THE IMPLEMENTATION AND ANALYSIS OF A PRACTICAL MEAN-VARIANCE PORTFOLIO OPTIMISATION MODEL FOR FUTURES CONTRACTS

by

Adam Clark Everett

Thesis submitted to the University of Plymouth in partial fulfilment of the requirements for the degree of:

MSc Data Science & Business Analytics

University of Plymouth
Faculty of Science & Engineering

September 2022

Masters Dissertation License

Copyright Statement

Abstract

THE IMPLEMENTATION AND ANALYSIS OF A PRACTICAL MEAN-VARIANCE PORTFOLIO OPTIMISATION MODEL FOR FUTURES CONTRACTS

Adam Everett

Project Supervisor: Dr. Matthew Craven

This project explores the implementation of a constrained mean-variance portfolio optimisation model using the open source Pymoo module in Python. We examine the impact of applying portfolio optimisation to futures contracts, an asset class that has historically received less attention within the relevant literature relative to other financial instruments. Using the evolutionary algorithm NSGA-II as implemented in Pymoo, we investigate the impact of algorithm parameters, population initialisation and normalisation procedures, as well as cardinality and maximum weighting constraints upon Pareto front formation and hypervolume. The in sample results of the portfolio solutions generated via optimisation are assessed against the model's objective value estimates to measure the excess return level among a variety of test configurations. We find that applying a set of parameters and constraints intended to produce realistically implementable portfolios to serve as the basis for simulating a short term momentum based futures trading strategy results in excess returns within sample which, on average, underperform the model estimates. We conclude with a series of proposals for alternative structures and additional refinements to the model which may result in opportunities for improved out of sample performance.

Table of Contents

Table of Figures

Index of Tables

Acknowledgements

I would like to extend my sincere thanks to Dr. Matthew Craven who has served as my dissertation supervisor throughout the course of this project. His guidance as I have explored a series of new and challenging subjects in the process its completion has been greatly valued.

My most heartfelt gratitude is due to my wonderful wife Imogen who has provided steadfast support during my entire journey through this Masters program and the long days and hours devoted to the production of this work.

1. Introduction

Seventy years subsequent to the initial publication of Harry Markowitz's essay "Portfolio Selection" in the Journal of Finance, the outlined model for portfolio optimisation which bears his name retains an eminent position in multiple areas of academic enquiry. Providing a solid foundation for further exploration, many works have rigorously expanded upon his initial ideas. Rather than presenting a fully developed model prescribing a quantitative process for portfolio construction to replace the typical intuitive or ad hoc methods of the time, he laid out a highly versatile framework, able to accept necessary modifications to accommodate the specific goals of its employer.

The investment landscape has changed dramatically over the decades following the model's introduction. Today, retail investors in many cases can participate in markets in which they would previously have faced much greater barriers to entry in terms of factors such as high transaction costs and minimum capital requirements. These developments, in combination with access to readily available tools for performing quantitative analysis have the potential to create new opportunities for the application of portfolio optimisation models by non-professional market participants.

In this project our primary objectives are twofold. Firstly, the implementation of a constrained mean-variance optimisation model utilising a freely available framework, the Pymoo Python module. We shall examine the impact of a variety of inputs, parameters, and constraints upon the quality and diversity of solutions which are generated. In addition, we seek to assess the performance of this model in a configuration meant to simulate a short-term futures trading strategy which uses varying time periods of historical return data to generate the required mean return vector and covariance matrix. The majority of research into portfolio optimisation has typically relied upon the return data of publicly traded equities for its analysis. However, Markowitz himself, in his original 1952 paper, outlined the potential he saw for the model to be applied to a multitude of different asset classes including bonds, stocks, and real estate (Markowitz, 1952). We will endeavour to assess, using very recent market data which spans a five and one half year period from 2017 to June 2022, whether such a strategy based on mean-variance optimisation demonstrates a capacity to generate returns that match or exceed the model's estimated objective values.

The intention of this work supposes that these analyses may provide insight or guidance into the process by which others may seek to employ a framework such as Pymoo in modelling and analysing portfolio or other forms of multi-objective optimisation. Additionally, we will draw certain conclusions as to the practicality of applying mean-variance optimisation to an asset class such as futures contracts, and describe further possible refinements to such a model beyond the scope of this project which have the potential to increase its value as an applicable portfolio optimisation tool for market participants.

2. Literature Review

2.1 Computational Research into Evolutionary Algorithms and Portfolio Optimisation

To a significant degree, the chosen objects of our analysis stand at the nexus of two independent areas of academic enquiry. An extensive body of literature exists which has explored various metaheuristic procedures and evolutionary algorithms as applied to constrained portfolio optimisation problems. It is apparent that the majority of these works are primarily focused upon investigating the computational and mathematical aspects of the applied techniques. Analysis is often directed towards comparing the efficacy of different procedures in terms of computational efficiency, quality of solutions, or the impact of applying different constraints to a given problem.

A common characteristic shared among them, and noted in multiple papers which have conducted thorough reviews of the relevant literature, is a very limited or completely absent focus upon the financial implications resulting from their experimentation (Kalayci et al., 2019; Metaxiotis & Liagkouras, 2012). Whether the models employed and tested are able to generate portfolio allocations that have some degree of predictive utility, remains unanswered in most cases.

However, many of the papers within this category focus their efforts upon exploring some of the constraints which may aid in providing for the mean-variance model to be applied in a more realistic and practical fashion, but at the cost of imposing significantly greater computational complexity upon the process of generating solutions. In particular, given its practical value in portfolio construction, and the very challenging structure it imposes on an optimisation problem from a mathematical standpoint, investigation of cardinality constrained problems are a primary topic of focus among the set of commonly applied constraints.

Chang et al. (2000) examines the application of three metaheuristic procedures: genetic algorithm, tabu search, and simulated annealing to a cardinality constrained portfolio optimisation problem utilising equity data from a number of major stock indices. They provide multiple examples of the discontinuous structure of the efficient frontier that will result under such a constraint, particularly among the segments of the population with lower return and variance objective function values. The impact upon the Pareto front of a wider set of constraints including upper and lower bounds upon individual asset allocation is illustrated in Chiam et al. (2008). This work provides multiple examples of the effect of such constraints upon the Pareto front, relative to that of an equivalent unconstrained problem. Changes in the structure of portfolio composition under differing sets of constraints are also investigated, with a focus upon encouraging practical asset allocations.

Another highly relevant work tested a total of five different procedures including the NSGA-II algorithm employed by this project, upon a cardinality constrained problem with a cardinality limit of ten using multiple asset pools consisting again of major stock index components (Anagnostopoulos & Mamanis, 2011). Their findings placed NSGA-II in second place behind the Strength Pareto Evolutionary Algorithm (SPEA2) in terms of computational performance.

It is evident that the use of futures contracts as the subject of mean-variance optimisation and subsequent analysis comprises a rather insignificant minority of the published literature in this area. Kalayci et al. (2019) examined in excess of 60 works published between 2004 and 2019 to classify the data sets used in these projects. They span a wide range comprising different assets from markets across the world, but all utilise publicly traded equity prices or hypothetical datasets.

Pai & Michel (2014) outline a pair of metaheuristic strategies they term the multi-objective evolution strategy (MOES) and multi-objective differential evolution (MODE). The asset pool available for portfolio selection was composed of 37 financial futures, including equity indices, currencies, and bonds covering a period from 2004-2013. Their problem takes the form of a minimum variance optimisation, with a detailed risk budgeting constraint among the three categories of assets, formulated as an additional objective function based upon constraint violation minimisation. They found that the impact of the utilised metaheuristic procedures on a problem with a complicated set of constraints and a similar set of assets to our own project was able to generate consistent results in terms of Pareto front formation.

2.2 Statistical Research into Parameter Estimation

The vector of mean returns and covariance matrix are the core inputs which influence the results of any mean-variance portfolio optimisation model. A large body of work has been published focused upon the challenges of constructing a covariance matrix that will accurately represent the true relationship between the underlying set of assets. Huo et al. (2012) discusses the distorting effect that outliers in a dataset have upon the calculation of covariance. This can give rise to what they term “spurious correlation” in which conventional methods of calculating covariance under such conditions may produce large values which in reality are zero. Robust estimation procedures are proposed as a means of countering this effect.

The employment of robust statistical methods, intended to address some of the challenges that arise in the utilisation of data that is non-normal in distribution and may contain an abundance of outliers, to the optimisation of commodity futures portfolios is explored in Adhikari et al. (2020). This work compares the outcome and performance of optimisation based commodity portfolios against an equally weighted alternative over a variety of historical lookback and holding periods. Their findings include the assessment that the value of historical data in parameter estimation declines significantly when periods longer than twelve months are used. Optimisation results are measured over holding periods ranging from one to twelve months. They conclude that the robust optimisation procedure employed offers significant improvement to portfolio performance in comparison to the equal weight method within the sample period covering a range from 1986 to 2018.

Ledoit and Wolf (2003; 2004) outline the covariance matrix shrinkage procedure designed to mitigate the problems that result from estimation error among outlying data. This process will be applied in our own project and discussed in further detail within the body of the work.

2.3 Financial Research into Portfolio Optimisation

Many researchers have written on the fundamental nature and structure of futures contract pricing, and the significant differences they possess in contrast to more traditional asset types which typically form the basis for Markowitz optimisation such as equities or fixed income instruments. Equities are representative of an ownership stake in an ongoing venture which can be expected to generate some degree of positive return across time so long as the underlying organisation continues to operate profitably. This may be present either in the form of dividends or capital appreciation. Whether or not there exists any risk premium or other quality to futures contracts which would be indicative of a positive expected return across time is a subject which has been widely investigated and remains an unsettled topic of research to the present day. Erb and Harvey (2016) seeks to investigate the proposition that long only commodity investment may provide an attractive alternative to equities for investors. They find support for the assertion that across time the excess returns generated by individual commodity futures contracts are approximately equal to zero. However, they do conclude that a number of momentum based strategies have indeed achieved excess returns outperforming equities over various periods within their dataset spanning 1969 through 2004. Looking at the same time period, Gorton and Rouwenhorst (2006) find that an equally weighted index of commodity futures achieves excess returns of approximately 5% relative to the contemporaneous US risk free rate. They attribute this to the existence of an identifiable risk premium in the chosen set of instruments. Such findings highlight the complexity of discerning the nature and prospect of excess return potential among this asset class, and the conflicting conclusions that arise in various projects seeking to investigate this question.

Black (1976) argued that a futures position is much more accurately conceptualised as akin to a sports bet than an equity investment. No money is exchanged at the point of initiation of a position; only upon the conclusion of each trading day are the gains and losses of each position settled. He notes the importance of the fact that the difference between the total number of outstanding long and short positions must always be zero, and as a result, any distinction in the nature of a long vs. short position that exists with regards to equities or any other tangible asset does not apply in the case of futures contracts.

Gorton et al. (2013) investigates the potential factors that may predict the existence of a risk premium in the futures market including inventory levels and the positioning of traders as available in published data. They identify a negative non-linear relationship between the convenience yield and levels of inventory in the case of physical commodities. With regards to the positioning of market participants, they note some patterns of change correlated with price movement, but conclude that such data does not possess any predictive validity. They conceded that available inventory data is subject to significant estimation error and as such is likely to provide little in the way of predictive ability to identify the existence of states that would imply positive expectation of return.

Given the decidedly unclear nature of futures contract value prospects across time when compared against equity or fixed income investment, some researchers have dispensed with such considerations and focused upon alternative factors which may be predictive of future price movement. Miffre and Rallis (2007) model and evaluate a variety of futures trading strategies focused upon price momentum as the primary indicator used to direct portfolio allocations. They utilise an asset pool of 31 of the most liquid futures contracts traded on US exchanges, overlapping significantly with those available in our own model. With a historical dataset ranging from 1979 to 2004, they find their momentum strategies to perform best over a shorter term one to twelve month holding period, generating average returns of 9.38% annually.

In a later article, Fuertes, in addition to Miffre and Rallis (2010), continue their analysis by examining a series of short term futures trading strategies based upon momentum and term structure signals. They again find that the tested strategies generated positive returns over the holding period, with an annual return in excess of 10% in contrast to a passive long-only strategy yielding only 3.4%. In their conclusions it is also noted that the long-short portfolios exhibit substantially higher levels of risk in contrast to those that are long only, and stress the importance of protective strategies such as the employment of stop-losses on positions in order to lessen the potential for catastrophic loss under adverse market conditions.

A prevalent criticism of portfolio optimisation models in general is the frequent experimental finding of poor out of sample performance relative to much simpler methodologies that may be employed for asset allocation. DeMiguel et al. (2009) tests a total of 14 different optimisation models against a *1/N* strategy in which all available assets are given equal weighting within a portfolio, also referred to as naive diversification. They find that the out of sample performance of optimised portfolios which utilise mean-variance as objectives and rely upon a sample covariance matrix is inferior to a *1/N* portfolio as measured by Sharpe ratio. They attribute this primarily to the significant effects of estimation error within both expected returns and variance. Exploring a variety of proposed methods for addressing this problem, they conclude that the seven tested methods have little impact in improving out of sample performance. Their findings are reassessed in Behr et al. (2013) in which a minimum variance model utilising a shrinkage procedure based on Ledoit and Wolf’s method is compared against a similar *1/N* strategy. In their tests, this method outperforms the equally weighted portfolios, providing an example of the disagreement which may frequently occur in the performance of such strategies between different research projects.

Kritzmen et al. vigorously oppose the assertion that the challenges encountered in optimisation models provides justification for their replacement by methods of naive diversification such as *1/N*, describing this notion as a ‘capitulation to cynicism’ (2010, p. 37). They stress the importance of investor judiciousness when relying upon data which results in improbable expectations of return and variance compared to historical norms. Also posited is the value of categorising historical periods of data on the basis of volatility regimes. This information may then be used in conjunction with investor expectations for an upcoming time period.

Throughout the literature devoted to Markowitz portfolio optimisation and its many derivatives, is the continued finding of inconsistent or contradictory results as to its usefulness as a tool for constructing realistic investment portfolios. Even though many decades have passed since its introduction, there undoubtedly still remains significant capacity for further research into the portfolio optimisation model from both a computational and financial standpoint.

3. Methodology

3.1 Dataset

Serving as the basis for our investigations is the Continuous Futures dataset, accessed via the Nasdaq Data Link website under academic license (Stevens Analytics, 2022). It consists of pricing data for 78 futures contracts traded on leading international exchanges. Within this dataset we have access to daily settlement prices for the included contracts covering just over fives years from 2017 through June of 2022.

Constructing a continuous price series for a publicly listed equity which has traded without interruption throughout the desired period is a simple and intuitive task. In contrast, each individual futures contract has both a specific date of inception, upon which it is first listed, as well as one of settlement or expiration. In order to maintain a position in relation to a specific instrument beyond the settlement date of a particular contract, it is necessary to simultaneously close the current position and open a new one in the subsequent contract, a process commonly described as ‘rolling’.

In precisely the same manner, a continuous futures price series must employ a roll methodology to link individual contracts together. The simplest, and a historically common means of doing so, involves switching from the price of the current contract on the day of expiry to the next contract with nearest settlement. This method, however, has been shown to result in excess volatility around these rollover dates which may induce a distorting effect upon the series (Carchano & Pardo, 2009). A variety of more sophisticated methods for producing continuous series have been developed which seek to minimise distortion or abrupt jumps in price and are more likely to be analogous to strategies a market participant seeking to maintain longer term exposure might employ.

