge to a near optimal solution in reasonable time. In the case of genetic algorithms - the heuristic method that is the focus of this paper - it takes very little time to actually implement the search for the optimum. In the case of a GA, the only time consuming part of the optimisation procedure is the computation of fitness values which will be introduced later, and in some circumstances this evaluation can be greatly accelerated by a-piori calculations. Careful selection of the fitness function means that these approaches can be very fast relative to other optimisation algorithms such as Greedy algorithms or grid search. Heuristic methods are often applied in stock selection problems when it is necessary to take realistic constraints into consideration.

While such constraints can often be formulated mathematically, it may be the case that they are non-linear. Chang et. al. [3] illustrated this problem in their analysis of cardinality constrained portfolio optimisation. The CCPP was introduced in [3]. This work illustrates the application of three optimisation heuristics applied to the CCPP - in particular a tabu-search, simulated annealing and a GA. The algorithm for the GA here forms the basis for the approach taken in this analysis. However, this paper differs in that the application of the GA is to perform feature selection for a mixed signal long-short strategy. In contrast, [3] focuses on the optimisation of weights and portfolio selection in CCPP.

The earliest work on GAs for optimisation purposes was that of Holland [9]. The algorithm itself is an attempt to optimise by mimicking the principles of evolution as observed by Charles Darwin. This is reflected in the lexicon that comes with GAs - solutions are represented by chromosomes and parents mate to produce a child. The attraction of a GA is largely due to flexibility. Chromosomes are evaluated with respect to some fitness function which can be defined in the context of the problem at hand - for example, in portfolio management this should relate to the risk adjusted reward of a portfolio. There is additional flexibility in the choice of crossover and mutation operations, and this is illustrated in [2] [16] [19] [18].

More recent work concerning the application of GAs to portfolio optimisation includes that of Jalota and Thankur [10]. In addition to the constraints seen previously, the also consider thresholds on the skewness of the return distributions of stocks that are constituent to the selected portfolio. Also, Srinivasan and Kamalakannan's work [21] applies the GA approach to multiple objective optimisation within financial risk management and provides an example of the use of GAs outside of portfolio allocation.

3 Single-Signal Long/Short Equities Strategies

3.1 The Data: Daily Price-Volume Data for US Listed Equities

The data used here was obtained from Kaggle [13] from the "Huge Stock Market Data" dataset. In it's entirety, the data comprises all stocks and Exchange Traded Funds (ETFs) listed on the NASDAQ, NYSE and NYSE MKT exchanges. Only the stocks are used here.

Each stock daily observation of it's opening price, closing price, daily high price, daily low price and daily volume recorded in a time series indexed by date. The data source states that prices have been adjusted for dividends and splits.

The time series for each stock has been recorded since that stock was listed. Consequentially, some of the stocks only have a few years or even months of data and cannot be meaningfully used in a strategy that can be meaningfully analysed and backtested. The data ends on October 11th 2017. A stock will only be used in any of the following strategies if it has ten years of data recorded and if it belongs to a list of stocks grouped by cluster. For example, Citibank (C) sustained huge drops in it's stock price around the financial crisis of 2008 and will have a disproportionate influence on any trading strategy so will be ignored here. The stock clusters were provided by Norges Bank Investment Management.

There is often a question of whether or not it is appropriate to include the financial crisis in any stock price analysis. The crisis will not be included in the data used here. While it is undoubtedly a significant event, it's inclusion would not be appropriate for the purposes of this work. Further analysis including extreme events in the training data is suggested in the Further Research section of this paper.

In the analysis that immediately follows this section, stocks will be grouped in clusters by the industrial sector to which they belong. The clustering is summarised in Table 1.

3.2 Relative and Absolute Stock Prices

The ranking system for the assets will use technical analysis to uncover pricing anomalies from historical price-volume data. In particular, formulae - called technical indicators - are applied historical price volume data to uncover patterns in the behaviour of an asset. The strength of the stock is ultimately quantified by a real number - or signal - and this signal drives the ranking system. All data used is historical price volume data: this means that at the current time the

Code	Industry	Number of Stocks (N_j)
C1	Oil & Gas Exploration & Production	26
C23	Industrial Machinery	17
C48	Healthcare Equipment	17
C50	Biotech & Pharmaceuticals	16
C55	Department Stores	19
C7	Pipeline Equipment	16
C70	Electric & Multi Utilities	18
C72	Banks	20
C74	Non-life Insurance	18
C81	Retail	19
C96	Software	21
C99	Semiconductors	19

Table 1: Summary of Stock Market Clusters used in Long-Short Strategies

data is deterministic.

Table 1 introduced the data considered here. In particular, there are twelve clusters. Assigning integer labels to these clusters gives the set of clusters $\{(1, 2, ..., N_{Clust}\}$ where $N_{Clust} = 12$ with this data. In each cluster, there are differing numbers of assets - for cluster j the number of assets is denoted with N_j.

The strategy involves applying technical indicators to historical price volume data form each stock a given cluster. To compare performance across clusters, it is convenient to keep track of the cluster to which a given stock belongs. Note that the strategy is applied to price volume data observed at a daily frequency so it is sufficient to view the price-volume data as a discrete-time stochastic prices over the time interval $[0, T]$. In particular, for asset $i \in \{1, 2, ..., N_j\}$ in cluster $j \in \{1, 2, .., N_{Clust}\}$, the price volume process is

$$X^{i,j} := \left(X_t^{i,j}\right)_{t \in [0,T]},$$

where $X_t^{i,j} \in \mathbb{R}^4$. In other words, the state space of the price volume process is $\mathbb{R}^4$. For ease of interpretation, it is helpful to consider this process into four $\mathbb{R}$ valued stochastic processes, as follows

$$X^{i,j} = \left(P^{i,j}, H^{i,j}, L^{i,j}, V^{i,j}\right).$$

The daily close price series observed across $[0, T]$, is

$$P^{i,j} = \left(P_t^{i,j}\right)_{t \in [0,T]}.$$

The remaining terms, corresponding to the daily low, high and volume, analogously generate the processes $L^{i,j}, H^{i,j}$ and $V^{i,j}$.

A technical indicator will be a mapping from a subset of the path of at least one the $\mathbb{R}$-valued series. In general it will be a mapping from a subset of the path of the multidimensional process $X^{i,j}$, but it may not use all four elements in the vector $X_t^{i,j}$. The size of this subset will determine how far into the stocks history the technical indicator looks as it attempts to infer information regarding the strength of the stock. Throughout this paper W will be used to keep track of the size of the subset considered, and the same W will be used at any time $t \in [W, T]$. This is a rolling window and the symbol W will be used interchangeably with the terms "window size" or "window."

An important consideration when applying a technical indicator is that it may be heavily influenced by particularly valuable or liquid stock in the cluster. To reduce this effect, the data to which the indicators are ultimately applied is the path of $\tilde{X}^{i,j}$. This process denoted the relative prices - relative to the cluster-wide average. The transformation is introduced now. First, all stock close prices are rebased to start at \$100 and the high and low prices are shifted to agree with this.

Fix $t \in [1, T]$ and $j \in \{1, 2, ..., N_{\text{Clust}}\}$. For any asset $i \in \{1, 2, ..., N_j\}$, the absolute return on asset i is

$$r_t^{i,j} = \frac{P_t^{i,j} - P_{t-1}^{i,j}}{P_{t-1}^{i,j}}.$$

Next, the mean return across the cluster j is

$$\bar{r}_t^j = \mathbb{E}[r_t^j] = \frac{1}{N_j} \sum_{i \in \{1,...,N_j\}} r_t^{i,j},$$

The following shift essentially demeans the cluster returns observed at time t

$$\tilde{r}_t^{i,j} := r_t^{i,j} - m_t^j.$$

and then the stock prices are recovered though

$$\tilde{P}_t^{i,j} := \tilde{r}_t^{i,j} * P_{t-1}^{i,j} + P_{t-1}^{i,j} \quad \text{for } i \in \{1, 2, ..., N_j\},$$

and $\tilde{P}_0^{i,j} = 100$ for all stocks. These are the relative prices. Applying the same shift to the high and low series yields the transformed series $\tilde{H}_t^{i,j}$ and $\tilde{L}_t^{i,j}$ respectively. The stochastic process $\tilde{X}$ is then given by

$$\tilde{X} = \left(\tilde{P}^{i,j}, \tilde{H}^{i,j}, \tilde{L}^{i,j}, V^{i,j}\right).$$

Note that it is common practise to scale the volume using market capitalisation data. As this data was not available here, the analysis proceeds with no such transformed volume but future work should investigate the effect of this possible extension.

3.3 Design of a Single-Signal Long-Short Strategy

A key determinant of a long-short strategy's success is the signal-generating function. This function drives the rankings system and thus determines the strategies ability to find pricing anomalies to exploit in the capital markets. At any time when a trade can occur, the signal generator is a map from some subset of the path of the process $\tilde{X}^{i,j}$ stopped at the current time t. Note that, at time t this path has been resolved - the data is historic and so there is no stochasticity, only a set of observed values. In other words, the path of the stochastic process $\tilde{X}^{i,j}$ is completely deterministic until the present time t. If a technical indicator is used to generate buy or sell signals, it is sufficient to consider only the path of the deterministic stopped process from time 0 up to an including time t to be a set of observed $\mathbb{R}^4$ vectors given by

$$X^{i,j}_{[0:t]} := \{\tilde{X}^{i,j}_0, \tilde{X}^{i,j}_1, ..., \tilde{X}^{i,j}_{t-1}, \tilde{X}^{i,j}_t\}.$$

In the field of technical analysis, there is debate regarding whether a signal is computed on the day of trade execution or on the day that precedes it. In this paper, the convention will be that the signal is computed on the day of trade execution. Equipped with these observations, a technical indicator is a map

$$g : X^{i,j}_{[0:t]} \to I \subseteq \mathbb{R} \quad x \mapsto \xi^{i,j}_t.$$

where $i \in \{1, 2, ..., N_j\}$, $j \in \{1, 2, ..., N_{Clust}\}$ and t is an admissible day of trade execution in the usual interval. The real number $\xi^{i,j}_t$ is the signal that governs the ranking system.

The goal of the strategy is to construct a portfolio of stocks in a given cluster j at time t, denoted Π^j_t, based on these rankings. To this end, the stocks are ranked in order of their technical signals, where the ranking system (ascending or descending) depends on the nature of the function. In general, a subset of stocks - in the half of the set ranked stocks that corresponds to the overbought stocks - is assigned to the short portfolio. From the other half, an equivalently sized subset is assigned to the long portfolio. Given a ranking system, the collection of ranked assets at time t in cluster j is given by

$$\mathcal{A}^j_t := \Big\{ i \text{ such that } i \in \{1, 2, ..., N_j\} \text{ and } \xi^{1,j}_t \geq \xi^{2,j}_t \geq ... \geq \xi^{N_j,j}_t \Big\}.$$

and then a portfolio is constructed by

$$\Pi^j_t := \{i \in \mathcal{A} \text{ such that } \ i \leq [N_j/K]\} \cup \{j \in \mathcal{A} \text{ such that } j \geq N_j - [N_j/K]\},$$

where [.] is the Iverson bracket. As shown previously, this is a feasible solution to the CCPP. Thus, the weight constraints are satisfied. However, there is no reason to assume that it is optimal or even close.

The long-short strategy is now presented as a tuple below, where the signal generator g is a mapping of the type introduced previously, and both W and δ are positive integers.

$$(g, W, \delta)$$

This tuple fully describes a long short strategy. The remaining parameters are the window size and the rebalance horizon are sufficient to complete the characterisation. The use of a rolling window ensures that the signals are not heavily reliant on features of the data that are likely to be outdated. The integer W truncates the domain of g. In particular, the domain of the generator is refined from $X^{i,j}_{[0:t]}$ to $X^{i,j}_{[0:t]} \setminus X^{i,j}_{[W+1:t]}$. This means that the most recent W observations of the price volume data are used in the calculation of the technical signal.

The final component of the strategy is the rebalance horizon, δ. This is a positive integer that determines how long the portfolio Π^j_t is held before it is liquidated. In technical analysis driven strategies, this is typically a short period of time. In this analysis it will be no longer than ten days. Conventionally, when the current portfolio is liquidated, another is constructed in the same manner and the process will repeat itself. This will be the same protocol adopted here.

The particular type of long-short strategy implemented in this analysis will be a quintile strategy. That means both the long and short portfolios will consist of one fifth of the number of available stocks in a cluster. In the language of the CCPP, this means that the cardinality is constrained by $K = \frac{2}{5}N$.

3.4 Technical Analysis for Cardinality Constrained Portfolios

Recall the previous but general definition of the signal generating mapping g. This function will be a technical indicator in this analysis, which is presented more precisely here as a mapping from the historical price-volume data for an asset into some subset of R. For any stock $i \in \{1, 2, \dots, N_j\}$ in cluster $j \in \{1, 2, \dots, N_{Clust}\}$, a technical indicator is used to generate a signal $\xi^{i,j}_t$ as follows

$$g : W^{i,j}_{[0:t]} \setminus W^{i,j}_{[0:W]} \rightarrow I \subseteq \mathbb{R},$$

$$\{X^{i,j}_{t-W+1}, X^{i,j}_{t-W+2}, \dots, X^{i,j}_t\} \mapsto g(\{X^{i,j}_{t-W+1}, X^{i,j}_{t-W+2}, \dots, X^{i,j}_t\}) = \xi^{i,j}_t$$

A rolling window is applied when implementing the long-short strategy so at any time t the indicator is applied only to the current observation of the price-volume data and the preceding $W-1$ observations. This means that the domain for the technical indicator is not biased towards relatively extreme observations that occurred long before the current time. The real value $\xi^{i,j}_t$ is called the signal. This values is generated in by the technical indicator and - depending on the specific indicator that is used - infers information about the stock. This information may be an estimate of how overbought the stock is in relation to the market average, or it may be information that indirectly allows a conclusions to be made about how overbought the stock is.

With the long-short strategy, the highest ranking must be associated with the overbought stocks. This is essential as these stocks will be assigned to the short portfolio. The ranking system corresponding to each of the indicators used will be made explicit throughout the remainder of this section.

Throughout this section, the following notation applies.

- $\xi_t^{i,j}$ is the signal generated by the technical indicator for stock j in cluster i at time t.
- $\tilde{P}_t^{i,j}$ is close price of stock j in cluster i at time t.
- $\tilde{H}_t^{i,j}$ is daily high of stock j in cluster i at time t.
- $\tilde{L}_t^{i,j}$ is daily low of stock j in cluster i at time t.
- $V_t^{i,j}$ is daily traded volume of stock j in cluster i at time t.
- W is the window size.

3.4.1 Moving Average Indicators

A moving average (MA) is the average of the previous observations in a time series across a specified period. The formula used to compute the average value may be the arithmetic mean - in which case the indicator is a simple moving average(SMA) - or an exponential moving average(EMA) can be used.

Simple Moving Average: SMA

$$\xi_t^{i,j} := \frac{1}{W} \sum_{\tau=1}^{W} \tilde{P}_\tau^{i,j}.$$

Exponential Moving Average

$$\xi_t^{i,j} := k\tilde{P}_t^{i,j} - (1-k)\,\xi_{t-1}^{i,j} + \quad \xi_{t-L+1}^{i,j} = c_0$$

where c_0 is the initialisation parameters and k is the smoothing parameter, chosen to be $2/(W+1)$. This is a recursive definition.

MA indicators are typically trend following, or momentum, indicators. A higher signal is taken as indicative of the underlying stock continuing to perform strongly until the end of the trading horizon. Hence, the long short strategy ranks stocks in descending of MA signals when the trading decision is made. The main difference between the SA and EMA is that the EMA allows more recent values to have a greater influence on the signal, whereas the SMA treats all observations equally.

3.4.2 Volatility Indicators

Volatility indicators are used to identify the stocks that carry the greatest uncertainty. An asset whose price exhibits a high volatility will show a lot of fluctuations around it's average price whereas a less volatile stock will fluctuate less. This can be interpreted a less volatile stock having across stable price across the periods considered in the rolling window, and generally this is taken by technical analysts as an indication that the observer may be certain of the price.

Several volatility focused indicators involve plotting two lines alongside a moving average of the stock price. In such cases, divergence of the lines implies that there is more volatility and convergence is indicative of a reduction in volatility. This is capture by taking the band width as an indicator - higher values imply grater volatility and lower values imply less volatility

Acceleraion Band Width

The upper ABAND is

$$U := \mathrm{SMA}\left(\tilde{H}_t^{i,j}\Big(1+4\,\frac{\tilde{H}_t^{i,j}-\tilde{L}_t^{i,j}}{\tilde{H}_t^{i,j}+\tilde{L}_t^{i,j}}\Big)\right),$$

and the lower ABAND is

$$D := \mathrm{SMA}\left(\tilde{L}_t^{i,j}\Big(1-4\,\frac{\tilde{H}_t^{i,j}-\tilde{L}_t^{i,j}}{\tilde{H}_t^{i,j}+\tilde{L}_t^{i,j}}\Big)\right),$$

where both $u \mapsto \mathrm{SMA}(u)$ is the W-period simple moving average of the series u, U is the upper band and L is the lower band. Generally, the standard simple moving average of the close price is plotted alongside the upper and lower band giving a channel. The indicator itself may be either the upper band value at time t, the lower band value at time t, or the band width at time t. The ABANDS Width is

$$\xi_t^{i,j} = U - D,$$

and this is the indicator used here.

Bollinger Band Width

First define $u \mapsto \sigma_W(u)$ to be the W-period historical standard deviation of some time series u. Then let

$$U = \mathrm{SMA}(\tilde{P}^{i,j}) + y\,\sigma_W(\tilde{P}^{i,j}),$$

be the upper Bollinger bands. Again, the simple moving average has period W and is taken to be the middle band. where M is a simple moving average with window size W. The function σ_W is the standard deviation over an the previous W observations of the close price and the parameter y is a scaling parameter taken to be 2. The lower band is analogously defined with

$$L = \mathrm{SMA}(\tilde{P}^{i,j}) - y\,\sigma_W(\tilde{P}^{i,j}).$$

Again the Band width is the reported indicator, and the signal at time t is given by

$$\xi_t^{i,j} = U - D.$$

In the long/short strategy, stocks are ranked in ascending order of signals generated by band width indicators. In active management, this would be the attitude of a risk averse chartist - the assets that stay close to their average prices enter the long portfolio and those that do not are assigned to the short portfolio.

Volatility (Standard Deviation)

The historical standard deviation in the price series is another technical indicator. This is the is direct quantification of volatility - standard deviation and volatility are interchangeable terms in finance - and this is interpreted as the width based indicators were. In particular, an ascending ranking system is used meaning the high-volatility stocks are believed to be overbought. The formula is simply to compute the W-period historical standard deviation at time t.

$$\xi_t^{i,j} = \sigma_W(\tilde{P}_t^{i,j}),$$

3.4.3 Volume Based Indicators

Volume focused technical indicators aim to summary the supply and demand for or a particular asset in the market. The idea is that highly liquid stocks - i.e. those that have high volume across the rolling window - are more in demand than those that are traded at lower volume. "In-demand" refers to demand in the traditional sense used in economics theory. Economics theorises that there is a fair price for an asset that is achieved when the supply of the asset available to market participants is exactly equal to the demand that exists for the asset. This price is called the equilibrium price. If more market actors are buying the asset, the demand for the asset will increase and ultimately this will either move the equilibrium price upwards or demand will have to because actors will realise that they are overpaying.

Accumulation Distribution

$$\xi_t^{i,j} = \xi_{t-1}^{i,j} + V_t^{i,j} \times \mathrm{CMF}_t^{i,j},$$

where

$$\mathrm{CMF}_t^{i,j} = \frac{(\tilde{P}_t^{i,j} - \tilde{L}_t^{i,j}) - (\tilde{H}_t^{i,j} - \tilde{P}_t^{i,j})}{\tilde{H}_t^{i,j} - \tilde{L}_t^{i,j}}.$$

and

$$\xi_{t-W+1}^{i,j} = 0$$

This is a recursive definition with when the indicator is applied over a window of size W.

Money Flow Index

$$\xi_t^{i,j} = 100 - \frac{100}{1 + \mathrm{MF}_t^{i,j}}, \quad \mathrm{MF}_t^{i,j} = \frac{\mathrm{PMF}_t^{i,j}}{\mathrm{NMF}_t^{i,j}}.$$

Here PMF is the positive money flow over an the window of length W, and NMF is the negative money flow over the same window. The PMF is the sum of the Raw Money Flow (RMF) values in that window on days where the typical price (TP) increases. The RMF and TP are given by

$$\mathrm{TP}_t^{i,j} = \frac{1}{3}(\tilde{L}_t^{i,j} + \tilde{P}_t^{i,j} + \tilde{H}_t^{i,j}), \quad \mathrm{RMF}_t^{i,j} = V_t^{i,j} \times \mathrm{TP}_t^{i,j}.$$

The negative money flow is analogously defined with days on which the typical price experiences a negative change from it's previous value.

On-balance Volume

$$\xi_t^{i,j} := \mathrm{OBV}_t^{i,j} = \sum_{\tau=t-W+1}^{t} V_\tau^{i,j} \mathbb{1}_{\{\text{Price Moves Up}\}} + (-1) * V_\tau^{i,j} \mathbb{1}_{\{\text{Price Moves Down}\}}$$

Price Volume Trend

$$\xi_t^{i,j} := \mathrm{PVT}_t^{i,j} = \frac{\tilde{P}_t^{i,j} - \tilde{P}_{t-L+1}^{i,j}}{\tilde{P}_{t-L+1}^{i,j}} \times V_t^{i,j}$$

For the indicators considered here, high signals imply that the asset is overbought as the price exceeds the price that satisfies the market equilibrium and so has to fall. The belief is that this fall will occur before the portfolio is liquidated. Conversely, assets with low signals are under-bought. The ranking system sorts the assets in descending order of their technical signals. The strategy take long positions in the assets that give low signals (overbought) assets and short the high signal (under-bought) assets. The volume based indicators are presented below. For this class of indicators, the ranking system is inverted.

3.4.4 Convergence-Divergence Indicators

Convergence divergence indicators are often technical indicators that are composed of two other indicators. The simplest case is the moving average convergence divergence which is presented below. This is simply the difference at a given time between the value of two moving averages where the former has a shorter period than the latter. If the former, or fast, moving average is above the value of the slow moving average, then the difference is positive and the averages may be converging or diverging. If the divergence is increasing, it will likely achieve a maximum divergence before converging once again. If the curves are converging, the lines will intersect and the divergence will be negative.

Since the ranking system in the long short strategy considers magnitudes as well as signs when ranking the stocks, stocks which have small positive MACD signals will be ignored by the portfolio so the procedure ignores potential crossovers. The MACD is therefore used as follows. A large positive value in MACD implies that the stocks has been trading well above it's historic average. It may continue to do so, but this is unlikely. The long short strategy will take short positions in stocks that exhibit such behaviour. Stocks that exhibit low negative values will populate the long portfolio following a similar rationale.

The moving averages used may be either simple or exponential. The latter is more sensitive to recent prices than the latter. The general form of a MACD indicator is presented below, and these indicators are fully characterised by the choice of moving average type and the two moving average parameters.

Moving Average Convergence Divergence (MACD)

Let n_{fast} be an additional parameter that is an integer smaller than the window size W, then the MACD is given by

$$\text{MACD}_{n_{\text{fast}},W}(\tilde{P}^{i,j}) = \text{MA}_{n_{\text{fast}}}(\tilde{P}^{i,j}) - \text{MA}_W(\tilde{P}^{i,j}).$$

In other words the MACD is the difference between a fast and a slow moving average of the close-price series. The signal, $\xi_t^{i,j}$ is simply the value of the above series at time t. There are many choices of fast and slow periods that parameterise this indicators. Some common choices combinations are 9, 12, 26, or 52 days. Note that the moving average used may be a simple or exponential moving average. The convention is that the fast and slow moving averages are of the same type.

Percent Volume Oscillator: PVO(n_{fast}, W)

The percentage volume oscillator is another MACD, but now the moving average is applied to the volume series and normalised to give a percentage score. The formula is

$$\text{PVO}_{n_{\text{fast}},W}(V^{i,j}) = \frac{\text{MA}_{n_{\text{fast}}}(V^{i,j}) - \text{MA}_W(V^{i,j})}{\text{MA}_W(V^{i,j})} \times 100.$$

The process for generating the signal $\xi_t^{i,j}$ is again simply the value of the PVO process at time t. The interpretations of the signals sign and magnitude are analogous to previous case. High values now can be interpreted as abnormally high volume and hence increased recent demand. Following a similar rationale to the previously considered volume indicators, the strategy is to short the high signal stocks and long the low signal stocks.

3.4.5 Miscellaneous

The following indicators are common but belong to none or several of the previously considered classes. This section contains some of the most utlised indicators, such as the RSI.

Aroon Oscillator (ARO)

First, define the Aroon up and down curves. They are defined by

$$\text{ARU}_t^{i,j} = 100 \times \frac{W - \max_{\tau\in[t-W+1,t]}\{H_\tau^{i,j}\}}{W}$$

and,

$$\text{ARD}_t^{i,j} = 100 \times \frac{W - \min_{\tau\in[t-W+1,t]}\{H_\tau^{i,j}\}}{W}$$

respectively. Either of these curves may be used as an indicator, but usually the Aroon Oscillator is preferred, as it summarises both curves by reporting the difference between them. The Aroon Oscillator (ARO) is

$$\xi_t^{i,j} := \text{ARO}_t^{i,j} = \text{AR Up}_t^{i,j} - \text{AR Down}_t^{i,j}.$$

The Aroon Up indicator tracks the strength of an uptrend and the Aroon Down tracks the strength of a down trend. The difference between these two indicators is an oscillator that generates signals in the range of -100 to 100. A high score from the Aroon up means that the uptrend was not recent and thus it is likely to be weakening and close to it's end. Hence, such a signal is indicative of an opportunity to profit from a short in such stocks. The Aroon low is high when a downtrend is believed to be coming to an end thanks to a similar reasoning. As such, a high score from the difference between the Up and Down means that there is likely an uptrend coming to an end, and a low value signifies the same for an uptrend. As seen before, the ranking system is inverted and the short portfolio consists of stocks that generate a high Aroon oscillator signal.