The Stevens Continuous Futures dataset offers price series’ calculated using a total of 14 different methodologies, the parameters by which they are categorised being: date of contract roll and price adjustment procedure. The most suitable rolling strategy may often be very specific to the underlying instrument and any standardised approach with an aim of simplicity must make compromises in this regard. For the purposes of testing we will rely primarily upon a method which utilises a first of the month of expiry roll strategy in combination with a calendar weighted price adjustment procedure. The calendar weighting procedure produces a price which is a weighted average of holdings based on a process by which 20% of contract value is shifted to the subsequent contract over the course of five days prior to the chosen roll date. On the fifth day, 100% of the value has been transferred to the new contract and the rollover process is complete. This method allows for the most advantageous series from among the available options, providing an accurate representation of the true price movement of the underlying instrument across time and avoids the significant volatility which may occur close to the settlement date among certain asset types.

The complete dataset includes futures contracts traded on exchanges located in Canada, China, Europe, and the United States. For the purposes of our analysis, we will restrict our available asset pool to those contracts traded on the Chicago Mercantile Exchange (CME). The CME is the largest futures exchange in the world as measured by number of contracts outstanding. As a result, the ability for investors or speculators to access these markets and take positions in the underlying instruments directly or through derivatives is widely accessible in many major jurisdictions. The use of this asset set is intended to provide a wide variety of allocation options, including some of the most liquid and widely traded futures contracts in the world, whilst maintaining a manageable number of decision variables for our optimisation model.

Table 1: Stevens Analytics Continuous Futures data

Quandl Code	Exchange	Symbol	Depth	Method	Date	Settle
CME_AD1_FW	CME	AD	1	FW	2018-06-18	0.74
CME_BO1_FW	CME	BO	1	FW	2018-06-18	29.57
CME_BP1_FW	CME	BP	1	FW	2018-06-18	1.33
CME_C1_FW	CME	C	1	FW	2018-06-18	356
CME_CD1_FW	CME	CD	1	FW	2018-06-18	0.76

To provide an efficient framework from which we may query our dataset in order to construct the mean return vector and covariance matrix for optimisation, data from the relevant categories is loaded into an SQLite database. As a result, we are able to perform SQL queries directly from our optimisation scripts, relying upon the sqlite3 module from the Python standard library to generate a data frame composed of daily price settlement data for each asset selected for inclusion within a chosen time period.

Table 2: SQLite database format

date	quandl_code	settle
2018-06-18	CME_AD1_FW	0.74
2018-06-18	CME_BO1_FW	29.57
2018-06-18	CME_BP1_FW	1.33
2018-06-18	CME_C1_FW	356
2018-06-18	CME_CD1_FW	0.76
2018-06-18	CME_CL1_FW	65.79

In its basic form, as shown in Table 2, the dataset consists of settlement prices for the chosen set of assets upon the end of each trading day within the selected range of dates. To generate the required inputs for our optimisation we must first calculate a series of daily returns as a percentage for each asset.

Different futures contracts may, for a variety of reasons, not always trade on an identical schedule to each other. Utilising a set of assets all traded on the same exchange mitigates the challenges that this can produce, but does not entirely eliminate them. In order for the values in the covariance matrix to correctly represent the relationship between different assets based on daily returns that align and are the result of trading activity from the same day, we must take precautions in the data restructuring process. To address this the data frame is indexed on the basis of date, resulting in a structure in which each column represents a particular futures contract and each row the settlement price for a particular date. Prior to the calculation of daily return values, rows containing null values will be removed. Certain lightly traded contracts which have a large number of missing settlement prices relative to the majority of the set will be excluded from our analysis. Following this, we are left with a pool of 40 individual futures contracts where the total number of dates in which there are missing settlement prices amongst them totals less than ten from a total period consisting of over 1200 trading days. This set of futures contracts includes underlying assets such as commodities, currencies, bonds, and equity indices.

Table 3: Daily settlement price for each asset indexed by date

date	CME_AD1_FW	CME_BO1_FW	CME_BP1_FW	CME_C1_FW	CME_CD1_FW
2021-12-31	0.7278	56.4840	1.3531	593.25	0.7905
2021-12-30	0.7258	55.9880	1.3505	596.00	0.7847
2021-12-29	0.7254	56.7840	1.3486	605.50	0.7815
2021-12-28	0.7231	56.5740	1.3423	604.75	0.7802
2021-12-27	0.7200	56.7200	1.3441	614.75	0.7814

The final stage in preparing our data for use in producing the required optimisation inputs is to calculate the daily percentage return of both a long and short position for each asset. For each column we calculate the return for each date relative to the preceding one. As our optimisation will also permit short positions, for each column of daily returns, a corresponding one will also be created representing a short position in the same contract, with a daily return opposite to the long position, demonstrated in Table 4.

Table 4: Daily percentage returns calculated for each asset, long and short, indexed by date

date	CME_AD1_FW_long	CME_AD1_FW_short	CME_BO1_FW_long	CME_BO1_FW_short
2021-12-31	0.00282	-0.00282	0.00886	-0.00886
2021-12-30	0.00055	-0.00055	-0.01402	0.01402
2021-12-29	0.00318	-0.00318	0.00371	-0.00371
2021-12-28	-0.00152	0.00152	-0.00257	0.00257
2021-12-27	-0.00131	0.00131	0.02309	-0.02309
2021-12-23	0.00485	-0.00485	0.01113	-0.01113

Our data is now structured in the final form required for parameter estimation. Each value represents the daily percentage return for a long or short position in each asset. All returns correspond to the same trading date for all assets and no null values are present. For the specified period of historical data chosen, we can now easily calculate the mean daily return for each position, and construct a covariance matrix with dimensions corresponding to the number of individual positions available for portfolio allocation.

3.2 Mean-Variance Portfolio Optimisation Model

The mean-variance, or Markowitz portfolio optimisation model takes the following standard form:

N = total number of positions which may be allocated weight within the portfolio

μ_i = the expected return for asset i (i = 1, … , N)

σ_{ij} = the covariance between assets i and j (i = 1, … , N; j = 1, … , N)

w_i = the decision variable denoting the proportion of the total portfolio held in asset i

$$Minimise \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \tag{1}$$

$$Maximise \sum_{i=1}^{N} w_i \mu_i \tag{2}$$

This structure comprises a bi-objective optimisation problem in which we simultaneously seek to minimise the overall variance of the portfolio and maximise mean return.

This basic form of the Markowitz model, which will often be described as “unconstrained” typically does assume the presence of the constraint

$$\sum X_i = 1 \, (i = 1, \ldots, N) \tag{3}$$

which requires that the sum of all portfolio components is equal to one. Absent this constraint, values less than one would entail leaving a proportion of the capital allocated to the portfolio uninvested. As the inherent assumption in conducting a portfolio optimisation exercise is that one has a specific sum of capital to invest, the allowance of solutions that do so only in part is contradictory to the fundamental intention of the process. Conversely, sums greater than one would represent a capacity to invest a more than the allocated capital, an obviously irrational condition to permit.

Such a constraint that may be outlined mathematically in a simple manner does, however, pose significant challenges when attempting to generate solutions via evolutionary algorithms. Within Pymoo's framework, all constraints must be structured as expressions that evaluate to less than or equal to zero in order to be satisfied and allow for the generation of a feasible solution. The portfolio summation constraint as implemented in this Pymoo, can be expressed in the form

$$\left(\sum X_i - 1\right)^2 (i = 1, \ldots, N) \tag{4}$$

which will evaluate to zero and satisfy the constraint only if the sum of all portfolio weightings is equal to one. A strict equality constraint such as this is exceedingly difficult for our algorithm to satisfy. Under such a constraint, it will require an extremely large number of evaluations in order to produce even a limited set of feasible solutions. To address this, our model will employ a procedure in which all asset weightings will be normalised prior to each objective function evaluation, ensuring that they always sum to one. With this in place, such a constraint can be excluded from the optimisation code.

Portfolio optimisation models may use different time periods over which returns are calculated: daily, monthly, yearly, etc. The appropriate range will vary depending on the length of historical data employed in estimating model parameters, as well as the time frame of the proposed investment or trading strategy. As we seek to utilise our model in the simulation of a short term futures contract trading strategy, all input and output values will be structured on a daily basis.

3.3 Representing Short Sales within the Model

The proper method for representing short sales within the mean-variance portfolio optimisation framework has been a subject of disagreement. One commonly used technique presents serious problems in attempting to realistically model the procedure. It might be presumed that the short sale of a particular asset could simply be modelled by permitting a negative portfolio weighting in the corresponding decision variable. Doing so in sole conjunction with a $\sum X_i = 1, (i = 1, \ldots, N)$ constraint is described by Markowitz himself as “an absurd representation of real-world short positions” (Markowitz, 2010, p. 3). A short position represented by a negative weighting would, in effect, create additional capacity for weighting to be applied to other long positions within the portfolio. This would imply that the proceeds of a short sale are immediately made available to an investor, leaving them free to apportion these funds among other positions. Of course, as Markowitz noted, this is completely unrealistic. In the case of equity short selling, proceeds of the sale will be held by the broker and typically invested in some form of risk free security for the duration in which the position is open. In the case of futures contracts, no money at all is exchanged between the parties upon the initiation of a position, and the implications of a negative weighting would be equally inappropriate.

The short sale of equities is an operation conducted in multiple steps. A market participant who wishes to engage in the practice must, typically through a broker, arrange to borrow the shares to be shorted. They may then be sold and the seller takes on an obligation to replace the shares in the future. The short seller seeks a decline in the underlying instrument’s price. If this occurs, they may replace the shares at a discounted price and the difference, less any associated transaction costs, will be their profit.

Equity short selling is subject to much more complexity and regulation than the simple purchase of the same shares. Given these challenges of implementation, the majority of portfolio optimisation models have historically disallowed short selling by imposing a constraint prohibiting negative weightings (Jagannathan & Ma, 2003). Such a constraint in the form $X_i \geq 0, (i = 1, \ldots, N)$ stipulates that all portfolio weightings must be greater than or equal to zero.

Upon initiation, a futures contract is essentially a bet on subsequent price, whichever direction the movement may be in. That historical mean returns are indicative of what they are also likely to be in the future is one of the primary axioms encompassed by the Markowitz mean-variance framework. If this premise is accepted, a model with the capacity to generate optimised portfolios made up of both long and short positions will clearly provide greater flexibility and opportunity to generate superior objective values compared to an approach which limits available allocation options by half.

To contend with the challenges this presents to an optimisation framework, and to provide for an accurate representation of long/short futures portfolios, we shall utilise a method described in Pogue (1970) and examined in further detail in Jacobs et al. (2005; 2006). The authors refer to this formulation as the "*2N*" approach, in which the number of decision variables is doubled from the long only structure. An additional variable is added for each asset's long position, representing a corresponding short position. By doing so, a lower bound of zero can be maintained for all positions, and thereby avoid the aforementioned problems which arise when negative weightings are permitted.

Under this structure, the portfolio return may be represented by the equation

$$R_p = \sum_{i=1}^{n} r_i x_i + \sum_{i=n+1}^{2n} (-r_{i-n}) x_i, (i = 1, \ldots, N) \tag{5}$$

as described in Jacobs et al. (2006).

The first term represents the contribution to total portfolio return of the long positions and the second term for those which are held short. In the same paper, the condition of 'trimability' is defined as being satisfied when a portfolio generated via such an optimisation structure does not simultaneously allocate to both a long and short position within the same security. In the event that this occurs, it will be necessary to eliminate overlapping long and short positions from the same asset and conduct a normalisation procedure to adjust weightings to maintain relative proportions before recalculating objective values.

3.4 Covariance Matrix Estimation

A multitude of works have noted the crucial consideration that, given the standard procedure for estimation of parameters within the Markowitz model, many optimised portfolios could just as accurately be described as ‘error maximised’ portfolios. Absent efforts at mitigation, mean variance optimisation will systematically generate portfolios that give heavy weighting to variables with the highest mean return and lowest variance. The corresponding parameters for these variables are precisely those most likely to be subject to the highest degree of estimation error (Michaud, 1989).

Ledoit and Wolf have collaborated together on a long series of published works spanning decades focused upon identifying and addressing weaknesses inherent in reliance upon the sample covariance matrix within the mean-variance optimisation model. They reiterate the vital consideration that variables whose mean return and covariance is estimated with data encompassing the most extreme outliers within a dataset will often be the primary targets to which an optimiser will be drawn (Ledoit & Wolf, 2003).

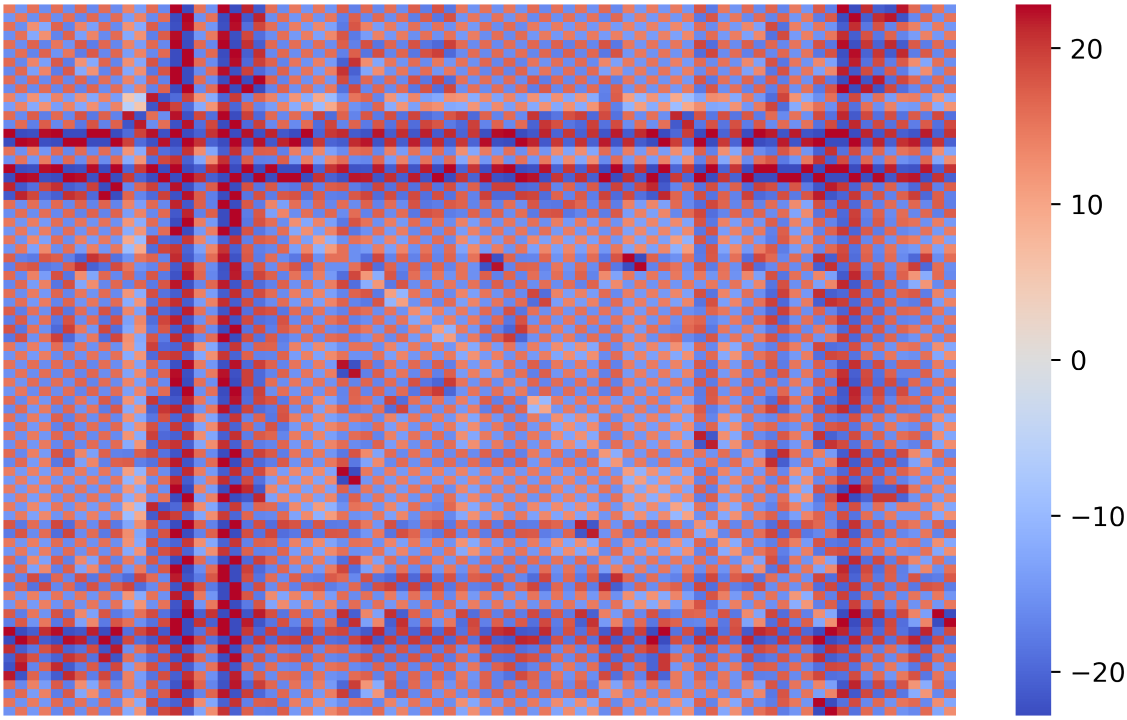

Figure 1: Sample covariance matrix of daily returns from 2017-2022 for 40 asset set, long and short positions. Log_2 of values taken to improve heat map contrast. Negative scale values correspond to negative covariance.