Chande Momentum Oscillator

The Chande Momentum Oscillator was introduced by Tushar Chande and it is defined by

$$\xi_t^{i,j} := \text{CMO}_t^{i,j} = \frac{S^U - S^D}{S^U + S^D} \times 100$$

where

$$S^U = \sum_{\tau=t-W+2}^{t} P_\tau^{i,j} \mathbb{1}_{\{P_\tau^{i,j} > P_{\tau-1}^{i,j}\}} \quad \text{and} \quad S^D = \sum_{\tau=t-W+2}^{t} P_\tau^{i,j} \mathbb{1}_{\{P_\tau^{i,j} < P_{\tau-1}^{i,j}\}}$$

S^U and S^D are interpreted as the sum of the close prices on days that the price increased from the previous value, and the sum of the close prices on days that the price decreased from it's previous value respectively. The former is a believed to be a measure of the strength of an uptrend and the latter a measure of downtrend strength. Like the Aroon oscillator, the CMO gives a value in the range of -100 to 100. The ranking system is again inverted with higher values indicative of over-bought stocks and lower values indicative of under-bought assets. Higher values of CMO correspond to longer periods of uptrend in an asset like with the Aroon oscillator. The interpretation of the indicator is very similar, and it would not be surprising if returns generated by strategies that use the CMO or ARO as a signal generator were correlated.

Relative Strength Indicator (RSI)

To define the RSI, it is first necessary to define the relative strength of an asset. To this end, let $\text{RS}_t^{i,j}$ be the relative strength of asset i in cluster j at time t over the rolling window of length W be

$$\text{RS}_t^{i,j} = \frac{\overline{\text{G}}}{\overline{\text{L}}},$$

where

$$\overline{\mathrm{G}} = \frac{1}{W}\sum_{\tau=t-W+2}^{t}(\tilde{P}_{\tau}^{i,j} - \tilde{P}_{\tau-1}^{i,j})\mathbb{1}_{\{\tilde{P}_{\tau}^{i,j} - \tilde{P}_{\tau-1}^{i,j} > 0\}},$$

and

$$\overline{\mathrm{L}} = \frac{1}{W}\sum_{\tau=t-W+2}^{t}(\tilde{P}_{\tau}^{i,j} - \tilde{P}_{\tau-1}^{i,j})\mathbb{1}_{\{\tilde{P}_{\tau}^{i,j} - \tilde{P}_{\tau-1}^{i,j} \leq 0\}}.$$

Then, the RSI normalises the relative strength too give a percentage score between 0 and 100. in particular, using the RSI as the signal generator gives

$$\xi_t^{i,j} := \mathrm{RSI}_{i,j}^t = 100 - \frac{100}{1 - RS_t^{i,j}},$$

where a value of 100 indicates that the price has only been increasing over the period considered and a value of zero indicates that the asset price has only decreased. The RSI is a very popular indicator and it is believed to be over-bought if the RSI score exceed 70 - stocks that generate such a signal are assigned to the short portfolio and under-bought stocks tend to generate and RSI score of 30 or below. High values of RSI are typically indicative for a strong uptrend that is due to end but it may be the case that such values are accounted for by market conditions that are generally bullish.

Stochastic (STOCH))

The stochastic fast (STOCHF or %K) is used to generate the following signal at time t.

$$\xi_t^{i,j} := \%\mathrm{K}_t^{i,j} = \frac{\tilde{P}_t^{i,j} - \min_{\tau\in[t-W+1,t]}\tilde{L}_{\tau}^{i,j}}{\max_{\tau\in[t-W+1,t]}\tilde{H}_{\tau}^{i,j} - \min_{\tau\in[t-W+1,t]}\tilde{L}_{\tau}^{i,j}} \times 100.$$

As usual, this is using a rolling window of size W to compute the signal. This a a special case of the stochastic family of indicators. In the fast, the smoothing parameter that is usually present in a stochastic indicator is set to zero. When the smoothing parameter is non-zero, the indicator is called a full stochastic or %D. The stochastic with smoothing parameter of 3 may be referred to as a Stochastic slow. The general formula for the stochastic indicator with smoothing parameter $n \geq 1$ is

$$\xi_t^{i,j} := \mathrm{STOCH}_t^{i,j} = \mathrm{SMA}(\%\mathrm{K}^{i,j}),$$

which is a simple moving average of the series generated by the stochastic fast where the simple moving average has period n.

Stochastic indicators signify overbought stocks when the signal exceeds 80 and under-bought socks when the signal falls below 20. The indicator suffers from the same ambiguity in it's signals as the RSI and is very similar in terms of it's interpretation. The Williams's %R is an indicator derived from the Stochastic Fast - in fact it is the inverse and it is interpreted in the natural way.

William's %R

The William's %R is the inverse of the Stochastic Fast and another commonly used indicator. It's interpretation is completely analogous to that of the Stochastic fast, although with the inversion values above -20 are now perceived as over-bought and values of less than -80 suggest that the stock is under-bought. The formula is

$$\xi_t^{i,j} := \%\mathrm{R}_t^{i,j} = -100 \times \frac{\max_{\tau \in [t-W+1,t]} \tilde{H}_\tau^{i,j} - \tilde{P}_t^{i,j}}{\max_{\tau \in [t-W+1,t]} \tilde{H}_\tau^{i,j} - \min_{\tau \in [t-W+1,t]} \tilde{L}_\tau^{i,j}} \times 100$$

Momentum (MOM)

This is simply the difference between the observed current price and the price observed at the start of the rolling window.

$$\xi_t^{i,j} := \mathrm{MOM}_t^{i,j} = \tilde{P}_t^{i,j} - \tilde{P}_{t-W+1}^{i,j}$$

It is often the case that high values of momentum is indicative of an uptrend so the assets that cause such behaviour should be bought. However, trial and error suggested that the opposite is true for the rebalance horizons considered here. It may well be the case that the MOM did reveal uptrends but the rebalance horizon was two long to exploit them before they ended.

Rate of Change (ROC)

The rate of change is a scaled version of the momentum. It has a completely analogous interpretation, and the same convention of inverting the ranking system as is done for the momentum indicator is adopted with the rate of change. The formula is

$$\xi_t^{i,j} := \mathrm{ROC}_t^{i,j} = \frac{\tilde{P}_t^{i,j} - \tilde{P}_{t-W+1}^{i,j}}{\tilde{P}_{t-W+1}^{i,j}}$$

Average True Range (ATR)

The true range is the process $\mathrm{TR}^{i,j}$ where each the value taken at time t is given by

$$\mathrm{TR}_t^{i,j} = \max\big((H_t^{i,j} - L_t^{i,j}), (|H_t^{i,j} - P_t^{i,j}|), (|L_t^{i,j} - P_t^{i,j}|)\big)$$

Then the average true range is simply the average at time t taken using a rolling window of length W, and this average gives the signal.

$$\xi_t^{i,j} := \mathrm{ATR}_t^{i,j}(W) = \mathrm{Average}(\mathrm{TR}_t^{i,j}, ..., \mathrm{TR}_{t-W+1}^{i,j})$$

The average taken here is the usual arithmetic mean, but this can be altered. There is little that can be definitively said of what this indicator signifies. The indicator is known to be successful in some cases but views on what inference it makes from the stock data are often subjective. Nonetheless, it is used and the appropriate ranking procedure in the long-short strategy was found - by trial and error - to be such that the short portfolio is built from assets with high ATR scores.

Average Directional Indicator (ADX)

The ADX suffers from the same issues as the ATR in that it has an ambiguous interpretation and is prone to generating evidence. However, when it does work, it can be very useful. It's calculation is detailed below. This requires a sequence of calculations and a a larger window size than the previous calculations. Typical window sizes for this indicator are multiples of 15 less than 60.

First, define the positive and negative directional indicators as

$$+\text{DI} := \frac{\text{MA(+DM)}}{\text{ATR}} \times 100,$$

and

$$\text{-DI} := \frac{\text{MA(-DM)}}{\text{ATR}} \times 100,$$

respectively. The denominator in the above fractions is the average true range as introduced previously. This immediately gives some intuition as to why this indicator suffers from the same issues as the ATR. The numerator terms are called the positive and negative directional movement and they are defined as follows.

$$+\text{DM} := \tilde{H}_t^{i,j} - \tilde{H}_{t-1}^{i,j}$$

$$\text{-DM} := \tilde{L}_{t-1}^{i,j} - \tilde{L}_t^{i,j}$$

In both +DM and -DM, negative values are nullified. Next, the previously introduced terms are used in deriving the directional indicator, given by

$$\text{DX} := \frac{|+\text{DX} - \text{-DX}|}{|+\text{DX} + \text{-DX}|}$$

Finally, the average directional indicator is defined by a moving average to the DX process. This may be done by any smoothing procedure - typically Wilder's smoothing is used but an exponential moving average worked best on the given data so this version was implemented here.

More information on the indicators used here and more can be found at [23].

4 Backtesting and Implementation of Single-Signal Long-Short Strategies

Intuitively a high quality strategy is one that achieves as great a profit as possible given on the investor's initial capital. There is uncertainty regarding the future performance of any position the investor holds so they will take on risk as a consequence of any actions they take. The amount of risk that the investor will be willing to take on is governed by that person or entities risk aversion, and this will vary form one investor to the next. A good trading strategy should achieve a high return and incur low risk. There are many metrics that aim to quantify to what degree a strategy achieves these goals.

Another important consideration with trading strategies is that participation in financial markets is not free. In a long-short strategy rebalances can involve the creating positions in entirely new stocks for the new portfolio and simultaneously liquidating a position in the previous portfolio. The worst case scenario is that this happens with all of the positions in the current portfolio at time t and this behaviors is due to the ranking system. Therefore, the choice of indicator also affects portfolio turnover and consequentially maintenance costs. The turnover is introduced in this section although, for the purposes of this research, the emphasis is on risk-weighted return.

4.1 Returns

Backtesting a trading strategy often involves statistical analysis of the returns distribution. In the case a long-short under the assumption that the portfolio is always leveraged with an initial constant amount of capital C, the return generated is easily obtained from the initial capital and the liquidation values of the portfolio as follows. From the cardinality constraint imposed on the portfolio, it is clear that the positions in each of the assets are

$$S_t^{i,j} := \frac{C}{K},$$

for any asset $i \in \Pi_t^j$ in cluster j at time t. This is independent of the choice of cluster and stock. Note that the positions that compose the portfolio are fixed until the next rebalance date, $t+\delta$. At this rebalance date, the portfolio is liquidates and another is created.

From creation until liquidation, the change in the value of the any position in stock $i \in \Pi_t^j$ is given by

$$\Delta V_t^i = S_t^{i,j}(P_{t+\delta}^{i,j} - P_t^{i,j}),$$

if the strategy is to take a long position in that stock, or

$$\Delta V_t^i = S_t^{i,j}(P_t^{i,j} - P_{t+\delta}^{i,j}),$$

is that strategy is to take on a short position. The overall change in the value of the portfolio is then given by

$$\Delta V_t^{\Pi} := \sum_{i \in \Pi_t^j} \Delta V_t^i$$

and the portfolio return is then simply the quotient of this quantity and the initial cost of opening the portfolio, which is known to be constant C from the construction of the portfolio. In fact, the initial cost may not be exactly C due to rounding but serves as an adequate approximation and makes no real difference to the analysis if returns are taken to mean returns on the amount of available capital rather than returns on the initial investment. Formally, the return on he portfolio in cluster j is defined as

$$r_t^j = \frac{\Delta V_t^{Pi}}{C}.$$

The return in cluster j, r_j, is a random variable. The construction and subsequent liquidation of a portfolio gives multiple observations of this random variable and this constitutes a distribution which can be analysed.

4.2 Backtesting Metrics

Several metrics used in backtesting are simply descriptive statistics of the returns distribution. For example, the mean return, volatility(standard deviation), skewness and kurtosis are often reported. These statistics have the usual interpretation and the formulae are provided for their sample estimates. Throughout, n is the number of time observations. This will vary depending on the window size and rebalance horizon that parameterise the long-short strategy.

$$\bar{\mu}^j = \frac{1}{n}\sum_t r_t^j$$

$$\bar{\sigma}^j = \left(\frac{1}{n-1}\sum_t (r_t^j - \bar{\mu}^j)^2\right)^{\frac{1}{2}}$$

$$\bar{\mathcal{S}}^j = \frac{\frac{1}{n}\sum_t (r_t^j - \bar{\mu}^j)^3}{\left[\frac{1}{n}\sum_t (r_t^j - \bar{\mu}^j)^2\right]^{\frac{3}{2}}}$$

$$\bar{\kappa}^j = \frac{\frac{1}{n}\sum_t (r_t^j - \bar{\mu}^j)^4}{\left[\frac{1}{n}\sum_t (r_t^j - \bar{\mu}^j)^2\right]^{2}}$$

For the returns generated by the strategy in any cluster j, the Sharpe ratio is

$$\text{Sharpe}^j = \frac{\bar{\mu}^j}{\bar{\sigma}^j}.$$

The conventional definition of the Sharpe ratio often has the difference between the mean return and some benchmark return as the numerator. The subtraction is suppressed in this definition. This is because, while in the continuously compounded case it would make sense to consider the return in relation to the risk free rate - in the case of a leveraged portfolio it does not. The initial capital - for the purposes of his analysis - is assumed to be readily available and it's purpose is to be invested in the capital markets. At no point is there a decision on whether to accrue interest on the capital in a risk free setting or invest in the stock-cluster in adherence with some other strategy.

As well as properties of the return distribution, the following metrics will also be acknowledged: the maximum drawdown, periods to recovery, success rate and average turnover. The remainder of this section will define these terms.

First, to the end of defining drawdown, consider the series generated by the cumulative sum of the returns. Note that the returns are additive under the assumptions that come with a leveraged portfolio. In cluster j define

$$\rho_t^j = \sum_{\tau \le t} r_\tau^j,$$

to be the cumulative return and then define the maximum drawdown as

$$D^j := \max_t \{\rho_t^j - \min_{\tau \le t} \rho_\tau^j\}.$$

In other words, the maximum drawdown is the greatest peak-to-trough distance seen in the plot of the cumulative portfolio returns. The periods to recovery gives the number of periods that the portfolio requires to recover from the loss incurred at maximum drawdown and it is defined as follow.

$$\text{PTR}^j := \sum_{t=t'}^{\infty} \mathbb{1}_{\{\rho_t^j < \rho_{t'}^j\}} + 1$$

where t' is the time on which the peak cumulative return introduced with the drawdown was attained, and the final 1 in the above formula is added to signify the first period that the price exceeds the previous peak price. In the case that the portfolio value never reattains it's maximum cumulative value, the periods to recovery is taken to be infinite.

The success rate is simply the proportion of portfolio creation and liquidation cycles that lead to a positive return which is a simple computation. Finally, the turnover - also know as the churn - is defined as follows. Again taking t to be any time on which a rebalance can occur, the turnover on at that date is given by

$$T_t = \sum_{i \in \Pi_t^j} P_t^{i,j} |S_t^{i,j} - S_{t-\delta}^{i,j}|.$$

In particular, this quantity is the sum of the turnovers of each position in the portfolio. For each position, the turnover is the product is the product of the current price of that asset a time t and the change in the quantity of that asset in the portfolio from pre-rebalance to post-rebalance. This difference exhibits one of three possible behaviours. Firstly, it may be the case that asset i is in the portfolio at that is liquidated and the next portfolio that is being simultaneously created. Should this be true, an adjustment in the position needs to take place; at most some stocks are bought or liquidated. Second, it may be the case that a new stock enters the portfolio. If this happens, an entire position has to be liquidated and a new position of a relatively large size has to be created. Obviously, this is a far greater contributor to the turnover than the first possible case.

The mean turnover is the average of all turnover values observed over the time that the strategy is active. As for interpretation, a high turnover is indicative of a trading strategy that would incur high transactions costs in reality. Transaction costs are relevant to any financial institution but these are not the focus of this analysis. It is enough to know that low turnover is preferable to high turnover, and to observe that the turnover is influenced by the parameterisation of the trading strategy.

4.3 Backtest of a Single-Signal Long Short Strategy in One Cluster

This section illustrates the backtesting procedure from a single-signal long-short strategy applied to a cluster of stocks. The stock cluster used is a collection of US corporate equities from the NYSE. With the labelling introduced in Table 1, the data is from cluster C1. The These stocks relate to US Oil and Gas industry firms and they are analysed analysed over a eight year period, from 2009 until 2017. Note that this time window does not include the financial crisis of 2008. Figure 1 shows the close price for the 26 equities in this cluster. They are unlabelled for for

Figure 1: Raw close prices for US corporate equities in the oil and gas cluster.

the sake of presentation but it is clear that the time series for each of the stocks exhibit similar patterns. This is expected as the variables that influence the stock price of each cluster are likely the same for different corporations in the same cluster; for example, it is very likely that commodity prices are very associated with the prices of these firms. While the patterns in the processes are similar, the magnitude does vary from one firm to the other, as does the start price. The transformation from raw prices to relative prices aims to blunt the effect of this on the forthcoming analysis. The transformation is illustrated in figure 2. The strategy is applied to the data illustrated in figure 2. The data seen here are the transformed relative prices. This transformation is may be thought of intuitively as shifting and demeaning the close-prices series in the cluster. The low and high series are transformed to agree with the close price transformation.

Having applied the strategy using the relative prices, portfolio returns are easily obtained and are analysed during backtesting. Recall that the strategy used is completely characterised by three parameters - the choice of indicator, the window size over which the indicator is applied and the rebalance period. These quantities dictate the frequency of the observations and their start date. In this example, the quantities are fixed as stated below, where g is the indicator, W

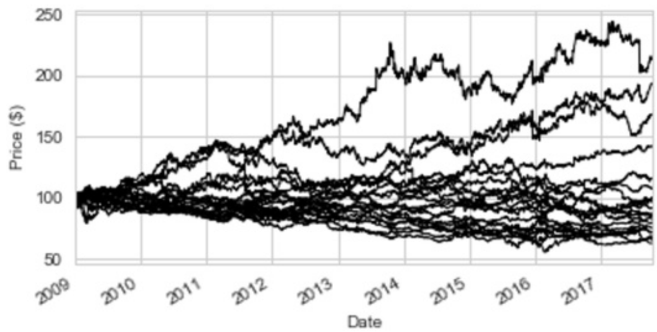

Figure 2: Relative close prices for US corporate equities in the oil and gas cluster.

is the window size and δ is the rebalance horizon.

$$g := \text{RSI}, \quad W = 20, \quad \delta = 3$$

Applying this strategy over the time interval from 01 January 2009 until 11 October 2017 generates the annualised percentage returns shown in Figure 3. Figure 3 illustrates the performance

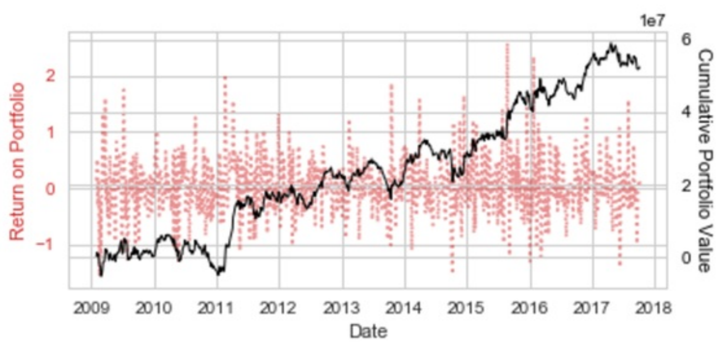

Figure 3: Returns given by applying a RSI based long-short strategy to Oil and Gas US Corporate Equities. The strategy uses 20 day rolling window and portfolios are rebalanced every 3 days.

of the strategy described when applied to the oil and gas cluster. Most importantly, the strategy does achieve a positive value. Over the eight year period, the strategy accumulates approximately \$60 million. Note that the portfolio is assumed to be leveraged and as such there is a large maintenance cost required to generate this value. In this example, the cash injection at each rebalance is \$100 million dollars into both the long and short portfolios, which gives a cost of C =\$200 million, and this cost is incurred at each rebalance date due to the leveraged assumption. The annualised percentage returns, shown in red on the figure, are centered around zero and exhibit positive spikes that are mostly greater in magnitude than the negative spikes.

It is clear in Figure 3 that the RSI based strategy does not perform very well until 2011. For the first few years the strategy returns are quite symmetric around 0 and this is reflected in the portfolio value. It is possible that this is some lasting unpredictability from the crisis in 2008. There may be some merit in analysing the performance of other volatility based indicators over this window to uncover if this is indeed the case, or if the poor performance of this strategy is limited to the particular parameterisation evaluated here.

From the start of 2011 until the end of 2018, the strategy appears to give a more stable and positive performance. There is an upwards jump in portfolio value, and from then until the end of 2017 (until October 11th in particular) the trend is consistently upwards. This suggests that, perhaps, volatility in the markets became more regular as time passed from the crisis, or perhaps that the market in general moved in a positive direction.

To give context to Figure 3, consider the following trends observed in the oil sector between 2009 and 2017. Between 2009 to 2011, the market was bullish. However, in 2011 until 2014 in the prices seemed to plateau around \$100 per barrel. This seems to match what is seen for the strategy. Returns are positive over this time frame, but with the exception of a couple of large increases at the start of 2011, it could be argued that the portfolio sees two systems at work in it's upward cycle.

First, the jump in value from negative to positive may be accounted for by the strategy capitalising on the decay of an upwards trend. The graph shows a period with slightly less return spikes, suggesting that the strategy was able to identify signals but the signals did not correspond to meaningful shifts in the price. Conversely, from the end of 2014 the strategy seems to find more opportunities to exploit. In this period, the market price for oil moved from \$100 to as low as \$20 a barrel. Crucially, there was again volatility in the oil market and this leads to more exploitable opportunities for the RSI based strategy. The rationale behind RSI as a technical indicator is that it uses volatility to identify trends so it is reasonable that the performance picks up in times of high volatility.

The last consideration that ought to be made in evaluating a trading strategy is the cost of maintenance which is described by the turnover. Figure 4 shows the turnover of the strategy across the time window considered here. Each time rebalancing occurs, the behaviour of the strategy is to either recommend liquidating an asset and replacing it in the next portfolio, liquidating and potentially shorting the same assets if it moves from the long to short portfolio or adjusting the amount of an asset held to correspond with the price changes. Of these cases, the latter requires the least amount of capital to be moved. The first case involves a full liquidation

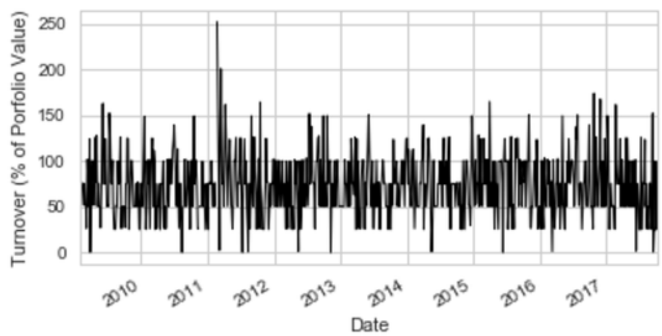

Figure 4: Turnover for the RSI based strategy applied to the oil and gas cluster over the eight year period from 2009 to 2017.

and then purchasing of another stock so becomes expensive and the movement of a stock from long to short is by far the most expensive, as essentially the stock is bought and then sold immediately. The turnover as a percentage of total portfolio value is shown in Figure 4 and is frequently in excess of the 100%. This is high and indicative of a strategy with high maintenance costs. The most notable turnover spike in Figure 4 occurs towards the start of 2011. This occurs

Statistic	Value
Annualised Percentage Returns	
Mean	3.11%
Volatility	4.89%
Skewness	37.39%
Kurtosis	174.37%
Value	
Maximum Drawdown	5.79%
Periods To Recovery	66 days
Success Rate	53.7%
Transaction Costs	
Mean Annualised Turnover (as a percentage of C)	6379%

Table 2: Backtest results for the application of the RSI driven long-short strategy to US oil and gas equities between 2009 and 2017.

in the long portfolio and is likely due to a dramatic change in price of one of the stocks. Further research is needed to confirm exactly what was responsible, bu that is not the focus of this work.

Overall the turnover is quite high for this strategy, but this is reflective of what was observed in most non-trend following indicators. The mean annualised turnover is 6379% as a percentage of the maintenance capital of the strategy.

4.4 Comparison of Efficacy Among Different Strategies in a Cluster of Stocks

It is often the case that while some indicators will perform well in a cluster, others will not. For example, consider the cluster of oil and gas stocks (C1 in the encoding tabulated previously). Figures 5 and 6 show the mean return and volatility respectively - both in annualised percentage terms - of various long short strategies in this cluster. Strategies are stated on the x-axis obeying the convention of stating the indicator first name, any additional parameters, and then the window size.

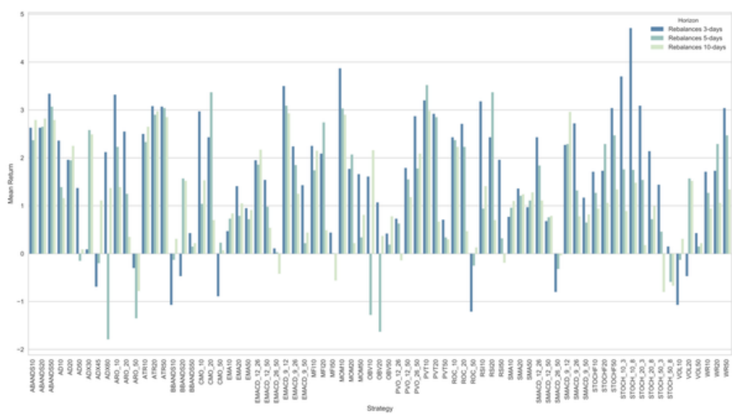

Figure 5: Mean Return in Annualised Percentage terms for Long-Short Strategies in the oil and gas cluster.

It is clear from Figure 5 that it is possible to make a positive return via a long-short strategy in the Oil and Gas cluster. The figures referenced show the outcome of applying a long short strategies that use a variety of technical indicators to drive the ranking system. The strategies were applied over the time window form January 1st 2009 until October 11th 2017. Both the mean return and volatility were estimated using the methods outlined in the previous section.