The time period covered by our dataset includes a number of examples of outlying data of historically extreme prominence. During April of 2020, as settlement neared for the CME's West Texas Intermediate crude oil contract, prices plunged and some positions were even closed at negative prices. The specifications of this instrument stipulate physical delivery of the underlying commodity, each contract representing an obligation to deliver, or take delivery of, 1000 barrels of oil to the major storage and distribution terminal located in Cushing, Oklahoma. At this time, expiry of the contract coincided with the initial period of lockdowns brought about by the COVID-19 pandemic. Demand for crude oil and related products declined at a dramatic rate which could not be matched by a corresponding fall in production within such a short period of time. Under such circumstances, with very limited storage capacity available, the normal opportunity for traders to profit by purchasing oil to be delivered in the short term, and then selling longer dated contracts at a higher level, arranging for storage of the product over the intervening period, was not possible as it would be under normal market conditions. As a consequence, many speculators holding long oil positions, without the capacity to facilitate the physical delivery or storage of the underlying commodity, were forced to exit their positions at an extreme loss. In the case of contracts closed at negative values, this entailed having to in effect pay others to take on the burden of accepting delivery under very constrained circumstances. Similarly, as a result of other unprecedented market conditions that accompanied the early stages of the pandemic, the price of the Random Lengths Lumber contract, also included within our analysis, rose in price from trough to peak by over 500% between the months of March and May 2020. Such extraordinary price movement over such a short period has few historical analogues.

It is essential to act to mitigate the impact of such outliers within the dataset to allow for our model to generate portfolio solutions not excessively concentrated in those assets that have experienced dramatic increases or decreases in price within the historical period of data used. For this, we will make use of Ledoit and Wolf's shrinkage function as implemented by the open source Python module Scikit-learn (Pedregosa et al., 2011). This function performs a shrinkage procedure upon the covariance matrix based upon Ledoit and Wolf's method of calculating a shrinkage coefficient. Applied to a sample covariance matrix, as pictured in Figure 1, this will conduct a transformation, pulling outliers towards more central values (Ledoit and Wolf, 2003; 2004).

Figure 2 provides a visual representation of the impact of the Ledoit Wolf shrinkage procedure upon a sample covariance matrix estimated using a five year set of return data from 2017-2021. Given the low magnitude of the values contained within the matrices, to achieve acceptable visual contrast we take the base two logarithm of each absolute value and then multiply by negative one the elements which were originally negative in order to restore the correct structure. The impact of the shrinkage procedure is most apparent along the diagonal representing the variance of each individual position. For example, in the case of a high variance asset such as the previously described WTI crude oil contract, variance falls from 0.022 to 0.0045 following shrinkage. In contrast, the 2 Year US Treasury Note contract, whose return variance within the dataset is much lower, rises from a daily variance of 3.35E-07 within the sample covariance matrix to 6.3E-04 after the application of Ledoit Wolf shrinkage. The covariance between a long position in each of these assets moves from -2.0E-06 to -3.5E-07 once shrinkage has been applied.

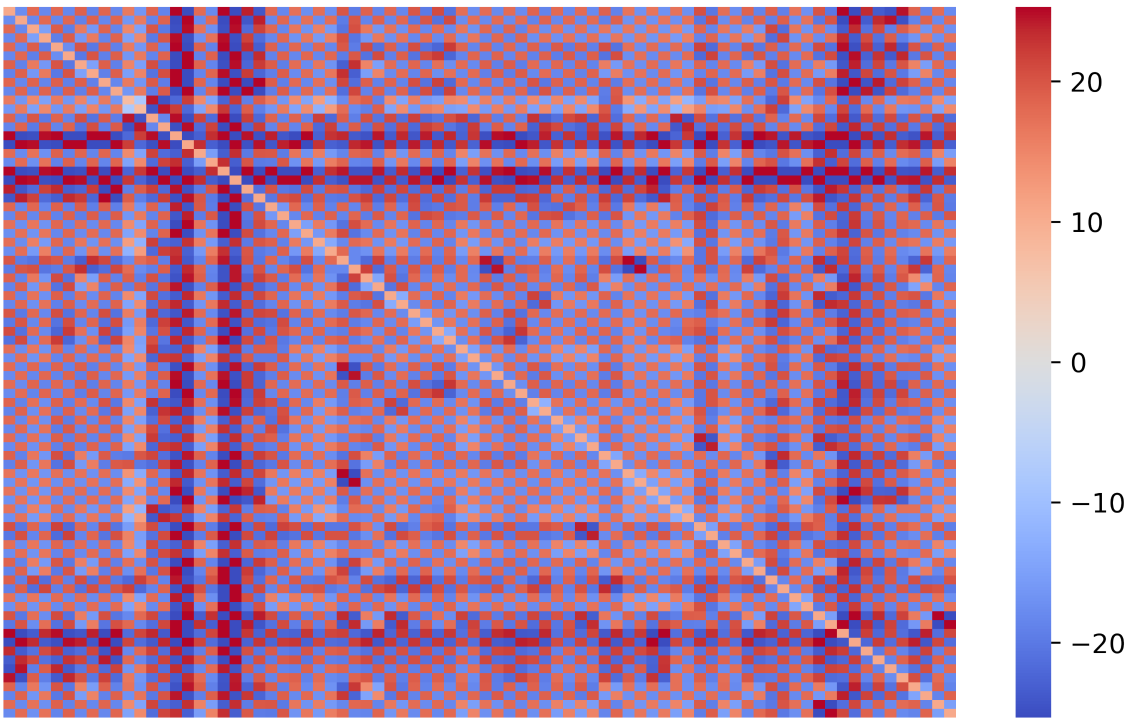

Figure 2: Covariance matrix from Figure 1 following Ledoit Wolf shrinkage. Log_2 of values taken to improve heat map contrast. Negative scale values correspond to negative covariance.

3.5 The Impact of Constraints

An appropriate set of constraints can greatly increase the usefulness of a portfolio optimisation model when it is to be applied to a practical investment or trading strategy. Of course, the imposition of additional constraints has only the potential to hinder the objective values that may be produced. However, doing so has been shown to provide the opportunity to achieve superior out of sample returns in relation to unconstrained examples (Kolm et al., 2014). For the purposes of this project, we will focus our attention upon two commonly applied constraints which, in conjunction, can produce more manageable portfolio allocations by limiting the number of securities held and also forcing a level of diversification that an unconstrained optimisation is unlikely to produce: the cardinality and maximum weighting constraints.

Primary among these, and one that has garnered among the greatest degree of attention of researchers, is the cardinality constraint. This involves placing limits or boundaries upon the number of individual holdings which may be included in optimised portfolios. The primary subset of our data used for analysis and testing of our model and strategy consists of 40 individual futures contracts in which one may take either a long or short position; equal to 80 decision variables in total.

An unconstrained optimisation will tend to generate portfolios comprised of a large number of very small positions. For example, in an unconstrained test using our previously outlined asset pool, some allocation is made to every single one of the 80 positions available. 77 of these are included with an average of less than 2% weighting among the population of optimised portfolios. Particularly in the case of the futures market, given the respective size of individual contracts, it would require very large amounts of capital to actually implement these portfolios. In the context of our proposed strategy which seeks to assess the usefulness of such a model to a non-professional market participant, the construction of such portfolios is unlikely to be feasible.

The cardinality constraint is typically formulated as

$$\delta_i \in \{0,1\}, i=1,\dots,n \quad \sum_{i=1}^{n} \delta_i \leq K \quad , \tag{6}$$

where δ_i is a binary variable corresponding to each of the problem's decision variables. A value of one represents inclusion of a particular asset within the portfolio and zero exclusion. The sum of these variables must be less than or equal to K, the cardinality constraint.

The constraint is implemented logically in Python by the expression

sum(asset > cardinality_threshold for asset in x) – cardinality_constraint

which logically implements the same requirement as the mathematical formulation in the form necessary for the Pymoo module. It will evaluate to less than or equal to zero only if the number of positions with a non-zero weighting is less than or equal to the specified cardinality constraint. A relaxation threshold is added to provide a margin by which variables not included within the cardinality limits of a particular solution may exceed zero by a very small degree, 1e-10, without resulting in a constraint violation and infeasible solution. Examples of the Python code used for data processing and optimisation within this project can be viewed at https://github.com/adamceverett/Futures-Portfolio-Opt.

While it has a beneficial effect in allowing portfolios with a manageable number of positions to be generated, a cardinality constraint also has a profound impact upon the computational complexity of the problem. A standard mean-variance optimisation problem, including any number of linear constraints, is a quadratic programming problem for which solutions can be generated via well established mathematical procedures. When a cardinality constraint is incorporated into this structure the result may be categorized as a mixed integer non-linear programming problem. As a result, the cardinality constrained mean-variance portfolio optimisation problem falls into the category of NP-complete problems. A corresponding proof is described in Bienstock (1996). Many other constrained portfolio optimisation structures may also fall into the same category. For example, Mansini

& Speranza (1997) present a mean semi-absolute deviation portfolio optimisation model with minimum transaction lot constraints. By performing a transformation from their portfolio feasibility problem to two known NP-complete problems: the Partition problem and Knapsack problem, they are able to prove its equivalency and inclusion within the category. NP-completeness characterises a class of problems for which there is no known algorithm in which the computational complexity of N, denoting the problem size, can be bounded by a polynomial function. In the context of cardinality constrained portfolio optimisation, this entails that most example problems of practical size are intractable to the computation of exact solutions in any realistic time period. Instead, practical approaches to such problems must employ alternative methods which have demonstrated the capacity to produce approximate solutions which are acceptable relative to their application and can be computed in a reasonable time period with the hardware resources available. Metaheuristic procedures and evolutionary algorithms are effective and widely used tools to meet this challenge.

In addition to the cardinality constraint, our model will also explore the application of a maximum weighting constraint upon each variable in the decision space vector. Regardless of one's degree of confidence in the ability of an optimiser to generate portfolios that will be likely to achieve the estimated level of return and variance, it is widely recognised that excessive concentration in a single or small number of assets is a source of significant risk. A cardinality constraint can limit the number of holdings to a level that is manageable relative to the capacity of the portfolio holder, but it does not perform any function to prevent heavily concentrated allocations.

A maximum weighting constraint can restrict the allocation of portfolio weighting among any single asset to a limit that is deemed acceptable. Previous investigations have found evidence that placing minimum or maximum bounds upon portfolio weightings did not result in any significant improvement to out of sample performance (Jagannathan & Ma, 2003). As a counter-argument to such findings, the concept of “model insurance” may be taken into account. This stresses the importance to any portfolio construction process of recognising the shortcomings of the underlying model, and the unavoidable estimation error in the relied upon parameters. Constraints, such as those which limit portfolio concentration, can be useful in this regard to provide a degree of protection against the true totality of risks that may be underestimated or unrecognizable to the model (Kolm et al., 2014).

In our Python code the maximum weighting constraint can be formulated by the expression

max((asset - max_asset_weight) for asset in x)

which will only evaluate to less than or equal to zero, and satisfy the constraint, in the event that all decision vector weightings fall below the specified maximum weighting for a given optimisation.

3.6 Evolutionary Algorithm: NSGA-II

To conduct the optimisations for this project we shall utilise the open source optimisation framework Pymoo. Pymoo, implemented in the Python programming language as suggested by its name, provides a highly versatile foundation from which to model multi-objective optimisation problems subject to a variety of different constraints. It includes a multitude of different pre-coded algorithms which may be applied to solve single or multi-objective problems, including some of the most widely used metaheuristic procedures and evolutionary algorithms including NSGA-II, Particle Swarm Optimisation, and MOEAD among others. A detailed description of Pymoo’s structure and functionality is provided in Blank and Deb (2020).

Solutions to our modelled optimisation problem will be generated via Pymoo's implementation of the Non-Dominated Sorting Genetic Algorithm (NSGA-II), first introduced in the paper Deb et al. (2002) as an improvement upon the original NSGA proposed in 1994.

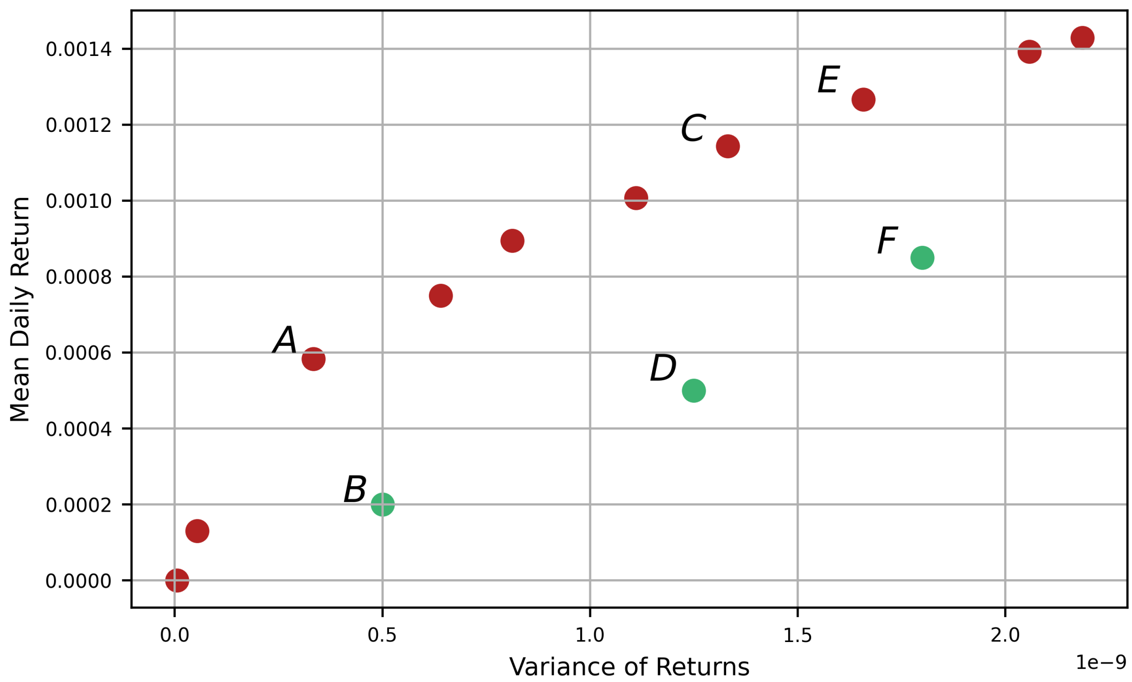

Figure 3: Examples of Pareto dominant solutions

The concept of Pareto dominance is essential to the function of NSGA-II and to the assessment of different solutions in a multi-objective optimisation problem. Figure 3 illustrates an example set of solutions plotted in relation to the two objective functions of our portfolio optimisation problem. Solution *A*, with respect to the problem's objective functions: return to be maximised, and variance to be minimised, can be said to dominate solution *B* if both of the following conditions hold true:

$$Return\ \ A \geq Return\ \ B \text{ AND } Variance\ \ A \leq Variance\ \ B$$

$$Return\ \ A > Return\ \ B \text{ OR } Variance\ \ A < Variance\ \ B$$

Examining Figure 3, we can see that both conditions do indeed hold, and we can therefore state that solution *A* dominates solution *B*. With respect to our objective functions, *A* is strictly superior to *B*. In the case of solution *C* and *D*, solution *C* possesses a greater mean return objective value than *D*, but also a higher level of variance. Therefore, solution *C* cannot be said to dominate *D*, as they represent a trade off between the two objectives. Based on the same criteria, it can be stated that solution *F* is dominated by solution *C* and *E*, but not *A*. Within the financial terminology of portfolio optimisation, a portfolio solution that is non-dominated is described as “efficient”.

When plotted, the total population of non-dominated solutions generated in a given optimisation will form a curve referred to as a Pareto front, or “efficient frontier” in the specific context of portfolio optimisation. The points falling along this front represent the highest quality solutions that the algorithm has been able to generate during a run. The choice of a specific portfolio from among these solutions will offer some level of compromise between the two objectives; the ideal balance between them being a matter of investor preference.

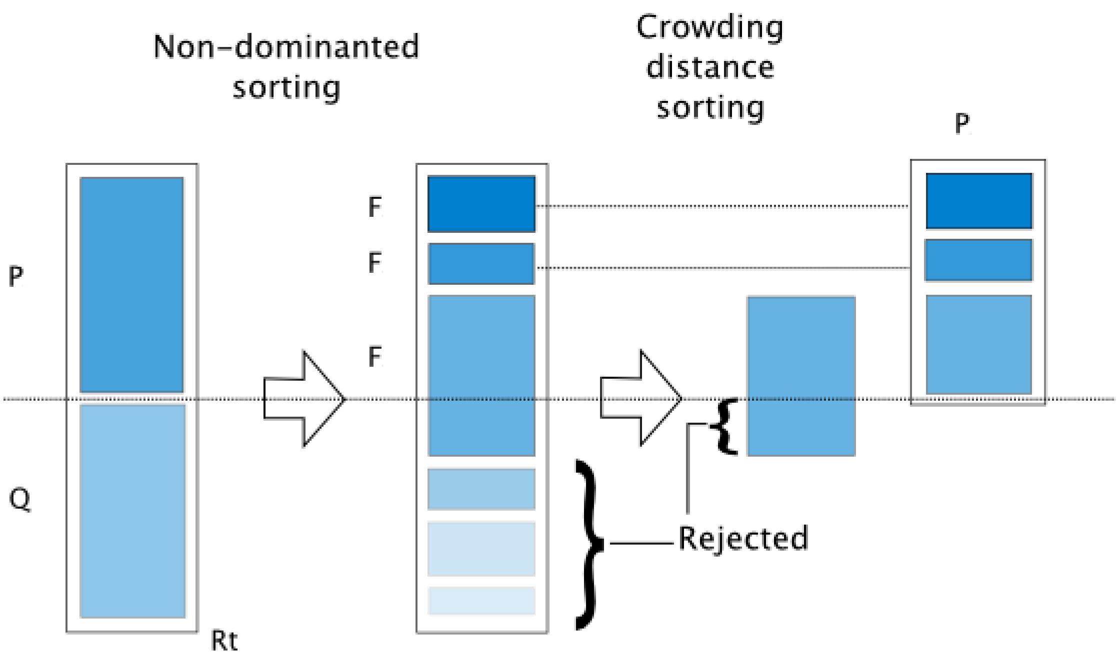

Figure 4: NSGA-II sorting procedure. (Blank & Deb, 2020)

NSGA-II shares many of the typical characteristics of other genetic algorithms, in which decision vectors are represented as individuals within a population and the respective values of their elements are analogous to genetic encodings. The main procedure which occurs during each iteration of the algorithm is depicted in Figure 4. An initial population of solutions *P*, of size *N*, is generated via a random or pre-defined initial sampling procedure. The non-dominated solutions among the population are grouped into a front, labelled *F*. The solutions in *F* are then excluded and this process is repeated so that all solutions will be ranked according to their membership in a given front. Binary tournament selection, crossover, and mutation operators are applied to create a child population, *Q*, of equal size to *P*.

Following the first generation, populations *P* and *Q* will be combined and sorted into fronts on the basis of non-domination as before. Solutions will be selected, starting with the first ranked front and working down in rank to the subsequent ones, for advancement to the next generation. Should *N* be satisfied by the partial inclusion of a ranked front, the solutions will be chosen on the basis of a minimum crowding distance calculation in order to encourage diversity. Once completed, the new population *P* will undergo a crowded tournament selection process, which utilises both Pareto front ranking and crowding distance, along with crossover and mutation operators to create a new child population. This process is then repeated for a certain number of generations or until a pre-defined termination criterion has been met. A detailed description of all procedures of NSGA-II is available in Deb et al. (2002).

The comprehensive review in Verma et al. (2021) found that NSGA-II has become one of the most popularly applied algorithms for generating solutions to multi-objective combinatorial optimisation problems and its occurrence in the relevant literature has been increasing steadily in recent years.

3.7 Initial Population Sampling Methods

To improve the performance of NSGA-II in the presence of constraints such as cardinality, which result in highly discontinuous search spaces, we utilise a procedure to generate an initial population which will satisfy all constraints and thereby generate feasible solutions immediately.

The default parameters for NSGA-II within Pymoo will create an initial population using a random sampling process. In the case of many unconstrained optimisation problems, this process is effective as it produces a population widely distributed throughout the search space which can then, utilising the algorithm's procedures, converge towards an ideal Pareto front.

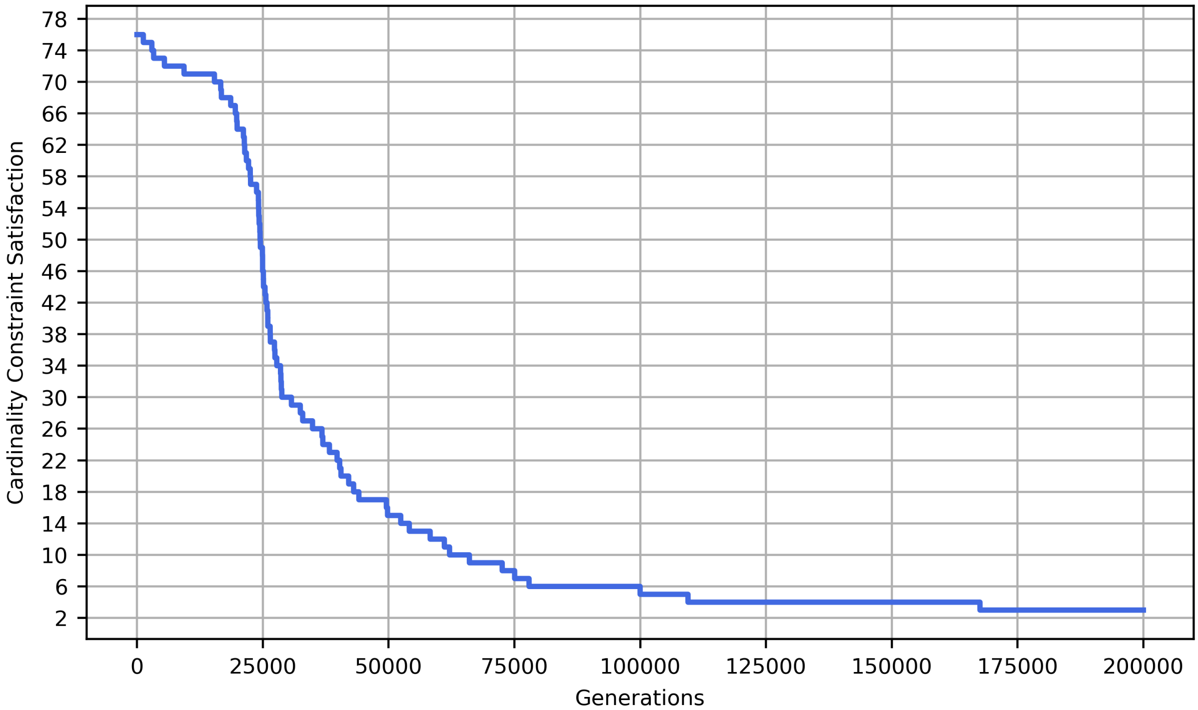

Figure 5: Convergence towards cardinality constraint. K=2 with initial population of 100 generated via random sampling. Cardinality constraint relaxation threshold of 1e-6 used.

In the case of a problem with constraints such as cardinality, random sampling provides an inefficient starting point as the individuals within the population will almost certainly begin their existence with decision space values that produce infeasible solutions. In the context of our problem, this often entails an arduous process of convergence that must be overcome simply to reach the feasible space, particularly in the case of stricter cardinality limits and even more so when combined with maximum asset weightings. Figure 5 exemplifies this challenge, where convergence towards a cardinality constraint of K=2, in the presence of a relaxation threshold of 1e-6, fails to converge towards satisfying the cardinality constraint and generate any feasible solutions even after 200 000 generations.

For an initial population, in the case of a cardinality constraint alone we must generate a set of vectors equal to the population size which contain K elements with random non-zero weighting. These values are then normalised so that the full vector sum will be one and we then shuffle these elements so that allocations are spread amongst the full range of variables.

When combined with a maximum weighting constraint, the procedure is similar but with the additional requirement that none of the elements are given initial weights exceeding the specified maximum. It is also necessary again that all values sum to one. To accomplish this, we make use of the Numpy random sampling module, specifically the routine which draws samples from a Dirichlet distribution.

A Dirichlet distribution across a vector $X_1, \ldots, X_n$ will always satisfy the conditions

$$X_i > 0 \quad \text{and} \quad \sum_{i=1}^{n} X_i = 1 \left(i = 1, \ldots, n \right) \quad . \tag{7}$$

The Dirichlet distribution takes a parameter *α*, which is a vector of the length of the desired sample size. The individual values of *α* may be varied in order to influence the relative concentration towards that element in the vector. This property provides for its common application as a prior distribution within Bayesian statistics. Varying the concentration of the values of *α* is not relevant to our purpose in which we simply seek a reasonably diverse sample equal in length to the chosen cardinality constraint and with each value below the upper bound of maximum weight. With all individual elements in the α vector equal, larger values will result in less variation between the individual components in the resulting sample. Smaller values for α will produce greater variance within the generated sample. As the product of the cardinality constraint and maximum weighting falls closer to one, it may be necessary to employ a larger α parameter, or potentially perform a large number of samples in order to achieve a set of values which fulfils both constraints. Taking heed of these factors which influence its results, the Dirichlet distribution provides an efficient procedure for producing a population which can produce feasible solutions immediately and is diverse within the bounds imposed by the two constraints.

3.8 Hypervolume

A number of quantitative measures have been developed to provide for a means of comparison among different non-dominated solution sets to a multi-objective optimisation problem. One such was first described in Zitzler and Thiele (1998; 1999), as 'size of the space covered'. Today, this procedure is most commonly described as hypervolume and has come to be a commonly used metric for assessing the quality of an evolutionary algorithm generated Pareto front.

Hypervolume is calculated for each individual solution by measuring the area or space dominated relative to a reference point. In the case of a bi-objective problem, each solution will cover an area in the shape of a rectangle which is bounded by the objective values of the solution and

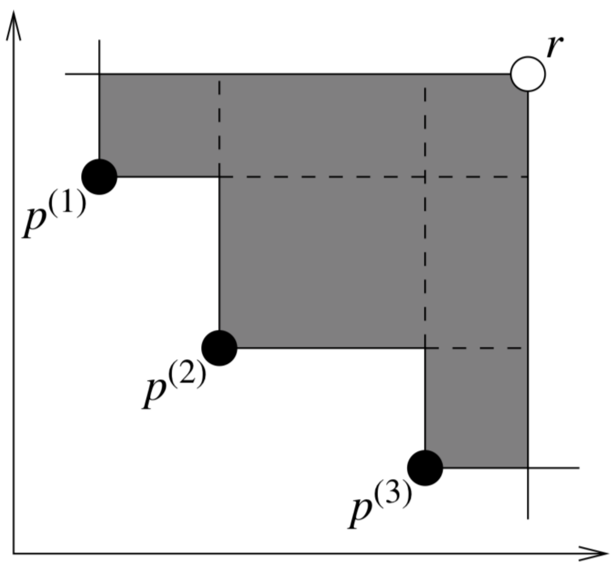

Figure 6: Hypervolume calculation for a bi-objective problem. (Blank & Deb 2020)

those of the chosen reference point. The union of these shapes produced by each solution will form a polygon whose total area represents the hypervolume for a given set of solutions. The process of calculation can be visualised in Figure 6. Hypervolume results are typically normalised to fall between the bounds of zero and one. Although most easily visualised in the context of a problem with only two objectives, the procedure for calculating hypervolume can be extended into any N-dimensional space as required in the case of a multi-objective optimisation problem.

One property of hypervolume, commonly cited as a primary drawback to its application in certain use cases, is the potential for a high degree of computational complexity. However, recent advances in the form of new procedures for more efficient calculation have mitigated this obstacle, particularly for problems with limited dimensionality such as is the case in a typical bi-objective mean-variance optimisation (Guerreiro et al., 2021). This corresponds to the structure of our problem and Pymoo's built in module for calculating hypervolume is able to generate results quickly for a problem of its size using non-specialised hardware.

In the case of problems for which the shape of the optimum Pareto front is known, hypervolume may be used as a means of assessing the performance of an evolutionary algorithm in approximating this ideal set. However, with problems such as will be approached in this project, which may include constraints which elevate the complexity to a level where the optimum Pareto front cannot be calculated, it can serve as a basis of comparison for tests that incorporate a varied set of parameters and constraints.

In comparing the quality of different Pareto fronts, hypervolume is subject to certain biases which must be taken into account when interpreting results. We will explore the impact of some of these factors upon the growth and maximisation of hypervolume and the resulting implications upon interpretation of this indicator.

3.9 Assessment of Portfolio Outcomes

The fundamental applied objective of such an optimisation process is to generate candidate portfolios which may exhibit evidence of potential to achieve superior returns compared to alternative methods of selection. It is therefore essential to assess the actual returns generated over a time period following the historical data sample used to generate the return vector and covariance matrix. The length of this period should correspond to the strategy one seeks to model, which in the case of our analysis, shall be a short term momentum based approach with excess portfolio returns calculated over a holding period of either 10 or 20 trading days.

A variety of methods for calculating the quality of a portfolio optimisation model's estimated vs. actual results have been used. Common examples include the Sharpe ratio and excess portfolio return relative to a contemporaneous risk free rate such as that of United States federal treasury bonds. For the scope of this project, we will focus upon performing a comparison of the mean daily returns achieved over the selected holding periods relative to the mean expected return objective values among the total population of solutions produced by a given optimisation test. Through a series of optimisations conducted with parameters and constraints demonstrated to produce a balance between speed of execution and peak hypervolume, we will calculate the in sample returns over the specified holding periods using varying lengths of historical data to estimate mean return and covariance for the available pool of assets. Daily returns, both optimised values and actual results, will be multiplied by the number of holding days and the mean optimised portfolio returns subtracted from the actual in order to produce the level of excess return.

4. Results

4.1 Characteristics of Data

Conducting a series of tests upon the full range of data used in our optimisation from within the Continuous Futures dataset provides valuable insight into its structure and reveals some crucially important details. The distribution of daily returns for the data which will form the basis of our optimisation tests may be visualised in a series of histograms in which the daily percentage return figures are categorised by year, visible in Figure 7. We then generate a series of statistical measures for each year's data in Table 5.