There are a few particular noteworthy observations. First, notice that the strategies all seem to be more effective with shorter rebalance periods. This is not surprising as most statistical arbitrage strategies are known to behave in this manner. In oil and gas cluster in particular, the stochastic indicator trading strategies mostly give a positive performance. They are less effective

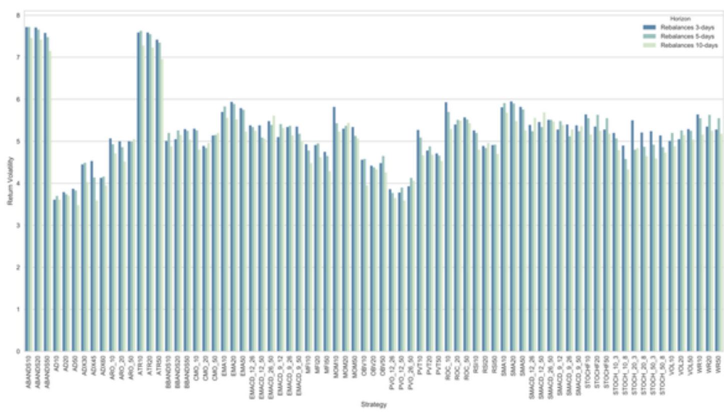

Figure 6: Return Volatility in Annualised Percentage terms for Long-Short Strategies in the oil and gas cluster.

with large window sizes and when the data used is over 50 days the average return becomes negative for stochastic based strategies. Note that the strategy encoded "STOCH_8_50" means that the strategy used a full stochastic indicator as the signal generator, and used a smoothing period of 8 days. This means that the indicator used 58 days worth of historical data and, given the negative output, this easy likely irrelevant.

The volatility plot shows that changing the rebalance horizon or the window size seems to have little effect on volatility. Should the efficacy of a strategy be measured in terms of the Sharpe ratio (as is often the case), then the return will have a greater influence than the volatility. The annualised volatility varies between 3% and 6% for most indicators and the highest volatility is seen in the acceleration bands and average true range. In terms of mean return, these indicators are also similar. The returns from these strategies could be correlated although further analysis would be needed to confirm this.

Another notable feature in these plots is that, while trend following indicators (SMA and EMA) annualised returns of around 1% and annualised volatility of between 5% and 6%, the momentum indicator which is traditionally trend following but worked better empirically as a mean reverting indicator in this analysis, outperforms them both. It achieves a similar volatility but around three times the return of these strategies. This is indicative of the subtleties that can arise with technical analysis. It appear that, at least in this cluster, the momentum may be trend following in very short horizons (at most two days) but after this period it should be treated as mean reverting. The moving averages seems pick up trends that are more stable over longer time horizons.

The most significant observation from the plots in the context of this analysis is that - in a given cluster of stocks - it appears that there are multiple indicators that perform reasonably well in a long-short strategy. This motivates the idea of combining these indicators in an attempt to find an improved signal generating function for a long-short strategy.

The risk-weighted return is quantified by the Sharpe ratio in Figure 7.

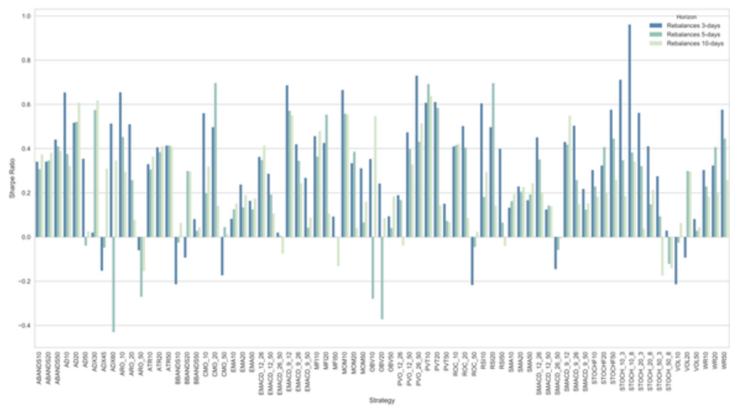

Figure 7: Sharpe ratios for Long-Short Strategies in C1

In asset management, often a Sharpe ratio of 1 is the minimum acceptable value that allows a strategy to be implemented. Although the stochastic based strategies come close to this, none of the strategies on the strategies attain this ratio. This motivates the necessity of improving these strategies. Mixing signals is a natural approach to improving the Sharpe ratio to an acceptable standard and this is the focus of the remaining sections in this paper.

5 Indicator Selection by Genetic Programming for a Mixed Strategy

5.1 Mixed Signal Strategies

The long-short strategies seen previously achieve varying levels of performance. From here forward, they will explicitly be referred to as single-signal strategies. This section aims to investigate whether mixing these strategies will lead to an improved performance. To be precise, the mixture of these strategies will be a long-short strategy where the signal is generated by a weighted sum of the signals generated in each of the single-signal strategies. The problem of choosing which strategies to mix to use will be addressed by a genetic algorithm (GA). This is an intelligent search algorithm that will completely automate the feature selection process for the mixed strategy.

For the single-signal strategies, success is contingent on the ability of the buy/signal's ability to reliably identify assets for which their prices exhibit inconsistency with the fair market price. The technical indicator used may be unreliable in certain situations - for example the ADX is notorious for giving false signals - and some only use a subset of the price-volume data, such as the PVO that only uses volume data. Further, in active portfolio management the trader may use multiple indicators so this mixture approach to mimic the behaviour of a traditional trader. These points serve as motivation for the mixed approach.

To formalise the mixed signal approach, consider a set of M single-signal long-short strategies. For simplification, assume the strategies have a common rebalance horizon. In practise, this constraint on the initial population may be relaxed. Differing rebalance horizons may be included in the population if there is agreement on how to implement the recommended strategy.

Observing the common horizons restraint, each strategy in the set is uniquely defined by an indicator g and a window size $W \in \mathbb{N}$. Note that strategies which use the same indicator are different if they have a different window size and strategies using indicators with an additional parameter are taken to be different from strategies that use the same indicator but have a different extra parameter. For example, a MACD strategy that uses a 9 day fast simple moving average and a 26 day slow moving average is viewed as a different strategy from one which is built from a MACD that compares a 12 day simple moving average and a 26 day slow simple moving average. This makes sense as the information inferred from these two strategies and their effectiveness is different.

A mixed strategy is constructed form a subset of the set of all single-signal strategies. The cardinality of the subset used in mixed-signal generation is constrained by equality. This means that if there are M single-signal strategies, the mixed-signal strategy is a mixture of $M' < M$ single-signal strategies. Let $\mathcal{G}$ denote the set of M strategies - for ease of implementation they will be given an integer encoding. Also let $\mathcal{G}'$ be a subset of $\mathcal{G}$ such that

$$\operatorname{card}(\mathcal{G}') = M' < M.$$

The set $\mathcal{G}'$ gives the strategies whose signal generating functions will be used in the generation of the mixed signal that will determine what action will be taken for a given stock in one of the stock clusters. The approach is the same for all clusters so fix $j \in \{1, 2, ..., N_{Clust}\}$ be fixed and superscript j will be omitted from here forward. Let t be an admissible trading time in t and i be a stock in the chosen cluster. Then, the mixed signal generated for stock i in cluster j is

$$\xi_t^i := \sum_{m=1}^{M} \gamma_t^{m,j} \xi_t^{m,j} \mathbb{1}_{\{g_m \in \mathcal{G}'\}}$$

where $\xi_t^{m,i}$ is the signal from the mth indicator for stock i and $\gamma_t^{m,i}$ is the weight corresponding to that signal. In other words, the mixed signal for a stock i at time t is a weighted sum where contributions are annihilated in the case that the subset of indicators $\mathcal{G}'$ does not include that indicator. The cardinality constraint enforces that there are M' signals exactly in the combination. If and only if the mth indicator in the sum is present in the set $\mathcal{G}'$, then the weight and signal strength is considered for that indicator.

The implementation and backtesting of a mixed-strategy is conducted exactly as previously. It is clear that the diving force behind the strategies is the ranking system and ultimately this is determined by the weights an choice of indicators. Note that the mixed signal long short strategy still represents a feasible solution to CCPP. This is because, as was shown earlier, the following the long-short strategy yields a feasible solution to the CCPP. With mixed signals instead of isolated signals seen previously, the portfolio found by strategy at a given time is still as before, only the signal generation procedure is altered.

5.2 Returns from a Mixed-Signal Strategy

Consider the set of technical indicators corresponding to the strategies in $\mathcal{G}', \mathcal{I}' = \{g_1, g_2, \ldots, g_{M'}\}$. Assume that each technical indicator, g_m for $m \in \{1, 2, \ldots, M'\}$ may be fit over a rolling window of size W_m - the windows need not be the same. More precisely, the domain of the mapping g_m on the admissible trading day t is

$$\{\tilde{X}_{t-W+1}^{i,j}, \tilde{X}_{t-W+2}^{i,j}, \ldots, \tilde{X}_t^{i,j}\} \quad \forall m \in \{1, 2, \ldots, M'\}.$$

Time t is admissible if $t \in [W^{\text{MAX}}, T]$ where

$$W^{\text{MAX}} = \underset{m \in \{1,2,\ldots,M'\}}{\text{maximum}} W_m,$$

meaning that the strategy may only be applied when there is enough observations of price volume data to facilitate the computation of the indicators. The mixed signal on an admissible trading day t is then available for any stock i in some cluster of stocks as

$$\xi_t^i := \sum_{m=1}^{M} \gamma_t^{m,j} \xi_t^{m,j} \mathbb{1}_{\{g_m \in \mathcal{I}'\}}.$$

The strategies in this set also include window sizes and rebalance horizons which are implicitly acknowledged in the formula here.

The returns from a mixed-strategy with this weighting imposed on the signal may be decomposed into a linear combination of the returns of the constituent mixed signal strategies. This approach is detailed further in [17]. With the same weights as those present in the signal-generating sum, the return of a mixed strategy is the random variable

$$r^{\text{Mixed}} = \frac{\sum_{m=1}^{M'} \gamma^m r^m}{\sum_{m=1}^{M'} \gamma^m} = \sum_{m=1}^{M'} \gamma^m r^m,$$

where r^m is the return of the mth single indicator strategy. The final equality is since the weights sum to one. As in the framework of Markowitz, suppose that the returns of the mixed strategy are normally distributed with mean μ^{Mixed} and standard deviation σ^{Mixed} and the returns of the mth single-signal strategy are normally distributed with mean μ^m and standard deviation σ^m. Then

$$\mu^{\text{Mixed}} = E\big[\sum_{m=1}^{M'} \gamma^m r^m\big] = \sum_{m=1}^{M'} \gamma^m \mathbb{E}[r^m] = \sum_{m=1}^{M'} \gamma^m \mu^m$$

and the variance of the mixed return is,

$$\mathbb{V}[r^{\text{Mixed}}] = \mathbb{V}\Big[\sum_{m=1}^{M'} \gamma^m r^m\Big] = \sum_{m=1}^{M'}\sum_{l=1}^{M'} \gamma^m \gamma^l \text{Cov}[r^m t^l] = \sum_{m=1}^{M'}\sum_{l=1}^{M'} \gamma^m \gamma^l \sigma^{m,l}.$$

With the vector of mean returns $\mu \in \mathbb{R}^{M'}$, vector of weights $\gamma \in \mathbb{R}^{M'}$, and covariance matrix $\Sigma \in \mathbb{R}^{M' \times M'}$, the above may be expressed as

$$\mu^{\text{Mixed}} = \gamma^T \mu \quad \text{and} \quad \sigma^{\text{Mixed}} = \sqrt{\gamma^T \Sigma \gamma}.$$

5.3 Presentation of a General Genetic Algorithm

The GA introduced here is a heuristic to select the set of single-indicator strategies that give the best mixed signal ξ_t^i. A skeleton of a GA is presented in Algorithm 1. The genetic algorithm that is actually implemented is customised for the problem at hand, however the algorithm closely follows that of Chang, Meade, Beasley and Sharaiha[3]. In particular, the construction of chromosomes, definition of the mutation operator, and the mechanism of forcing weights to lie within a given range are based on the Algorithm 2 in [3]. However, unlike this algorithm, the GA here is not used to optimise for portfolio but rather a choice of single-signal strategies and the weights representing their relative importance to a mixed signal.

Algorithm 1 Skeleton Genetic Algorithm

```
begin
    Create a population of solutions from the gene-pool.
    Evaluate all solutions in the population using a fitness function, f.
    while IT < MAX do
        Create a new child chromosome using genetic operators.
        Evaluate the child using the fitness function.
        Replace the weakest population member with the child.
        IT ← IT + 1
    end
end
```

The objective of this algorithm is to optimise the choice of strategies that determine the mixed signal. To this end, take the gene pool, denoted with G, to be a collection of single-indicator strategies with a common rebalance horizons. Each constituent single-signal strategy in this set is a gene. If M strategies are considered then the gene pool has cardinality M. A chromosome, $\mathcal{C}$ is a binary vector in $\mathbb{R}^M$ which has exactly M' non-zero elements. If element $m \in \{1, 2, \ldots, M\}$ is non-zero, then the mth strategy in the gene-pool is present in a mixed strategy. This chromosome is part of the solution. The other part of a solution is a weight vector $\mathcal{W}$, again this is a vector in $\mathbb{R}^M$ with M' non-zero elements that match the indices of those in the chromosome. The non-zero elements are now real numbers $\gamma^m \in [\epsilon, \delta]$ where $0 < \epsilon \leq \delta < 1$ which give the weight of signal from the mth strategy in the mixed signal. The pair of a chromosome and a weight vector completely defines a possible solution to the optimisation problem.

To generate a single chromosome, M' integers are sampled from $\{1, 2, \ldots, M\}$. The corresponding indices an M-dimensional null vector are set to 1. The weights are initialised the chosen procedure. The population is then built by repetition of this process until the desired population size is achieved. The population may be different on different initialisations due to the random nature of the process.

5.4 Evaluating Solutions with the Genetic Algorithm

In a general GA, the evaluation stage amounts to applying some fitness function f to a candidate solution. However in this customised GA, there will be two components in the evaluation process. First, the weight constraints will be enforced. The weights supplied to this algorithm will be either randomly initialised or the result of the crossover operation between two parental weight vectors. In either case, they are not guaranteed to respect the fact that they must lie in the range $[\epsilon, \delta]$. The evaluation algorithm therefore ensures that this is the case.

The next stage of the evaluation is the application of f. Here the fitness function will be a weighted sum of two commonly used measures of a trading strategies performance. Note that it is assumed that - when applying this GA - one has access to the backtests for all candidate single-signal strategies. In particular, the returns series is needed to ensure that the evaluation stage is completed quickly.

Algorithm 2 **evaluate**$(\mathcal{S}, \lambda)$

.

$\mathcal{S}$**:** A solution in the population; composed of a chromosome $\mathcal{C}$ and weight vector $\mathcal{W}$.

$\mathcal{C}$**:** A binary vector of dimension M.

$\mathcal{W}$**:** A weight vector in $\mathbb{R}^M$ Indices of non-zero elements match those of the associated $\mathcal{C}$.

G **:** The Gene Pool.

M'**:** The maximum number of genes in a chromosome.

λ **:** The weighting parameter for the fitness function.

$f(\cdot, \lambda)$**:** The fitness function

(ϵ, δ) **:** The lower and upper bound respectively on non-zero elements in $\mathcal{W}$.

begin

  $Q := \{i \in G \text{ such that } c_i > 0 \text{ for } C_i \in \mathcal{C}\}$

  if $\sum_{i \in Q} \epsilon_i > 1$ *or* $\sum_{i \in Q} \delta_i < 1$ **then return**

  compute the current sum $L := \sum_{i \in Q} w_i$

  compute the free proportion $F := 1 - \sum_{i \in Q} \epsilon_i$

  update weights $w_i \leftarrow \epsilon_i + w_i \frac{F}{L}$

  let R be the set of fixed proportions $R := \emptyset$

  while $\exists\, i \in Q \setminus R$ *such that* $w_i > \delta_i$ **do**

    for all $i \in Q \setminus R$ **if** $w_i > \delta_i$ **then** $R := R \cup [i]$

    $L := \sum_{i \in Q \setminus R} w_i$

    $F := 1 - \sum_{i \in Q \setminus R} \epsilon_i + \sum_{i \in R} \delta_i$

    $w_i \leftarrow \epsilon_i + w_i \frac{F}{L} \quad \forall i \in Q \setminus R$

    $w_i \leftarrow \epsilon_i + w_i \frac{F}{L} \quad \forall i \in R$

  end

  return $f(\mathcal{S}, \lambda)$

end

The two measures of strategy performance are the Sharpe ratio and the diversification ratio which is inversely proportional to the between the returns generated by each of the single-indicator strategies in the mixed strategy. The reason for this is, while objective of the CCPP may be formulated in terms of the Sharpe ratio, in practise it may be problematic to use only correlated indicators. There is little benefit in doing so, and this situation may arise if a particular class of indicators has low volatility or high returns. With the previous notation, the Sharpe ratio for the mixed strategy is

$$\text{Sharpe}(r^{\text{Mixed}}) = \frac{\mu^{\text{Mixed}}}{\sigma^{\text{Mixed}}}.$$

The Diversification ratio was developed by Choueitfany and Coingart on 2008 [4] it is given by

$$\mathcal{D}(r^{\text{Mixed}}) := \frac{\gamma^T \text{V}}{\sigma^{\text{Mixed}}},$$

where

$$V = [\sigma^m]_{m=1}^{M'} \in \mathbb{R}^{M'}$$

is the vector of volatilities for the (M') single indicator strategies that compose a mixed strategy. These volatilities are the square roots of the diagonal elements of the covariance matrix. Note that this metric is inversely proportional to the correlation between the returns time series generated by the single-signal long-short strategies. This means that - if two strategies are correlated - there is a higher chance that only one of the two will be present among the highest ranked strategies in terms of their fitness values.

The fitness function against which solutions are evaluated is the weighted sum of these two quantities. Specifically, given any solution to the optimisation problem $\mathcal{S}$, the fitness values is

$$f(\mathcal{S}, \lambda) := \lambda \, \text{Sharpe}(r^{\text{Mixed}}) \; + \; (1 - \lambda) \, \mathcal{D}(r^{Mixed})$$

where $\lambda \in [0, 1]$ is a hyperparameter that dictates the influence of each of the two summands. In the above, a solution $\mathcal{S} \in \{0, 1\}^{M'} \times \mathbb{R}^{M'} \times [0, 1]$ is a collection of a chromosome representation of which vectors are present, a vector of weights that will be forced to satisfy the constraints, and a weighting parameter that governs the balance of the Sharpe ratio of diversification ratio in determining the fitness.

5.5 Genetic Operations

The genetic algorithm optimises a population through the systematic application of two genetic operations - Crossover and Mutation. The motivation is that crossover combines the genes of two chromosomes and produces another chromosome, known as a child. The two chromosomes that combine are the parents. Parents are chosen by a selection method. There are various methods selection procedures, the classical approach being by a binary tournament as is seen in [3]. This is not the method followed here - in this algorithm one parent is chosen as the solution with the greatest fitness in the population and the other is randomly selected from the remainder of the population. The reason for this method is that the fittest solutions presence in the mating pool should encourage the optimal genes to propagate thought the generations, while the randomly selected parent will encourage diversity in the child's genetic composition.

5.6 The Genetic Algorithm in Full

Algorithm 3 Genetic Algorithm for Feature Selection for a Mixed-Signal Long-Short Strategy

$\mathcal{P}$**:** The population - the set of all considered solutions.

$\mathcal{S}$**:** A solution in the population; composed of a chromosome $\mathcal{C}$ and weight vector $\mathcal{W}$.

G**:** The Gene Pool.

M'**:** The maximum number of genes in a chromosome.

λ **:** The weighting parameter for the fitness function.

$f(\cdot, \lambda)$**:** The fitness function

(ϵ, δ) **:** The lower and upper bound respectively on non-zero elements in $\mathcal{W}$.

d **:** The mutation factor.

IT **:** The maximum number of iterations.

begin
- initialise $\mathcal{P} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots \mathcal{S}_{N\_chrom}\}$ as detailed above.
- $\mathcal{F} := \{f(\mathcal{S}, \lambda)\}_{\mathcal{S} \in \mathcal{P}}$
- **for** *($i := 1$ to IT)* **do**
  - $p1 := \text{argmax}(\mathcal{F})$, randomly sample $p2 \in \mathcal{G} \setminus p1$.
  - $\mathcal{S}^* := \mathcal{S}_{p1} = (\mathcal{C}_{p1}, \mathcal{W}_{p1})$. Similarly $\mathcal{S}^{**} := \mathcal{S}_{p2} = (\mathcal{C}_{p2}, \mathcal{W}_{p2})$.
  - Apply crossover operation to $\mathcal{S}^*, \mathcal{S}^{**}$ to create new solution $\mathcal{S}' = (\mathcal{C}', \mathcal{W}')$.
  - $R := \{i \in \mathcal{S}'$ such that $c_i > 0$ for all $c_i \in \mathcal{C}'\}$ are indices for the indicators in the child.
  - $A := (\{c_i \in \mathcal{C}^* | c_i > 0\} \cup \{c_i \in \mathcal{C}^{**} | c_i > 0\}) \setminus R$.
  - **if** *i=1* **then** $w_i := (1-d)(\epsilon_i + w_i) - \epsilon_i \forall i \in R$ **else** $w_i := (1+d)(\epsilon_i + w_i) - \epsilon_i \forall i \in R$;
  - **while** $\text{card}(R) > M'$ **do** remove element j with smallest w_j for R;
  - **while** $\text{card}(R) < M'$ **do**
    - **if** $\text{card}(A) > 0$ **then**
      - add a random asset in A to R
    - **else**
      - add a random asset from $\mathcal{G}$ to R
    - **end**
  - **end**
  - $f' :=$ **evaluate**$(\mathcal{S}', \lambda)$
  - find $j \in \{1, 2, \ldots, M\}$ such that $f(\mathcal{S}_j, \lambda) = \underset{\mathcal{S}_i \in \mathcal{P}}{min} f(\mathcal{S}_i))$
  - $\mathcal{S}_j \leftarrow \mathcal{S}'$ and $\mathcal{F}_j \leftarrow f'$
- **end**

end

The crossover operation is a uniform crossover between the two parents. This means that crossover is determined by a case distinction approach. First, if any gene is present in both parents it is certain to be present in the child. Next, genes which are present in one of the parents and absent in the other are present in the child with probability 0.5. Finally, genes that are absent from both parents are certain to be absent in the child. The weights in the child are determined by an arithmetic mean of the weights in the parents. Note that the number of genes in any chromosome (or solution) s constrained by M'. The GA takes care of this constraint by the following mechanism which is the same as that of [3].

The precise method for enforcing the cardinality constraint on the non-zero elements in C is precisely given in Algorithm 3. Should the crossover yield a child with more genes than is permitted, then the genes with the lowest weights are sequentially removed until the constraint is satisfied. On the other hand, if there are too few genes in the child they are filled by randomly sampled weights from the genes in the parents that the child did not inherit. In the case that this set is depleted, which is unlikely, the genes are sampled randomly from the gene pool until the child has the appropriate number of genes present.

The mutation operator encourages diversity in the population by perturbing the weights in the new child. This corresponds to a scaling of the weights by some mutation factor $d \in (0, 1)$. The nature of the scaling - i.e. whether it is an increase or decrease - is determined randomly. Following the application of the genetic algorithm, the population may be updated in multiple ways. In this case, the child replaces the least fit member of the population. This completes one generational update and this process is applied iterative to the end of finding the optimal solution.

6 GA Constructed Strategy Performance

This section evaluates the performance of the mixed signal strategies that are found by the genetic algorithm. The GA aims to find a near optimal strategy with respect to the previously introduced fitness function.

Results were achieved by running the GA on a training set consisting of one cluster of stock prices observed between January 1st 2009 and December 31st 2014. The GA gives a choice of indicators and weights and these are then then used to parameterise a mixed-signal strategy that was applied to the data from the same cluster in a holdout set consisting of observations fro January 1st 2014 until October 11th 2017. Throughout this section, only three period rebalancing strategies are used in the initial population, but this may be altered and the analysis is the same.

The first feature of the GA analysed is the choice of weight initialisation method in Section 8.1. They are initialised using either a deterministic initialisation scheme where $\xi^m =: \xi = 1/M'$, a discrete uniform distribution over $\{\epsilon, \frac{1}{2}(\epsilon+\delta), \delta\}$, or by a truncated normal distribution centred at the midpoint of $[\epsilon, \delta]$.

Next, Section 8.2 investigates the quality of generalising the GA recommended strategy to the test data for each cluster. This is different from one cluster to the next and reasons as to why there is relatively poor or strong performance in certain clusters are suggested. Section 8.3 finally analyses the optimal number of indicators in some of the strongly performing clusters.

6.1 Influence of Weight Initialisation on Final Weight Recommendations

The following tables show the effect of various weight initialisation methods on the long term relative importance scores (weights in the weighted sum) of the signals used in the mixed strategy. These weights are initialised by one of three methods: either the weights are constant and chosen to be equal or the weights are random and have a specified probability distribution. In both cases, they must sum to 1. In the random case, two choices are illustrated in the last two tables. A discrete uniform distribution over the extreme values and their midpoint, and lastly a truncated shifted standard normal distribution centered at the midpoint and truncated at the extreme points.

To create these tables, the GA was run for 10000 iterations on the training data. The mutation factor was 25% and there were ten indicators allowed to be present in a mixed-strategy. The weights were constrained between 0.02 and 0.20 ($[\epsilon, \delta] = [0.02, 0.20]$). In the case of the equal initialisation, weights were all initialised at 0.10. Table 1 provides a deeper analysis of the behaviour of the weights in this experiment. The initial population consisted of 78 different single-signal long-short strategies with 3-period rebalancing and this is the case for the remainder of this paper. The mixed strategy was then backtested in the test data.