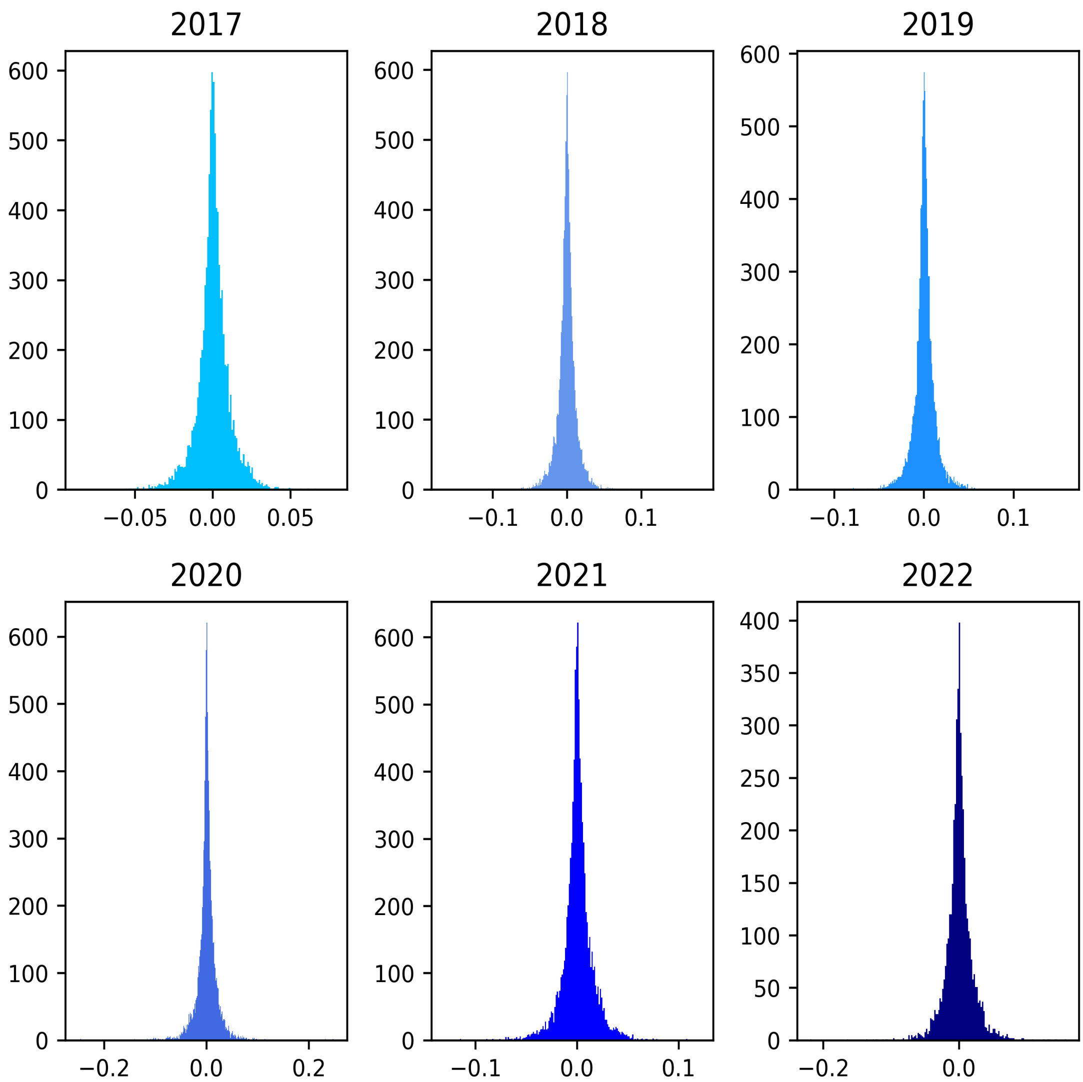

Figure 7: Distribution of daily returns for set of 40 Chicago Mercantile Exchange futures contracts by year. Scale of 2020 histogram limited to exclude significant outliers.

Table 5: Descriptive statistics of daily return data (2022 data through end of June)

Year	Mean	Minimum	Maximum	Variance	Skewness	Kurtosis
2017	0.04%	-8.6%	7.8%	0.000121	0.0022892	5.12
2018	-0.02%	-16.5%	17.9%	0.000175	-0.014343	11.78
2019	0.05%	-12.7%	15.9%	0.000162	0.3411903	11.01
2020	0.01%	-502.9%	25.8%	0.003385	-70.06453	6001.59
2021	0.06%	-13.1%	12.1%	0.000252	-0.224634	7.53
2022	0.02%	-21.9%	15.8%	0.000434	-0.195743	9.13

The mean daily returns within our dataset are small, falling within a range from a mean daily decline of 0.02% in 2018 to an increase of 0.06% in 2021. However, when the minimum and maximum values are viewed, it is readily apparent that our dataset contains a number of outlying observations of significant magnitude. For example, viewing the distribution of returns for 2020, the minimum value, a daily decline of 502.9%, represents an observation located at a distance in excess of 86 standard deviations from the mean, given a hypothetical assumption of normality for the distribution. This example is the product, as previously described, of a historically unprecedented decline in the value of a futures contract into negative territory as the result of extreme market conditions. Examining the minimum and maximum values within each year's subset of our data, the presence of multiple outlying observations far in excess of what would be indicative of a normal distribution is apparent. The difference in variance of returns between the years of the dataset is also wide, ranging from a low in 2017 to a significant high in 2020.

The skewness for our yearly data distributions exhibits no persistent pattern of either positive or negative values throughout the total period. Kurtosis of a normal distribution will present with a value of three; our results are in excess of this within all years of the dataset. Examining all of these considered measures, it is very clear that the structure of return data in 2020 is anomalous relative to the remainder of the dataset. Conducting a Shapiro-Wilk test upon each year's data results in extremely small p-values well below the threshold at which we may reject the test's null hypothesis of normality of the tested distributions.

4.2 Algorithm Parameters

The core parameters which impact the structure of an optimisation test conducted using the NSGA-II algorithm are population size and number of generations. The total number of function evaluations will be the product of these two values.

A review of the relevant literature on optimal population sizing within genetic algorithms reveals that any notion of an ideal population size is highly problem specific and invariably involves a number of trade-offs (Alander, 1992). For the purposes of a practical portfolio optimisation implementation, we must target a set of parameters which may converge upon a well formed Pareto front within a reasonable computation time using commonly available hardware resources.

Figure 8 provides a visualisation of hypervolume expansion in an unconstrained test of our modelled problem. In this test, population size is varied with the number of generations adjusted to achieve a common number of 1 000 000 function evaluations. The rate of convergence towards a plateau of hypervolume is inversely proportional to population size in this test with the exception of small population values less than the length of the decision vector doing so more slowly than a population of 100. The scale of Figure 8 is limited to 500 000 function evaluations as no further gains in hypervolume are observed among any of the tests beyond this point.

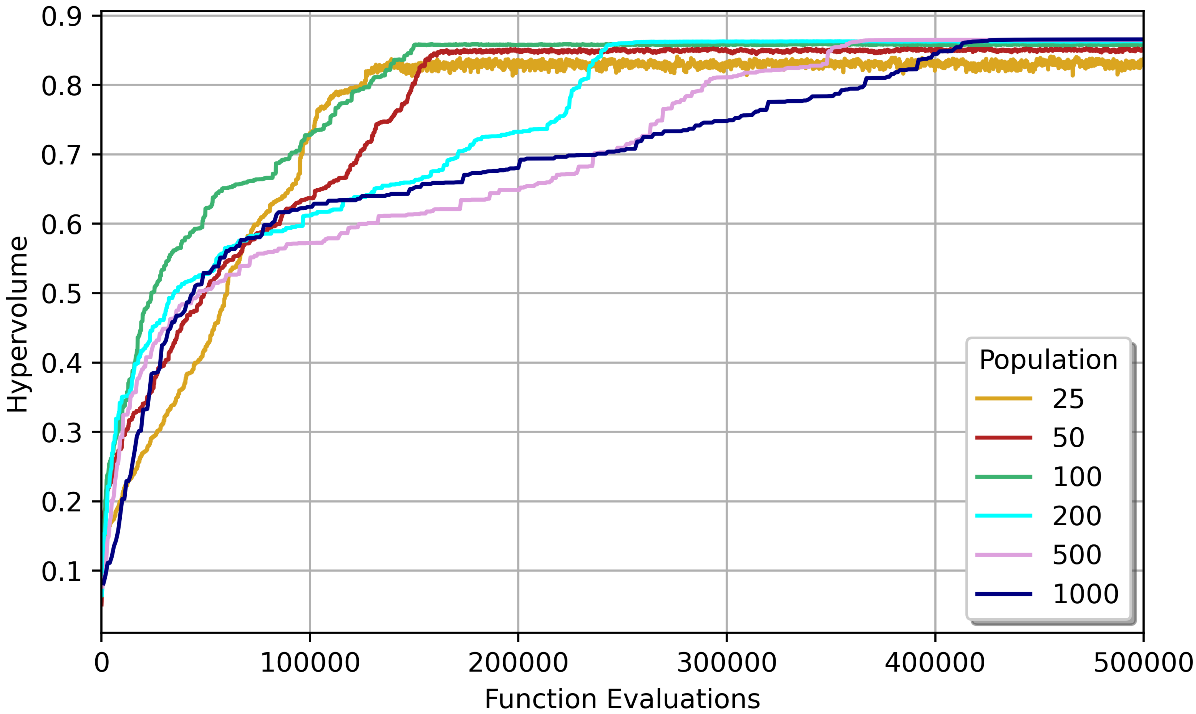

Figure 8: Impact of population size upon hypervolume expansion

As the population increases, a slightly higher peak hypervolume is achieved. Based on the manner in which hypervolume is calculated, for a given shape of Pareto front the value will be maximised as the population of solutions falling along it approaches infinity. However, as shown, for the given number of function evaluations, beyond a certain population the number of generations will be insufficient to reach a maximum level of hypervolume in contrast to lower values. As illustrated in Figure 8, beyond a population of 100, the number of function evaluations required to reach a peak increases as population size grows.

When different numbers of generations are applied to tests incorporating our two constraints, the Pareto front tends to expand to a further range when the algorithm completes a greater number of generations. However, there are clear exceptions as depicted in Figure 9. Given the complexity imposed by the application of the two constraints in conjunction, there are many discontinuities in the Pareto fronts, with significant gaps between different sections. In tests conducted with no constraints or a cardinality constraint alone, convergence towards a maximal Pareto front will be largely proportional to the number of generations until a limit has been reached. After this, limited to no gains will be made.

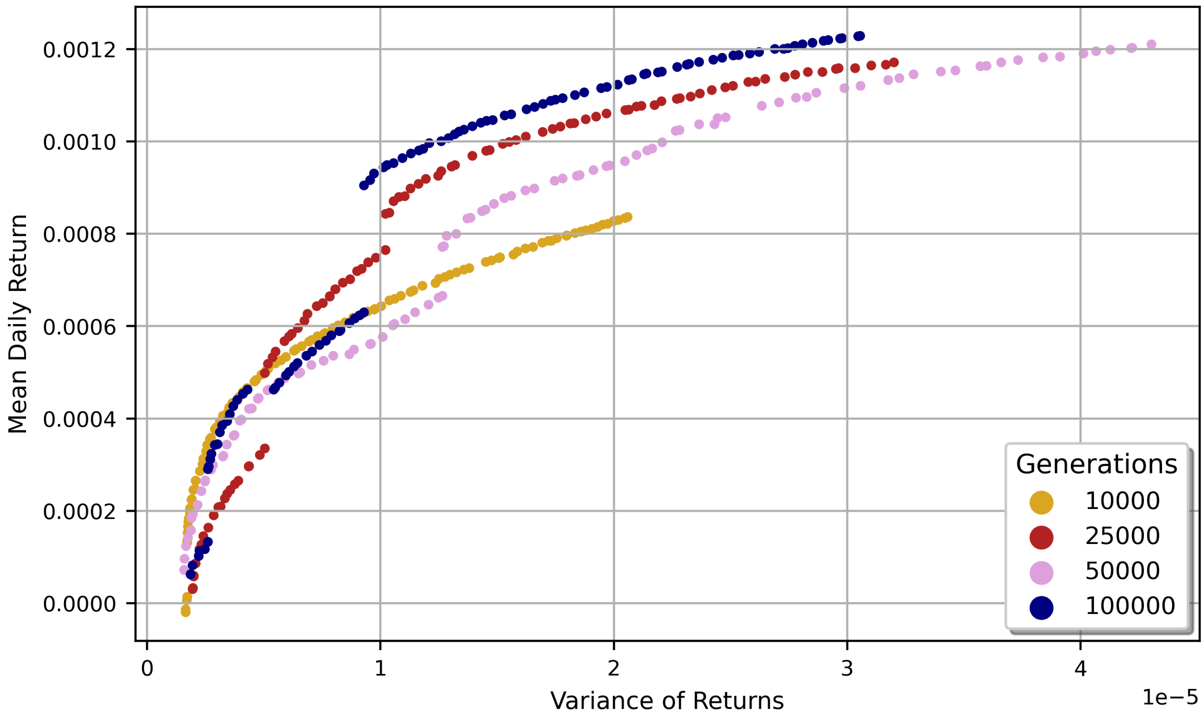

Figure 9: Impact of number of generations on Pareto front formation. Cardinality K = 5, Maximum Weighting 0.3

With the two constraints together, no such clear relationship exists and it is apparent that other factors are present which will influence Pareto front formation in optimisation tests conducted with a range of function evaluations that is bounded by what may be deemed practical given the available resources of time and hardware for an applied model such as we have proposed.

4.3 Impact of Constraints

As the two primary constraints which will be applied to our model are key to the construction of a realistic portfolio of futures contracts, it is essential to analyse their impact upon the objective function results that may be attained and the corresponding Pareto front formed.

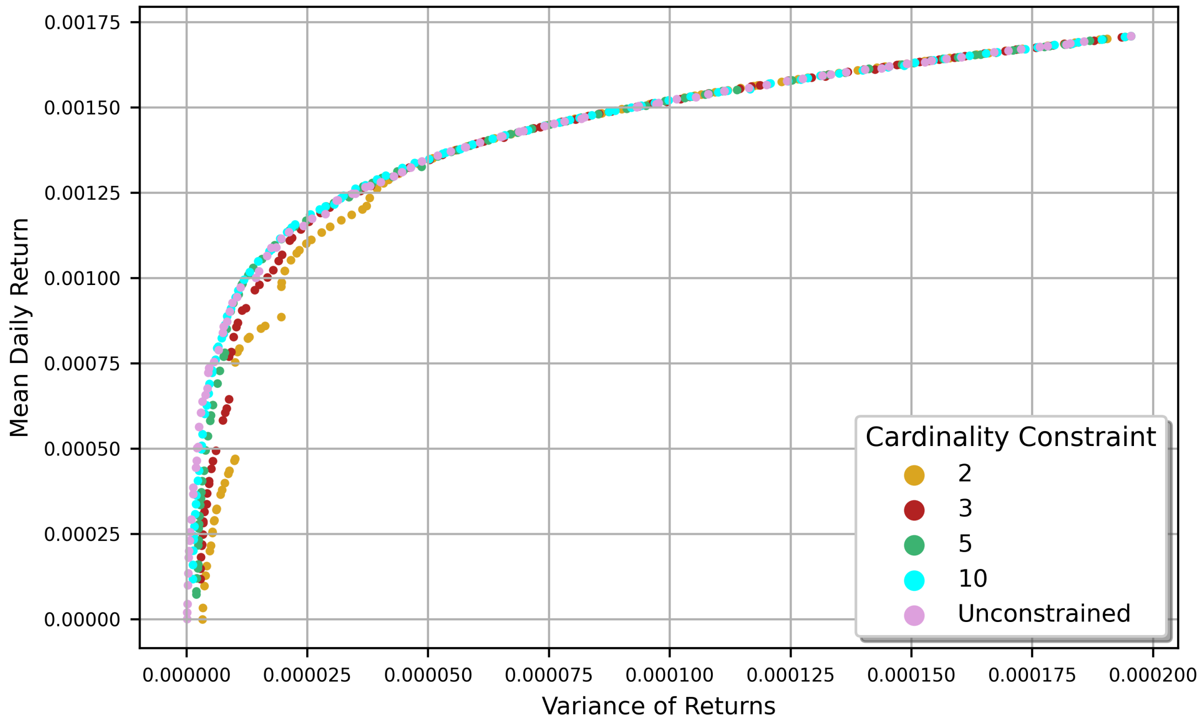

Figure 10: Impact of cardinality constraint upon Pareto front formation

Figure 10 illustrates the impact of varying cardinality constraints upon Pareto front formation with a series of tests conducted using a population of 100 for 50 000 generations. The previously described procedure of initial sampling is used. It is apparent that the most prominent impact of the cardinality constraint occurs among the solutions at the lowest levels of variance. Beyond a daily variance of approximately 5e-5, all populations converge with the unconstrained Pareto front and there is negligible distinction between their structures thereafter. Absent an additional constraint to control the weighting applied to particular positions, the portfolio solutions of the different populations are, regardless of cardinality, very similar beyond a certain threshold of variance. At the maximum end of return and variance, we will find portfolios made up almost entirely of the asset with the greatest mean return value. The unconstrained portfolio will allocate 100%

of its weighting, while those with cardinality constraints will apply the minimum allocation to other assets to fulfil the constraint function as structured. From a practical standpoint, and with respect to the objective functions, these positions will be so small as to be of negligible consequence.