Table 2 shows that when the Sharpe ratio is used as the fitness function and the diversification ratio is completely ignored ($\lambda = 1$), the weights that are eventually achieved by GA are close to being equal for seven of the strategies. In particular, seven of the weights lie in the range $[0.99, 1.03]$. There is less importance assigned to the AD and WR strategies and more to the EMACD(12,26). The strategy in this case generalises well to test data - the decrease from the in-sample to out-of-sample Sharpe ratio is about 30%. This similar to the figures seen for the mixed fitness function and better than when only diversification is used in the fitness function. The case of the mixed fitness function also causes the most diversity in the weights, with only two of them being close to what would be seen in the initial equally weighted mixtures.

Under normal initialisation, shown in Table 4, the optimal weights vector shows more dissimilarity than seen in Table 3 when the Sharpe ration is included in the fitness. With a pure Sharpe fitness function, six weights were close to equal with the equal initialisation seen in Table 2. Now,

	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0$
Indicators		**Weights**	
(AD, 10, 3)	-	0.067	-
(ADX, 30, 3)	-	-	0.107
(ADX, 45, 3)	-	-	0.089
(AR), 20, 3)	0.087	-	
(ATR, 50, 3)	0.101	-	
(BBANDS, 20, 3)	-	-	0.111
(CMO, 20, 3)	-	-	0.079
(MOM, 20, 3)	0.100	-	-
(PVT), 20, 3)	-	0.113	-
(PVT), 50, 3)	-	0.070	-
(VOL), 20, 3)	-	0.075	0.113
(VOL), 50, 3)	-	0.113	-
(WR, 10, 3)	0.100	-	-
(WR, 50, 3)	0.091	-	-
(EMACD(9), 12, 3)	0.101	0.112	-
(EMACD(12), 26, 3)	0.117	0.114	-
(EMACD(12), 50, 3)	-	-	0.094
(SMACD(9), 26, 3)	-	-	0.078
(PVO(9), 12, 3)	-	-	0.104
(PVO(9), 26, 3)	-	-	0.112
(PVO(26), 50, 3)	0.103	0.111	0.0112
(STOCH(3), 12, 3)	0.101	0.113	-
(STOCH(3), 52, 3)	0.099	0.112	-
In Sample Perfromance of GA Constructed Strategy			
Mean Return	3.110	2.582	0.625
Volatility	1.829	1.498	1.240
Sharpe Ratio	1.685	1.724	0.504
Out of Sample Performance Backtest Results			
Mean Return	4.126	3.241	1.713
Volatility	3.010	2.625	2.505
Sharpe Ratio	1.371	1.235	0.684

Table 3: Weights and strategies present for the optimal solution found by the GA when weights are initialised using deterministic equal weights.

	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0$
Indicators		**Weights**	
(ABANDS, 20, 3)	-	-	0.100
(AD, 10, 3)	-	0.108	-
(ADX, 20, 3)	-	-	0.099
(ADX, 30, 3)	-	-	0.099
(ARO, 45, 3)	-	-	0.098
(ATR, 10, 3)	-	0.069	-
(BBANDS, 50, 3)	0.113	0.124	
(MOM, 20, 3)	0.066	-	-
(OBV, 10, 3)	0.103	-	-
(OBV, 20, 3)	0.106	0.101	-
(SMA, 50, 3)	-	-	0.096
(STOCHF, 10, 3)	-	0.093	-
(WR, 50, 3)	-	0.081	-
(SMACD(9), 50, 3)	0.066	-	-
(SMACD(26), 50, 3)	-	-	0.090
(EMACD(9), 12, 3)	0.114	0.096	-
(EMACD(9), 26, 3)	-	0.117	0.101
(EMACD(12), 26, 3)	0.128	-	-
(EMACD(26), 50, 3)	-	0.093	-
(PVO(9), 12, 3)	-	-	0.101
(PVO(12), 26, 3)	-	-	0.116
(PVO(26), 50, 3)	0.110	0.121	0.100
(STOCH(3), 12, 3)	0.097	-	-
(STOCH(3), 52, 3)	0.098	-	-
In Sample Perfromance of GA Constructed Strategy			
Mean Return	3.008	2.700	1.046
Volatility	1.668	1.554	1.310
Sharpe Ratio	1.792	1.737	0.623
Out of Sample Performance Backtest Results			
Mean Return	3.598	3.264	1.382
Volatility	2.844	2.645	2.410
Sharpe Ratio	1.265	1.234	0.574

Table 4: Weights and strategies present for the optimal solution found by the GA when weights are initialised using truncated normal distribution.

	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0$
Indicators		**Weights**	
(AD, 10, 3)	0.077	0.112	-
(ADX, 30, 3)	-	-	0.087
(ADX, 60, 3)	-	-	0.094
(ATR, 10, 3)	0.074	0.089	-
(BBANDS, 20, 3)	-	-	0.099
(BBANDS, 50, 3)	0.104	0.113	0.138
(MOM, 20, 3)	0.118	-	-
(OBV, 10, 3)	-	0.076	-
(PVT, 20, 3)	-	-	0.090
(ROC, 20, 3)	-	0.106	-
(STOCHF, 50, 3)	0.078	-	-
(WR, 10, 3)	0.121	-	-
(EMACD(9), 12, 3)	0.078	0.107	-
(EMACD(9), 26, 3)	-	0.107	-
(EMACD(12), 26, 3)	0.130	-	-
(EMACD(12), 50, 3)	-	-	0.101
(SMACD(9), 26, 3)	-	-	0.100
(SMACD(9), 50, 3)	-	0.099	-
(SMACD(12), 50, 3)	0.103	-	-
(PVO(9), 26, 3)	-	-	0.096
(PVO(12), 26, 3)	-	0.078	-
(PVO(26), 50, 3)	0.118	0.112	0.100
(STOCH(8), 18, 3)	-	-	0.096
In Sample Perfromance of GA Constructed Strategy			
Mean Return	2.756	2.368	1.085
Volatility	1.557	1.438	1.334
Sharpe Ratio	1.750	1.647	0.813
Out of Sample Performance Backtest Results			
Mean Return	3.642	2.675	2.144
Volatility	2.673	2.442	2.437
Sharpe Ratio	1.363	1.095	0.880

Table 5: Weights and strategies present for the optimal solution found by the GA when weights are initialised using discrete uniform distribution.

only the stochastic strategies are close to equally weighted under normal initialisation. Note that with the diversification ratio only in the fitness function, the GA with normal initialisation converges to seven near equal weights. What is notable from the two tables though is that while there is an overlap in some of the optimal indicators and a a very similar in-sample performance, the different weight initialisation leads to a slightly different choice of optimal strategies. In particular, the ABANDS(20) strategy - acceleration bands applied over a window of size 20 - was only found in the optimal collection under normal initialisation, and not for either of the other choices. However, both weight initialisation methods seem to give similar MACD-type strategies in the optimal collection. Despite differences in the recommended strategy, the out-of-sample performance is similar to the equal weights case.

The discrete uniform initialisation method, shown in Table 5, also gives diversity in the weights. In the case of a pure Sharpe fitness function, it also achieves the best in-sample Sharpe and out-of-sample Sharpe ratios, and the smallest drop in decrease in Sharpe ratio when generalising to the test data. The decrease is about 22%. The performance for the mixed fitness is lower than seen previously. For the pure diversification fitness, the out-of-sample Sharpe is greater than the in-sample Sharpe.

These tables suggest that the choice of weight initialisation does have an effect on the optimal solution found by the GA. However, with enough single-signal strategies in the population, there is a possibility that similarly performing mixed-signal strategies may be achieved by different choices of weights and parameters. While the weight initialisation does have an effect, it is not the sole determinant of a strategies ultimate feasibility. With enough single-signal strategies, the the optimal chromosome may contain different single-signal strategies and achieve similar results. This is likely down to these strategies achieving similar results when applied to the train data. This is not an immediate cause for concern. If the user of the algorithm does deem this behaviour as unacceptable, it may be overcome by imposing a tighter constraint on the initial population of strategies. For example, they may only indicate a single SMACD, or a penalty may be incorporated into the fitness function to penalise inidicators from the same class.

6.2 Generalisation to Test Data In Different Clusters

This section aims to investigate if the GA constructed mixture strategies conform to this behaviour, and how well these strategies generalise to the test data. Note that, by the random nature of the GA, the results will not be the same each time.

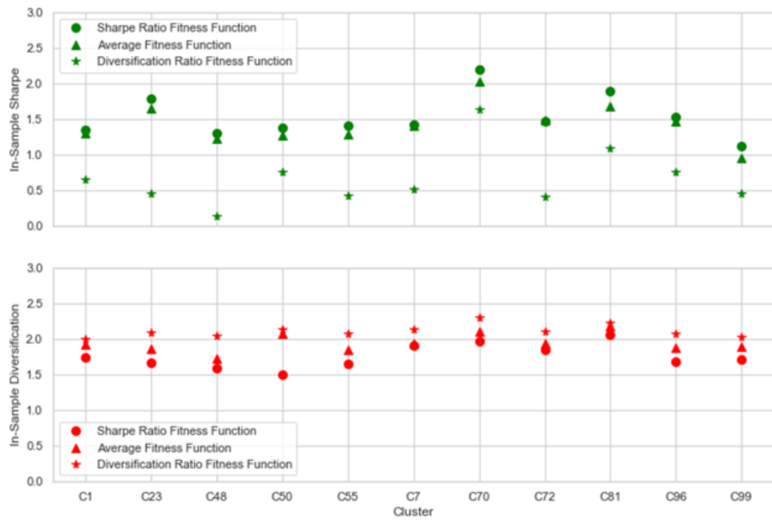

Figure 8: In-sample performance of the GA with 3-period horizon single-signal strategies in the initial population

Figure 8 presents the within sample performance of the GA in the top plot and the out-of-sample performance on the lower plot. For each cluster on the x-axis, the GA is run on the train data and then the strategy it returns is applied to the test data. The weights are initialised using an shifted standard normal distribution as detailed above. The experiment may be repeated for any of the other initialisation methods introduced. In each chromosome, ten genes are allowed to be present. This means that the mixed-strategy that the GA constructs is allowed to consist of exactly 10 unique indicator and window pairs. The weights are confined to the range [0.02,0.20]. The mutation factor was chosen to be 25% as it is throughout this paper.

The upper most plot shows the Sharpe ratio computed from backtesting the strategy on the data that was used to create it. Unsurprisingly, if the fitness function includes the Sharpe ratio at all, the mixed strategy achieves a promising Sharpe ratio. In all but one cases where the fitness function contains the Sharpe ratio, the in sample Sharpe is in excess of 1.0. In general, a strategy that achieves a Sharpe ratio in excess of 1.0 is seen as is viewed favourably. When the fitness is taken as the portfolio diversification, there is a drop in Sharpe ratio for all clusters. In C81 - the Retail Cluster - a strategy build using a diversification fitness function still achieves an in Sample Sharpe ratio greater than 1.0, but does still lose performance compared with the other fitness funcitons. This drop in performance is replicated other clusters too, most notably in the Healthcare equipment cluster - C48. Given the sector, this is plausible. It is likely that healthcare returns are correlated, which may lead to correlated returns from the strategies applied to this cluster. Intuitively, the task of diversifying while staying within the healthcare sector seems quite challenging.

The lower plot shows the diversification ratio of the mixed strategy found by the GA using the same three fitness funcitons as above. In all cases, diversification does not diminish as much as the Sharpe ratio when the fitness is changed. This is possibly due to the construction of the two ratios that compose the fitness function. Both have the same denominator, but the diversification completely discounts return. The behaviour in both these plots is suggestive that, in-sample, a weighted combination of the two rations can be useful in achieving a high return and reasonably uncorrelated selection of inidcators and weights to build a mixed strategy. However, it is more beneficial to have more weight assigned to the Sharpe component as this still achieves good dissimilarity scores.

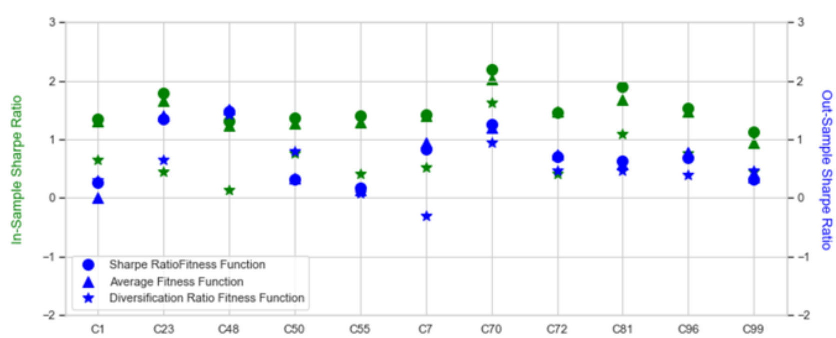

Figure 9: In and Out-sample Sharpe ration for all clusters using training set of data from 2009-2014 and test set from 2014 to 2017

Figure 9 compares the in-sample(green) Sharpe ratio for each of clusters with the out-of-sample Sharpe ratio. The fitness funcitons used in the GA are as seen in figure 8. Three clusters perform particularly well: industrial machinery, healthcare, and electric and multi utilities. A possible reason is each of these clusters exhibited relatively low volatility between 2009 and 2018. Healthcare supplies - in the US - have a relatively stable customer base and likely benefited from ObamaCare in 2009. Industrial Machinery is a relatively stable industrial sector and electric and multi utilities (C70) has seen an upwards trend across the cluster and low volatility. The Dow Jones average - included in the stocks used here - is the most significant stock in C70. It has been remarkably stable over the period considered and consistently in uptrend.