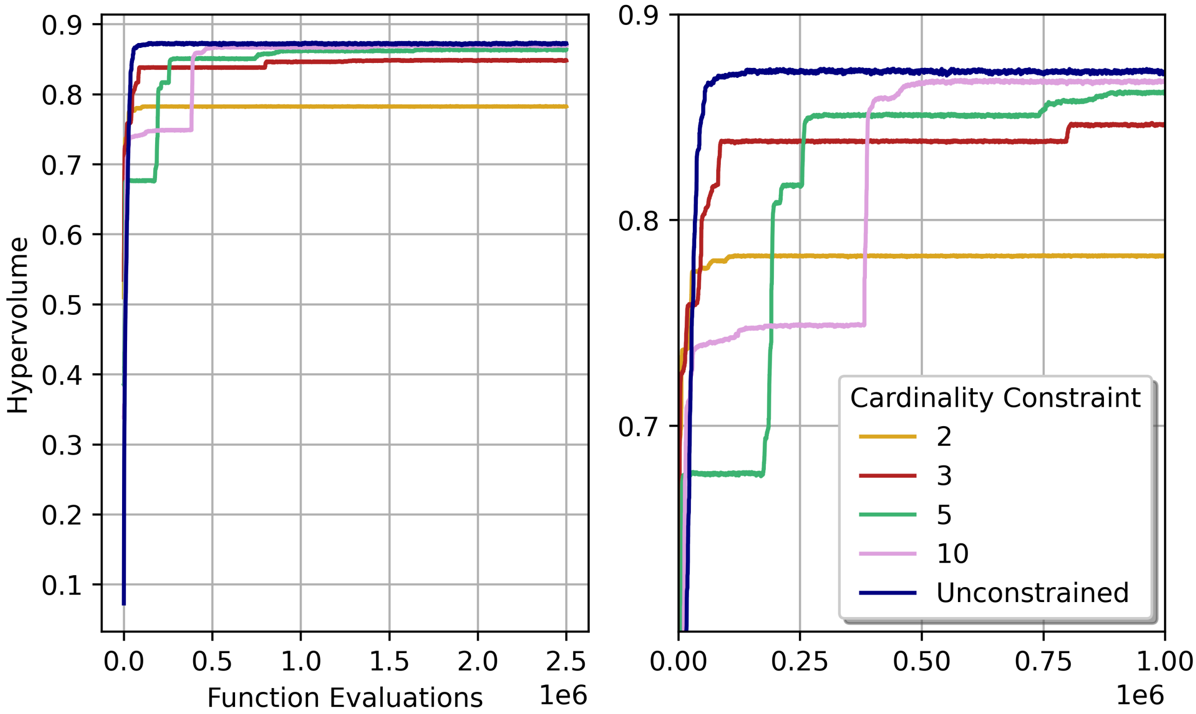

Figure 11: Hypervolume growth with varying cardinality constraints. Right:detailed view of area of growth within hypervolume range of 0.6 to 0.9 up to 1e6 function evaluations

The effect of different cardinality constraints upon hypervolume growth is illustrated in Figure 11. Provided with an initial population of feasible solutions, the tests conducted with cardinality constraints very quickly converge upon a hypervolume near the maximum they may ultimately achieve, with small additional gains sometimes taking place in rapid jumps at later points up to approximately one million function evaluations. The test in Figure 11 is conducted with a population of 100 for 25 000 generations. By the point at which the visualisations are bounded along the x-axis, all tests have reached a plateau and no further growth in hypervolume is observed out to a total of 5 000 000 function evaluations.

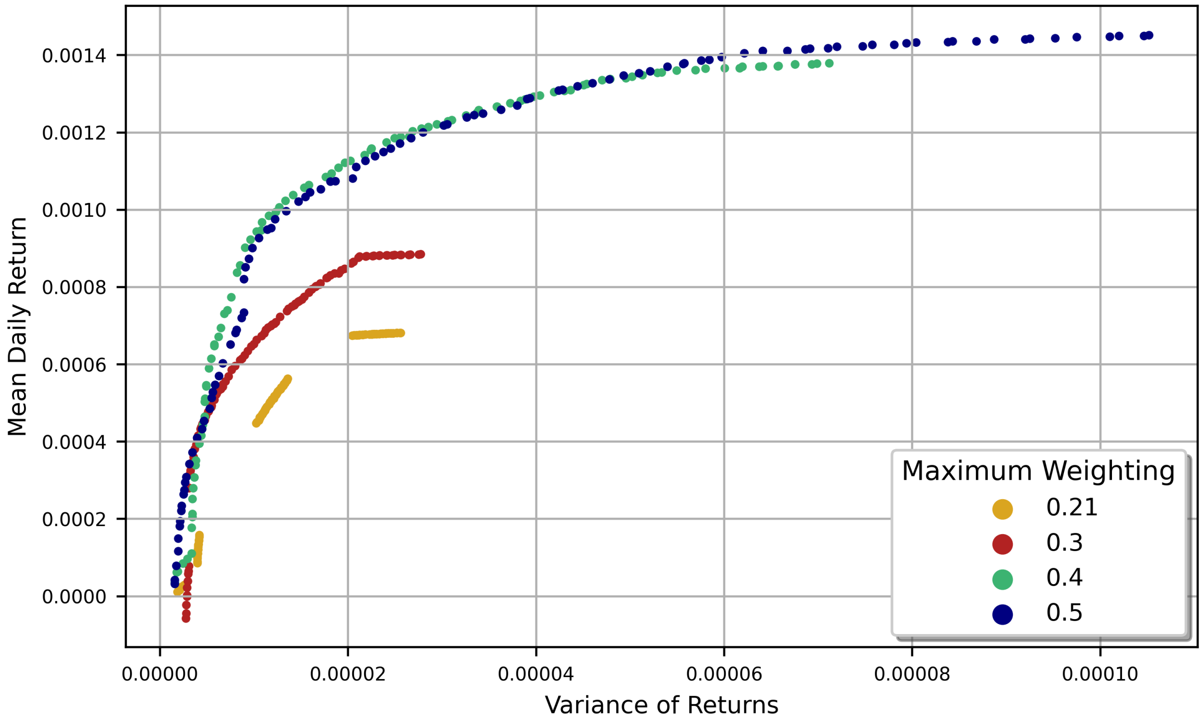

Figure 12: Impact of maximum weighting constraint on individual positions upon Pareto front formation. K = 5, Population 100, Generations 100 000

Optimisations conducted using a cardinality and maximum weighting constraint in conjunction produce a much greater degree of discontinuity across the Pareto front, growing more pronounced as the maximum weighting falls. Figure 12 illustrates the impact of a varied maximum weighting constraint upon a series of tests conducted with a population of 100 for 100 000 generations with a cardinality constraint of K=5. When combined, the product of the maximum weighting and cardinality constraints must be equal to or greater than one. Otherwise, of course, the generation of feasible solutions will be impossible as any solution that fulfils both constraints will achieve a total weighting of less than one and the normalisation procedure will push some weightings above the maximum. As a result, the closer this product is to one, the more restrictive the search space of feasible solutions is. Thus, we observe the more fragmented pattern of solutions in the above figure in which a cardinality constraint of K=5 is combined with a maximum weighting of only 0.21, leaving a very narrow feasible solution space.

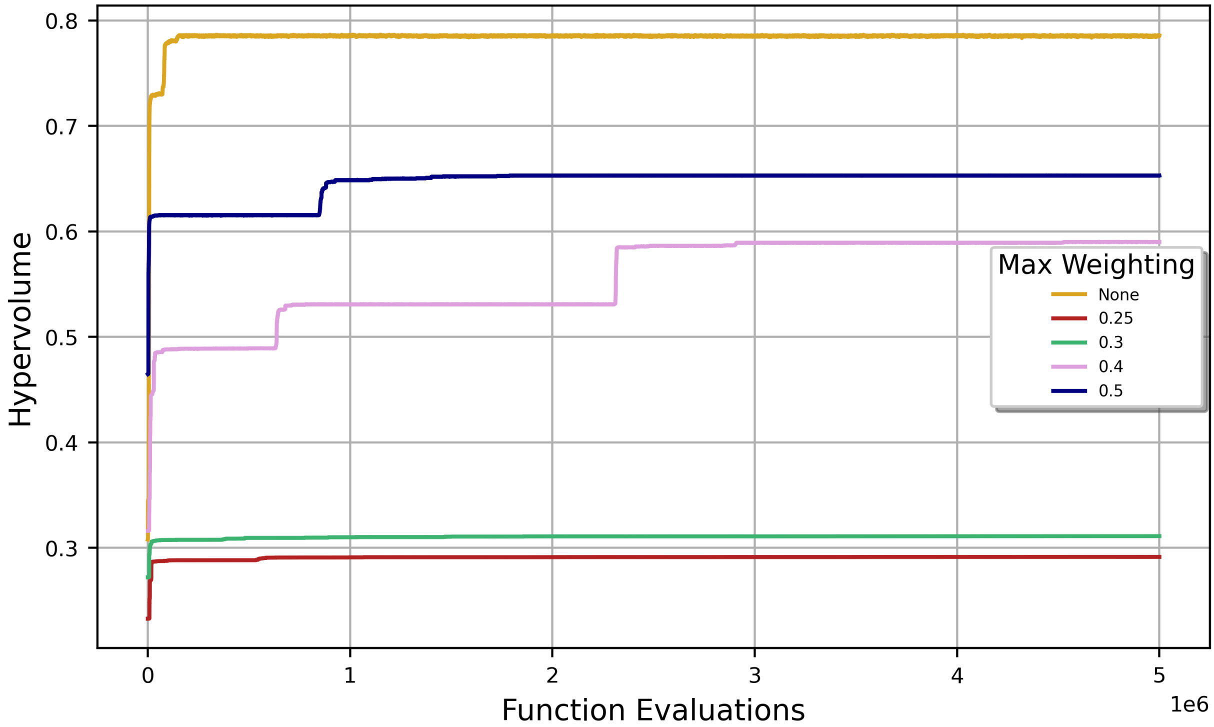

Figure 13: Hypervolume growth under varying maximum weighting constraints upon individual positions. K = 5, Population 100

The change in peak hypervolume as maximum weighting constraints are eased exhibits an expected pattern of increase. The tests in Figure 13 are conducted with a population of 100 and a cardinality constraint of K=5 out to a total of 5 000 000 function evaluations. Growth, again, is initially rapid under the effect of having provided an initial sample of feasible solutions. Following this, further increase occurs in a series of rapid jumps as new non-dominated solutions are found. The number of function evaluations required in order to reach a peak level of hypervolume is greater in the case of the combined constraints relative to cardinality alone given the increased complexity of the search space. In Figure 13, all tests have reached a hypervolume plateau by the point of 3 000 000 function evaluations, most having done so much earlier. As before, the majority of additional hypervolume achieved among tests with less restrictive maximum weighting constraints occurs on the basis of their ability to generate higher variance solutions, with substantial allocation to the assets with greatest mean return and variance. A market participant whose risk preferences are lower will be selecting a portfolio from the lower ranges of the Pareto front where there is significant overlap between the different maximum weighting tests. For these overlapping solutions which share very similar objective values, but represent portfolios with much different levels of concentration, those with more stringent maximum weighting constraints may provide a lower level of risk in reality due to being less concentrated among the most outlying assets which are more prone to estimation error.

4.4 Impact of Sampling and Normalisation

The previously described procedures for normalisation of our decision variable vectors and the sampling procedures used to generate initial populations that begin their search from a starting point of feasible solutions have been tested to assess their impact upon the efficiency of the test runs.

In Figure 14, we observe the hypervolume growth and Pareto front formation for an optimisation conducted with a population of 100 and cardinality constraint of K=2 for a total of 50 000 generations. One of the tests is conducted using a randomly sampled population and the other using our initial feasible sampling procedure. Both optimisations are conducted using a relaxation threshold of 1e-7 in contrast to the level of 1e-10 which is applied in most of our tests in order to allow the random sampled test to converge towards satisfaction of the cardinality constraint and generate a set of feasible solutions.

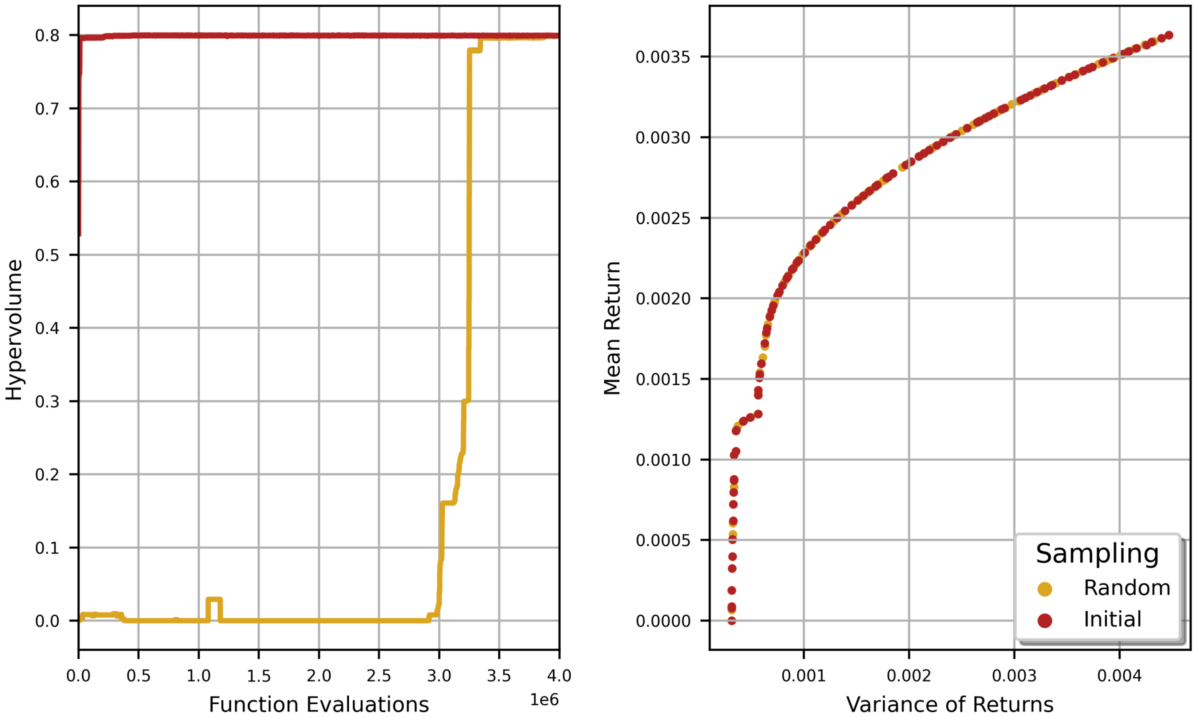

Figure 14: Hypervolume expansion and Pareto front formation for random vs. initial feasible sampling. Cardinality K = 2

The population with initial sampling very quickly converges upon a peak level of hypervolume and produces a well formed Pareto front with the discontinuity typical of small cardinality constraints. When random sampling is used, over 60% of the total number of function evaluations are required before the solutions reliably converge into the feasible space. Ultimately, the population of both tests are able to reach a set of solutions along a nearly identical Pareto front, and any further gains in hypervolume will be minimal thereafter even in the case of tests run for a much larger number of generations. Under such a configuration with a cardinality constraint alone, both an initial or random sampled optimisation test will arrive at approximately the same set of solutions, but the increased efficiency of the initial sampling procedure is evident.