The clusters with the worst out-of-sample performance are oil and gas (C1), and department stores (C50). For oil and Gas, the cluster was likely very inconsistent in terms of drift direction over the training window. Taking WTI (West Texas Index) as a proxy for Oil prices, the price was bullish, then plateaued and the appeared bullish in the training window. The GA would essentially be learning what indicators were important based on market constitutions that are characterised by three distinct time-periods: the lack of generalisation is not surprising. Retail was also likely very volatile in the training. The retail industry is largely driven by consumer spending. The train period began following the financial crisis of 2008 and it is likely that spending changed a lot in the five years following the crisis.

One final observation is that, with biotech and pharmaceuticals (C50) in the US. The out of sample performance is best for the diversification ratio GA. In this cluster, it is likely that customer sentiment switched between training and testing regarding US pharmaceuticals. Two significant events occurred - ObamaCare in 2014 and the first success lawsuit for deliberately inflating drug prices in the US in 2016. This illustrates the usefulness of diversification in a the selection of indicators: the diversification ration outperforms any strategy that is built with a returns influenced fitness function.

The main conclusion to be drawn from this analysis is that it is often possible to achieve a strong in-sample performance using the GA constructed strategy. However, the applying the strategy to the test data is not guaranteed to achieve similar performance. The performance of the strategy in the test set is often determined by the nature of the industrial cluster in question in this time period.

6.3 Finding an Optimal Number of Indicators

In clusters where the GA works well both in-sample and out-of-sample, it is possible that there are potential gains to be made form using more or less indicators in the mixed strategy than the ten considered previously. Similarly, more or less indicators may lead to feasible strategies in clusters where the GA was not successful in finding a profitable mixed indicator strategy.

Figure 10 compares the in and out of sample performance of the mixed-strategies that use different number of indicators in the industrial machinery cluster (C23). In other words, the number of genes that are allowed to exist in a chromosome is varied - where previously it was

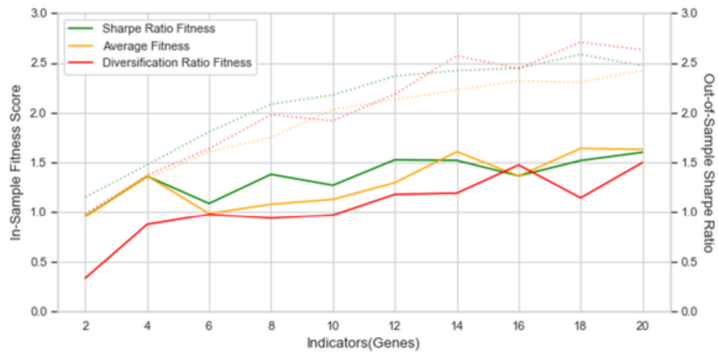

Figure 10: In and Out of Sample performance for varying number of indicators in a mixed signal strategy in the cluster of industrial machinery equities

10, it is now any number in the range 2,4,...,20. The in sample-fitness score is visible form the broken lines and the solid lines are the out of sample Sharpe rations. The analysis in figure 9 is using the industrial machinery cluster - the GA previously performed well on this data.

The strategy performs consistently well with fitness funcitons that include for all choices of the number of inidcators. The out of sample Sharpe is (almost) always over 1.0 and the performance appears to be maximised at 14 indicators. In this situation, the average fitness function is the optimal choice, but it only slightly outperforms the pure Sharpe ratio. The diversification ratio only fitness function gives worse performance for tighter gene constraints but achieves strong performance for 12, 14, and 16 genes. This curve attains it's maximum at 16 indicators. Regardless of the fitness function chosen, it appears that, in the industrial machinery cluster, between 12 and 16 gives the strongest performing strategies. Note that due to randomness, the exact locations of the maxima may differ which is why the range should be considered and not the single maximising value.

The next figure, Figure 11, recreates this plot using data for the semiconductors cluster. From the previous section, when ten indicators were used the mixed strategy did not generalise well to the test set in the semiconductors sector. In Figure 11, the out-of-sample Sharpe ration only stabilises above zero when 12 inidcators or more are used. For smaller number of indicators, the pure diversification fitness function outperforms the return weighted alternatives as has been the case previously. There is a possibility that, in volatile clusters, the diversification ratio may be a better choice of fitness function in certain clusters of stocks. In the semiconductrors sector, one of the major stocks - Qualcomm - exhibited change in behaviour between 2016 and 2018. After many years of a consistent upward drift, the price started to plateau. On the other hand, many

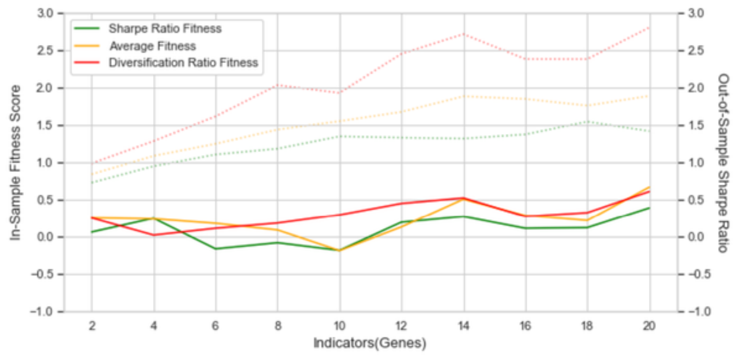

Figure 11: In and Out of Sample performance for varying number of indicators in a mixed signal strategy in the cluster of semiconductor equities.

other stock - like Intel - saw dram tic price increases across the same time frame. Regardless of the direction, it seems a lot of the major stocks in the cluster saw some change of systems in the the time frame used by the test data. This could suggest that, if a cluster is likely to see significant changes in trends, the diversification ratio may be the most suitable choice of the fitness function. This is not a proven statement though, rather a potential future research question.

The results presented in this section also suggest that the optimal number of indicators in different clusters will not be the same. Future research may build on the current GA approach to also find the optimal number of indicators in the cluster on which it is applied.

7 Further Research

While the GA approach has been successful in some experiments, there is opportunity for improvement in others. As hinted previously, many of the issues could be resolved with more data. For example, volume indicators may perform better if market capitalisation is used to scale the volume observations in the data. This would prevent high volume stocks from dominating the strategies.

Further, in this analysis the hyperparameters of the GA have been fixed and chosen subjectively. With more data and more computing power, it is possible that future work could employ a train, validation and holdout split instead of the train and holdout split employed here. Hyperparameters of the GA include the number of training iterations, the number of genes, whether the maximum gene count is constrained by inequality or equality, the choice of fitness function and the bounds on the weights. With the three part split the GA would be trained for numerous choices of parameters and the best collection may then be chosen for application to the hold-out set. This appraoch would be computationally costly but may yield improved out-of-sample performance.

Aside from increasing the train data, it is possible to better utilise that which is already there. One possible approach is to employ an expanding train-test split. A methods for doing this is as follows. First perform the single-signal backtests on the train data as usual and run the GA to achieve a set of inidicators that are optimal at the start of the test data. Then, at specified intervals, backtests could be carried out again and the GA could be re-trained before using it to find another mixed strategy parameterisation. This strategy would be the product of a GA trained on more recent data, so this procedure should be more robust to changes in the stock market data.

Finally, further analysis of the fitness function would also be beneficial. It is possible to incorporate the turnover or any other backtest metric into the fitness function which would have benefits in the applicability of this procedure. As discussed, a penalty term to penalise the inclusion indicators of the same family in the mixed strategy could be added.

8 Conclusions

The work in this thesis has achieved two goals. First, the presentation, implementation and backtesting of the long-short strategies shows that the Cardinality Constrained Portfolio Problem may be solved by statistical arbitrage strategies. There are two questions that must be considered in building these strategies. First, one must decide how the trading decision is automated. Second, one must decide how long to retain a portfolio before liquidating. In this analysis, technical indicators were used to automate the trading decision. The effectiveness of this approach varied depending on the indicator used in the strategy and the particular cluster of stocks to which the strategy was applied. To a lesser extent, rolling window sizes also had an effect. It is clear than within a cluster certain inidcators are more effective than others. As for the question of rebalance horizons, as is commonly the case for statistical arbitrage strategies, short horizons usually are more effective than those with longer rebalance periods. The strategies appear to be most effective when the rebalance period is less than 5 days.

The second area of research in this paper is addressing the question of whether or not a GA could be used to build an improved long-short strategy by mixing the single-signal strategies seen previously. In certain stock clusters, the GA built strategy definitely outperforms any of the single-signal strategies, but in others it does not achieve strong performance in terms of the out-of-sample Sharpe ratio. The performance appears to be very much data dependent. If a cluster behaves similarly in the training set and test set, then the approach seems to be more useful than if this is not the case. If there is a drastic change in market conditions the strategy does not always generalise well. These issues may be overcome by using more data - in this analysis the test set was three years in length and the train set was five. It is likely that these affects could be lessened by using a larger train set.

This research also investigated some features of the GA. In particular, different weight initialisation methods were introduced and scrutinised. Differences in weight initialisation procedures can lead to different optimal strategies but they achieve similar performance. With the deterministic approach, there is less variation in weight component of the optimal solution than with random methods. Also investigated was the optimal gene count in the chromosomes, that is how may strategies were allowed to be present. This number seems to vary by cluster, and in some cases it does not matter. In the two examples provided, it appears that at between 10 and 16 is a good choice.

Finally, this research has demonstrated that the GA optimisation heuristic introduced here does have potential to be a useful tool in asset management. Having achieved some promising results - and observing the notes made in the previous section - this approach to automating feature selection is a meaningful contribution to the existing literature on quantitative portfolio management.

References

[1] L Price A Sortino. "Performance measurement in a downside risk framework". In: *Journal of Investing* 3 (1994), pp. 50–58.

[2] T Back, D. B. Fogel, and Z Michalewicz. *Handbook of Evolutionary Computation.* Oxford: Oxford University Press, 1997.

[3] T-J Chang, N. Mease, J.E. Beasley, and Y.M. Sharaiha. "Heuristics for Cordiality Constrained Portfolio Optimisation". In: *Computers and Operations Research* 27 (2000), pp. 1271–1302.

[4] Y. Choueitfaty and Y. Coignard. "Toward Maximum Diversification". In: *The Journal of Portfolio Management* 35.1 (2008), pp. 40–51.

[5] Rama Cont. "Empirical properties of asset returns: stylized facts and statistical issues". In: *Quantitative Finance* 1 (2001), pp. 223–236.

[6] Gerard Cornuejols and Reha Tutuncu. *Optimization Methods in Finance.* Cambridge University Press, 2007.

[7] Eugene Fama. "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2 (1969), pp. 383–417.

[8] Eugene Fama. "Mandelbrot and the Stable Paretian Hypothesis". In: *Journal of Business* 36.4 (1963), pp. 420–429.

[9] J. H. Holland. *Adaption in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence.* Michigan: University of Michigan Press, 1975.

[10] H. Jalota and M. Thankur. "Genetic algorithm designed for solving portfolio optimization problems subjected to cardinality constraint". In: *The Society for Reliability Engineering, Quality and Operations Management* 9.1 (2018), pp. 294–305.

[11] Ghizlane Lakhnati Jamal Agouram. "A Comparative Study of Mean-Variance and Mean Gini Portfolio Selection Using VaR and CVaR". In: *Journal of Financial Risk Management* 4.2 (2015), pp. 72–81.

[12] Hiroshi Konno and Hiroaki Yamazaki. "Mean Absolute Deviation Portfolio Optimization Model and it's Applications to Tokyo Stock Market". In: *Management Science* 37.5 (1991), pp. 519–531.

[13] Boris Marjanovic. *MS Windows NT Huge Stock Market Dataset.* 2017. URL: `https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs` (visited on 11/16/2017).

[14] Harry Markowitz. “Portfolio Selection”. In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

[15] Harry Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* New York, NY. Wiley, 1959.

[16] M. Mitchell. *An introduction To Genetic Algorithms.* Cambridge MA: MIT Press, 1996.

[17] Robert Novy-Marx. “Backtesting Strategies Based on Multiple Signals”. Working Paper. 2015.

[18] C. R. Reeves. “Genetic Algorithms for Operations Researcher”. In: *Journal on Computing* 9 (1997), pp. 231–250.

[19] C. R. Reeves. *Modern Heuristic Techniques for Combinatorial Problems.* Oxford: Blackwell Scientific Publications, 1993.

[20] William Sharpe. “Mutual Fund Performance”. In: *Journal of Business* 39.1 (1966), pp. 119–138.

[21] S. Srinivasan and T. Kamalakannan. “Multi Criteria Decision Making In Financial Risk management With a Multi-Object Genetic Algorithm”. In: *Computational Economics* 52 (2017), pp. 443–457.

[22] Frantisek Stulajter. “Introduction to copula functions and their application in portfolio and risk management”. In: 2008.

[23] *Trading Technologies: List of Technical Indicators.* 2019. URL: `https://www.tradingtechnologies.com/xtrader-help/x-study/technical-indicator-definitions/list-of-technical-indicators/`.

BROWN_LEWIS_01593018

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60