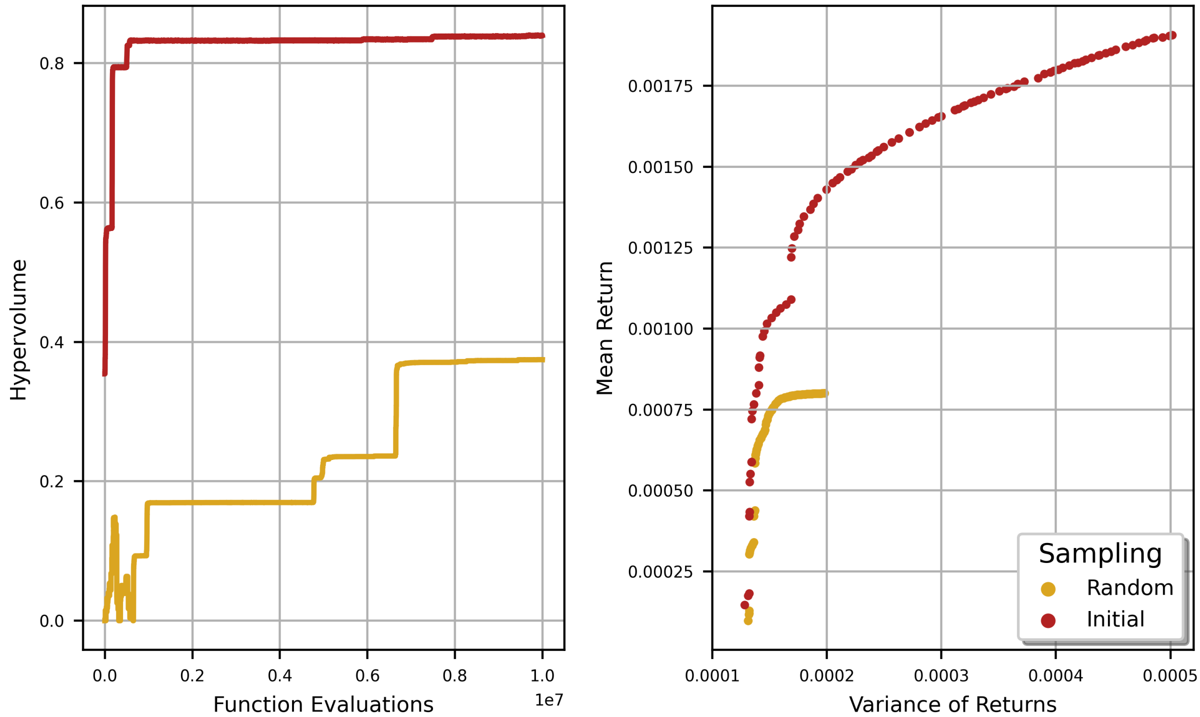

Figure 15: Impact of random vs. initial feasible sampling upon hypervolume and Pareto front formation. Cardinality K = 5, Maximum Weighting 0.3

In conjunction with a maximum weighting constraint, the positive impact of initial sampling upon Pareto front formation is similarly pronounced to the cardinality constrained tests. Figure 15 depicts such a test carried out with a cardinality constraint of K=5 and a maximum weighting boundary of 0.3. When random sampling is relied upon to generate feasible solutions for a test conducted with a population of 100 for 100 000 generations, a greatly reduced Pareto front is formed. The initialised population, outside of some of the small discontinuities which result along the Pareto front at the lowest variance levels, completely dominates the set of solutions produced via random sampling. Peak hypervolume for the population utilising initial sampling is significantly higher than in the random sampled test, and reaches a level very close to this within less than 10% of the total function evaluations carried out.

4.5 Impact of Outliers

One of the well known pitfalls that portfolio optimisation procedures are vulnerable to is the significant impact that outlier data can have upon allocations. This has led to some works exploring the notion that the actual impact of the optimisation process, if no measures are taken, could just as accurately be described as error maximisation (Kritzmen, 2006). The dataset for this project, as described previously, contains certain outliers of significant magnitude which provide an instructive basis from which to investigate this effect.

When assessing the entire history encompassed by our dataset, we observe that the mean daily return of a short WTI crude oil and long lumber contract position are respectively 0.0036% and 0.0013%, the two most significant outliers within the return vector with high mean return objective values.

The difference in structure of the resulting Pareto front produced by the removal or inclusion of these two assets with significantly outlying return and variance values is acute as illustrated in Figure 16. The pictured test compares both an unconstrained optimisation and one incorporating a cardinality constraint of K=5 and a maximum weighting of 0.3 for each asset set. The tests are all run for 100 000 generations with a population of 100. When the 40 asset set including the crude oil and lumber contracts is used, the set of non-dominated solutions is able to expand to a much wider range of return and variance. The 38 asset portfolio, in contrast, is limited to a very narrow range, particularly of variance, simply by the exclusion of these two assets from the selection pool.

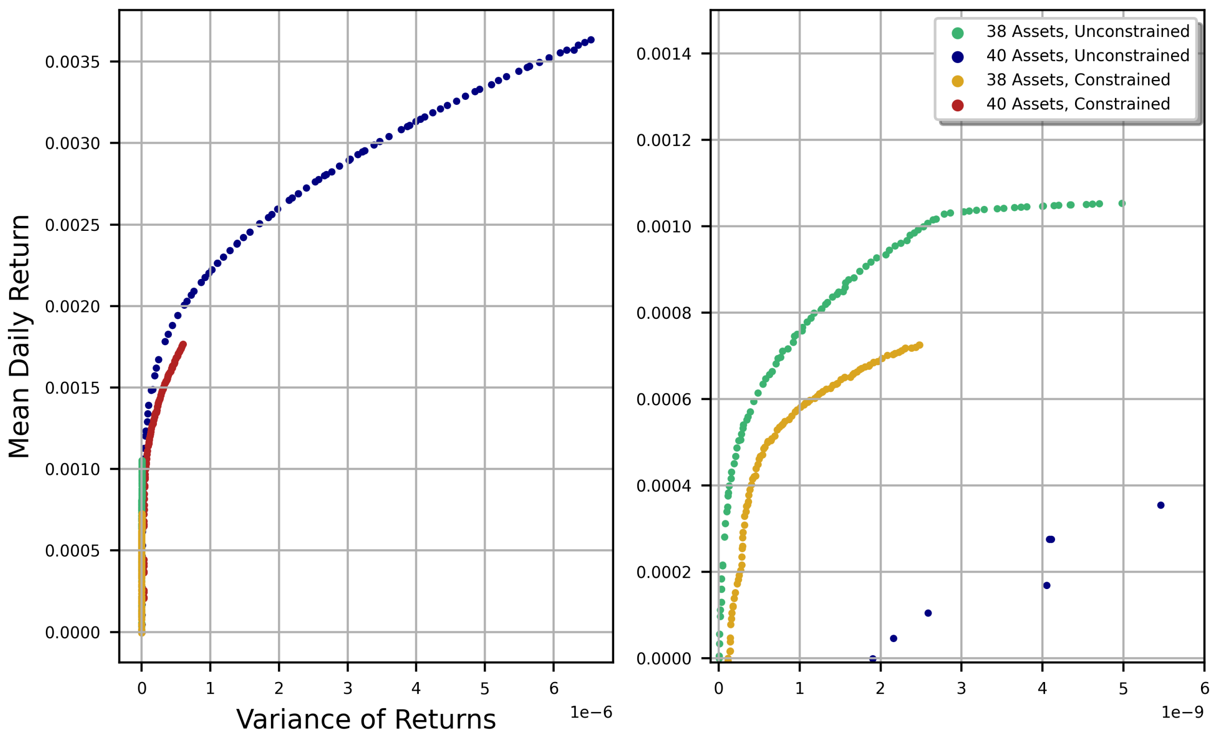

Figure 16: Impact of outliers upon Pareto front formation. Right: axis scales limited to provide detail view of 38 asset tests.

When the portfolio compositions are examined in detail, we observe the difference in degree of portfolio concentration that occurs when such outliers are present or excluded. Within the 40 asset unconstrained optimisation results, mean allocation among the population of solutions to a short position in the WTI crude oil contract is 48.7% while a long lumber position garners 20.9%. Thus, we have almost 70% of the average portfolio concentrated in only two positions, a result which would certainly not satisfy any common sense notion of being well diversified. The results of a cardinality and maximum weighting constrained test using the same data again produces a significant proportion of the allocated portfolios weighted heavily towards these two assets, limited by the bounds imposed by the constraints. This constraint configuration does produce a more diverse range of portfolio allocations containing, on average through a series of tests, approximately 25% of portfolio solutions with no allocation to either one or both of these outlier assets.

Importantly, these portfolio allocations are generated via the use of the five years of return history from within the dataset, from 2017 through the end of 2021, to estimate mean return and covariance. Practical implementations of a momentum based strategy will nearly always rely upon a shorter historical return data window and, of course, the impact of extreme outliers in a smaller set of data used for parameter estimation will only be magnified. The tests performed in Figure 16 provide a stark example of the significant impact of outliers on portfolio allocation, with the addition or removal of only two assets from a total of 40 having a profound effect on Pareto front formation and portfolio optimisation results.

4.6 Analysis of Allocated Portfolio Performance

It is, of course, essential to analyse the actual performance of a set of example portfolios allocated by our model in order to assess its viability as a basis for implementing a short term futures trading strategy. This shall be conducted utilising a series of tests which combine cardinality and maximum weighting constraints with the intention of producing portfolios that will be practical to implement.

The tables below display the mean excess percentage returns realised among all generated portfolio solutions relative to the estimated values for a series of optimisation tests conducted with varying parameters and utilising a 40 asset long/short selection pool. The results are calculated forward from the beginning of each year in the dataset based on optimisations which use a data history for parameter estimation of 3, 6, 9, or 12 months. Results are calculated on the basis of a short term portfolio holding period of either 10 or 20 trading days.

Tables 6 through 13 display the mean excess percentage return realised among all portfolio solutions over a given holding period relative to the mean return of the solutions generated via the corresponding optimisation parameters. The optimised and actual daily returns are multiplied by the holding period and the results represent the level of excess returns over that period.

Table 6: Excess returns: K = 5, Maximum Weighting 0.3, 10 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	1.32%	0.3%	1.16%	1.6%
2019	-10.03%	-3.48%	-0.93%	-4.43%
2020	-2.31%	-0.99%	0.01%	1.83%
2021	1.31%	-0.61%	-9.38%	-6.19%
2022	-1.98%	-4.47%	-3.74%	-1.71%

Table 7: Excess returns: K = 5, Maximum Weighting 0.3, 20 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	0.38%	0.81%	2.06%	0.99%
2019	-11.22%	-7.98%	-3.6%	-6.83%
2020	-4.94%	-2.69%	2.27%	1.33%
2021	-1.47%	-1.37%	-5.43%	-0.47%
2022	-5.88%	-1.63%	-2.73%	-2.17%

Table 8: Excess returns: K = 8, Maximum Weighting 0.25, 10 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	1.84%	1.19%	0.87%	1.89%
2019	-7.21%	-4.24%	-1.66%	-2.46%
2020	-1.43%	-0.33%	0.26%	1.48%
2021	0.67%	0.13%	-5.6%	-3.51%
2022	0.08%	-4.26%	-2.39%	-2.59%

Table 9: Excess returns: K = 8, Maximum Weighting 0.25, 20 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	-1.77%	-0.69%	-0.32%	2.07%
2019	-10.61%	-5.51%	-5.31%	-4.13%
2020	-0.51%	-4.4%	2.77%	-1.75%
2021	-1.17%	2.5%	-7.04%	-11.23%
2022	-4.71%	-3.61%	-0.59%	3.54%

Table 10: Excess returns: Unconstrained, 10 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	2.19%	0.08%	0.64%	1.36%
2019	-9.51%	-6.47%	-2.68%	-4.0%
2020	-2.27%	-0.53%	4.45%	4.79%
2021	0.95%	-4.32%	-18.58%	-18.91%
2022	1.85%	-8.02%	-7.51%	-4.29%

Table 11: Excess returns: Unconstrained, 20 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	-0.76%	-4.02%	-1.69%	-2.21%
2019	-14.04%	-14.89%	-9.95%	-6.83%
2020	-10.19%	-3.91%	3.71%	4.45%
2021	-1.37%	5.63%	-27.42%	-27.04%
2022	-19.99%	-10.17%	-4.7%	-2.39%

Table 12: Excess returns: Cardinality constraint only: K = 5, 10 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	2.6%	0.43%	0.86%	1.59%
2019	-8.88%	-6.6%	-4.25%	-3.84%
2020	-3.87%	1.98%	4.2%	4.81%
2021	0.54%	-3.37%	-18.84%	-19.93%
2022	1.97%	-7.61%	-7.52%	-4.69%

Table 13: Excess returns: Cardinality constraint only: K = 5, 20 day holding period

Year	3 Months	6 Months	9 Months	12 Months
2018	-0.53%	-3.54%	-1.45%	-1.99%
2019	-14.29%	-15.13%	-10.84%	-7.86%
2020	-10.23%	-1.3%	3.82%	4.08%
2021	-0.33%	5.39%	-29.64%	24.66%
2022	-19.04%	-10.11%	-6.23%	-3.08%

Over the entire range of tests, the mean daily excess returns for the portfolios produced by our model relative to the estimated results is -0.26%. No clear pattern emerges with regard to the impact of length of historical data used to estimate mean return and covariance. Categorised accordingly, the mean daily excess return ranges from a low of -0.3% with 9 months of data to an equal high of -0.22% for 6 or 12 month periods.

Among the eight configurations of optimisation constraints and holding periods tested, all achieve excess returns within the sample period which underperform the average level of estimated return among the corresponding optimised portfolio sets. When compared, it is however observed that the unconstrained optimisations, as well as those with a cardinality constraint alone, underperform those with a combined cardinality and maximum weighting constraint to a substantial degree. The unconstrained and cardinality only tests yielded excess returns of -0.36% on average compared with -0.15% for the tests with a cardinality and maximum weighting constraint. The difference between the use of a 10 or 20 trading day holding period among the tests appears to be minimal. The mean daily return among the tests categorised accordingly is -0.25% and -0.26% respectively.

4.7 Structure of Portfolio Allocations

Lastly, in Figures 17 and 18, we examine a series heat maps which highlight the level of allocation to each variable within a population of 100 in a selection of portfolios generated within the optimisation tests discussed in section 4.6. Each column corresponds to one of the long or short positions that may be allocated to amongst the set of 40 assets, totalling 80 decision variables. Rows in the plots represent the 100 individual portfolio solutions generated by the total population for each test.

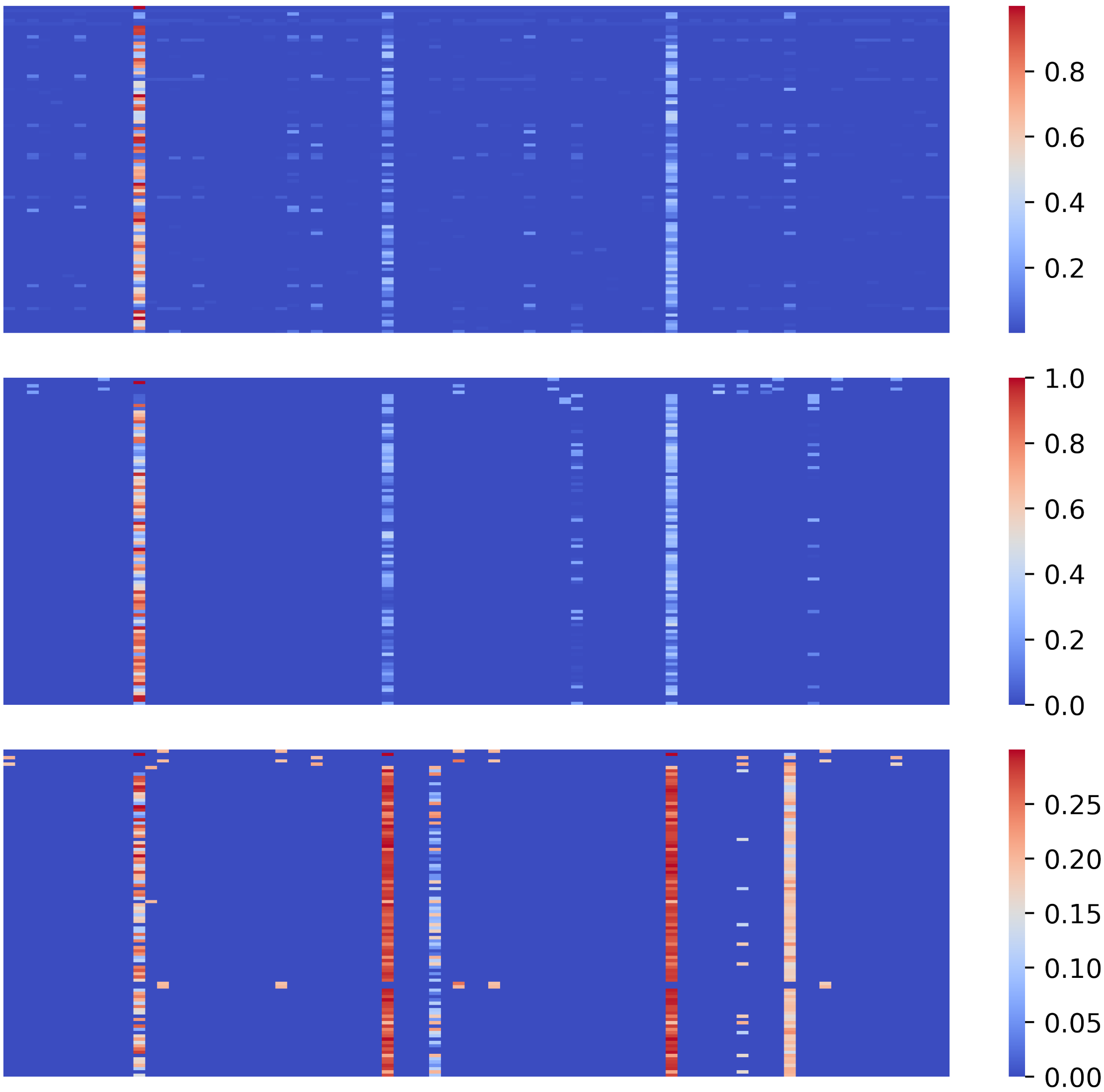

Figure 17: Portfolio allocations among optimisation population. From top: Unconstrained, Cardinality constraint only K=5, Cardinality K=5 and Maximum Weighting 0.3. 9 month historical data period Apr-Dec 2020.

The plots in Figure 17 showcase the results of the tests conducted using nine months of historical data previous to the beginning of 2021 with a ten trading day holding period. This date range coincides with the period of high market volatility following the beginning of the COVID-19 pandemic. For the visualised tests using this historical data period, realised excess return relative to optimised results are the lowest among all historical date ranges used to generate data for parameter estimates. The similar concentration of weightings between each test is apparent, with three assets: short WTI crude oil, long lumber, and long gasoline, being included in the majority of all portfolios amongst them. The large number of small positions, comprising less than 2% of the total portfolio in the unconstrained test, are largely shifted into greater weightings of the most concentrated positions in the centre test which employs a cardinality constraint alone. When the cardinality constraint is paired with a maximum weighting constraint, it forces the allocation of a more substantial weighting to the five assets that comprise the portfolio. There is still limited diversity among the total population and the majority of portfolios are made up of a variation of the same five positions.

As a contrast, observing Figure 18, we may examine the results for tests conducted with the same configurations as in Figure 17, but for the date range in which the realised excess returns most closely matched the optimised values. Figure 18 shows the allocations for tests conducted with six months of historical data previous to the beginning of 2018. In the case of the unconstrained test, the average weighting of the position with heaviest concentration among the portfolios declines from 54.12% to 43.33% relative to the corresponding test in Figure 17. A similar impact is seen in the cardinality only test, the top position falling from an average weighting of 55.48% to 45.25%. The combined cardinality and maximum weighting constrained test continues to allocate its largest weightings to a similar set of positions as in the other tests, but does include a greater number of portfolios which incorporate an alternative set of positions.

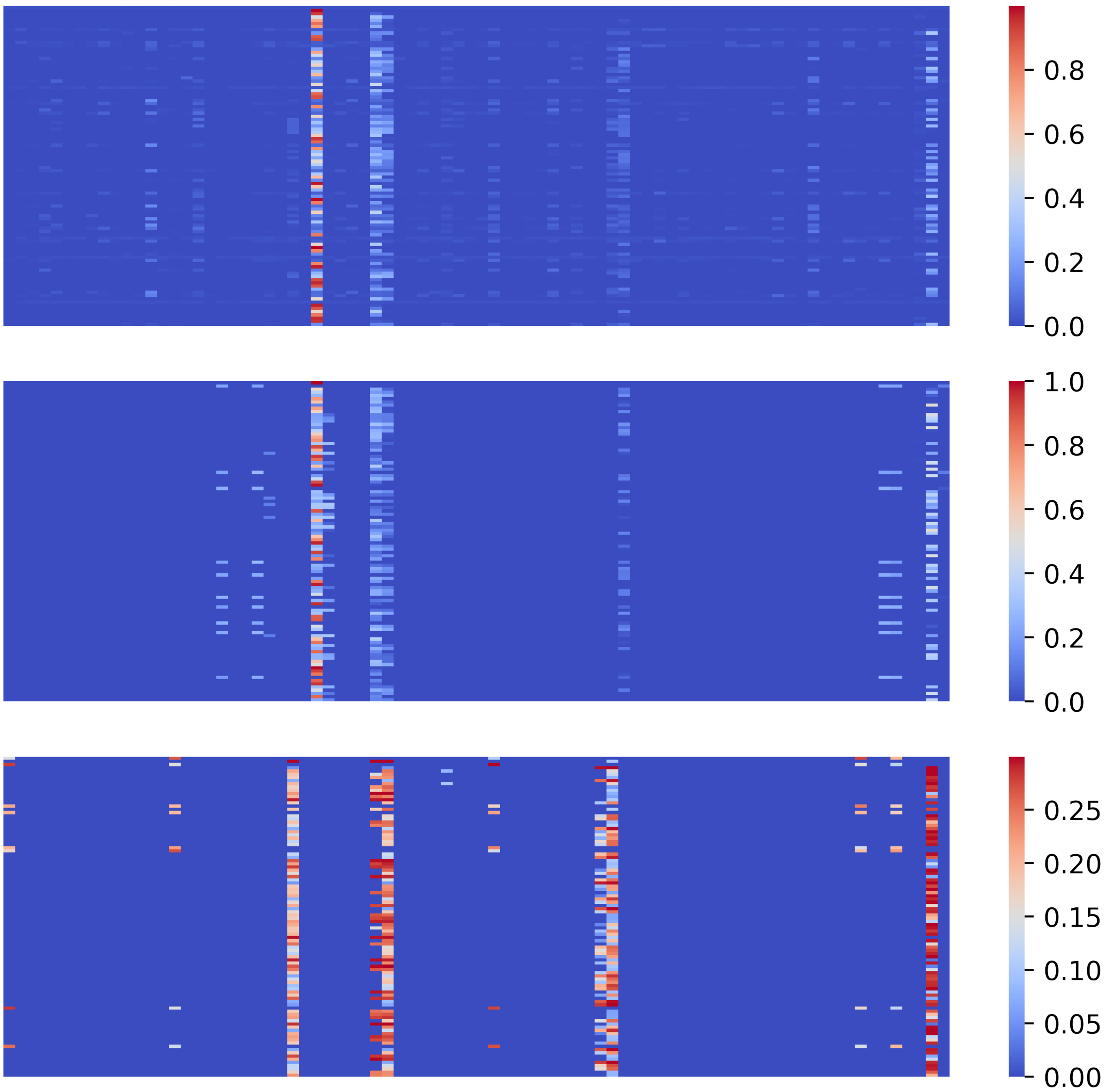

Figure 18: Portfolio allocations among optimisation population. From top: Unconstrained, Cardinality constraint only K=5, Cardinality K=5 and Maximum Weighting 0.3. 6 month historical data period Jul-Dec 2017

5. Discussion of Results

Analysis of the distribution of daily returns within our dataset for the time periods covered quite clearly supports a conclusion of non-normality, in line with the majority of findings for the distribution of asset price returns. One commonly cited challenge facing the application of the Markowitz model of portfolio optimisation is its assumption of normality in the distribution of the underlying asset price returns across time. In reality, the majority of samples of returns for a wide range of assets have frequently been observed to exhibit a much greater propensity to contain extreme outlying values than would be characteristic of a normal distribution, described statistically as leptokurtosis (Chang et al., 2000). Colloquially and within financial terminology, this is often referred to as a distribution with “fat tails”.

A cursory examination of the results contained in Table 5 immediately draws one’s attention to some very extreme values. As described earlier in the context of covariance matrix estimation, the precipitous drop in price of the WTI crude oil contract in April of 2020 represented a decline in excess of 500% in a single day, as represented by our continuous futures calculation methodology.

Kurtosis is often described as a measurement related to the ‘peakedness’ of a dataset in contrast to a normal distribution. However, Westfall (2014) argues that such interpretations are in fact inaccurate and kurtosis is descriptive only of the presence of outliers within a distribution, or the propensity to produce them in the case of a probability distribution. The typical method for calculating kurtosis was outlined by Pearson (1905) as

$$Kurtosis[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \quad . \tag{8}$$

Kurtosis, as computed via the Scipy Stats module in Python, is normalised such that it is equal to zero for a normal distribution. In the case of our full data history taken as a whole, or divided into the years which comprise it, the resulting calculations are all positive and for 2020 to an extreme degree. We may therefore confidently conclude that our return distributions, in accordance with the commonly stated characteristic of asset price return distributions, is decidedly leptokurtic in its structure.

With regards to the fundamental algorithm parameters of population size and number of generations, a pattern emerges of a trade-off between efficiency of generating solutions and peak results, as measured by hypervolume. The convergence of smaller populations upon the ultimate structure of the Pareto front they will form occurs more rapidly, but at the cost of a lower maximum hypervolume and a less complete coverage of the front which will result in a more granular set of portfolios available for selection. For example, a population of 25 will result in much greater gaps in risk/return objective values from one portfolio along the front to the next in relation to one generated with a larger population. This is subject to diminishing returns as when population increases the differences between adjacent portfolio composition will become more and more negligible. In Deb et al. (2002), the primary published work introducing the NSGA-II algorithm, their benchmark tests were carried out using a consistent population of 100. In our own tests, given the structure of the modelled problem, this value seems to represent a successful compromise between the aforementioned factors. We have therefore chosen to use this value as a standard parameter for the majority of all tests conducted, and in particular those used to calculate the excess return potential of our model.

When performing unconstrained or cardinality constrained optimisations, the impact of increasing the number of algorithm generations has the effect of progressively increasing hypervolume of the Pareto front up to a plateau, after which little further increase occurs. Some of the conducted tests which incorporate a maximum weighting constraint in addition to cardinality produce paradoxical results. Certain tests, such as the example in Figure 9, show optimisations carried out for a smaller number of generations having portions of the Pareto front which dominate segments of others run for a larger number. Within the scope of application for the relatively simple model we have employed using the Pymoo framework, where the capacity to generate solutions without excessively long computation times is a priority, we would limit most tests to those which can be performed in less than approximately four hours using typical consumer hardware. Within such a boundary, it seems likely that given the increased complexity of the objective landscape for optimisation problems with a cardinality and maximum weighting constraint applied in combination, there will be a stochastic element in the results generated. Conducting a series of tests and extracting the non-dominated solutions from among them may be a strategy to produce superior results in a more efficient manner.

The impact of our normalisation and initialisation procedures for constraint handling have been demonstrated in the performed tests to achieve a major increase in efficiency in the generation of solution sets. In particular, the process of generating solutions for a combined cardinality and maximum weighting constraint problem in the absence of an initial population of feasible solutions is extremely challenging for the algorithm to perform and our chosen procedure allows rapid hypervolume expansion and the generation of portfolio allocations which avoid the excessive weightings characteristic of an optimisation that is unconstrained or uses a cardinality constraint alone. The absence of a well structured and problem specific initial sampling procedure for the generation a population of solutions within the NSGA-II algorithm may ultimately achieve the same results in terms of hypervolume and Pareto front structure, but will often result in a greatly increased computational burden.

The most significant gains in hypervolume among tests conducted with constraints that allow larger numbers of assets to be held and in greater weightings, is concentrated among the set of solutions that fall upon the highest variance ends of the Pareto front. However, as these portfolios are typically composed of varying weightings of a small set of assets with extreme parameter values, and the corresponding probability of estimation error, higher peak levels of hypervolume can be deceiving when attempting to assess the relative quality of different Pareto fronts. When more moderate return and variance preferences are held, the limitations imposed by more stringent constraints are less pronounced, and as mentioned previously, may also provide portfolio allocations that are of significantly lower risk along facets which are not capable of being recognized within a simple bi-objective model.

The tests conducted which compare the impact of removing the two assets, crude oil and lumber, comprising the most significant set of outlying observations within the dataset, very clearly illustrate their impact. When included, a large segment of the Pareto front at the highest variance levels, which exceeds the range of the asset set in which they are excluded, is comprised to an overwhelming degree of portfolios heavily weighted to these two positions from a total of 80. Statistical techniques to mitigate this effect such as Ledoit Wolf shrinkage have a limited effect which can be seen in the structure of the covariance matrix generated. The tests conducted do however demonstrate that when outliers of such significant magnitude are present, such procedures are not sufficient to prevent outsized allocations. As a result, a proper set of constraints to enforce a chosen level of diversification within generated portfolios is crucial to avoid producing a set of optimised solutions whose composition is concentrated beyond a level which a rational investor's intuition would deem acceptable.

In testing the actual performance of our allocated portfolios using the chosen parameters and constraints, intended to achieve a balance between efficient Pareto front formation and practical utility, we find that our model is prone to overestimation of expected returns within the sample period. A very crucial consideration, which is typically absent from the intensive technical analysis which accompanies the portfolio optimisation process, is the underlying philosophy behind the assumptions of the mean-variance model. It assumes that assets are likely to repeat the performance of the historical time period used to estimate optimisation parameters. Based on the results encountered in our testing, of a tendency of optimised portfolios to underperform the model values over a range of configurations and time periods, it may be the case that it is generally not a sufficient basis for predicting futures contract returns with a reasonable degree of accuracy. It is feasible that given this underperformance an alternative formulation for this set of assets which utilised an objective based upon mean reversion could potentially produce a superior set of out of sample results.

The examination of portfolio allocations reiterates the fundamental fact that the optimisation process possesses no concept of diversification outside of that which occurs via a minimisation of variance. In order to avoid the proverbial sin of placing all of one's eggs in one basket, it is essential to employ measures such as covariance matrix shrinkage and a proper set of constraints if one seeks to use an optimisation model as the basis for an applied investment strategy.

6. Conclusions

Given the scope available for the model construction and analysis conducted by this project, it is necessarily limited in ability to replicate all of the important factors which would need to be taken into consideration before employing such an investment strategy. Further refinements to the model could be performed in order to provide a more effective optimisation platform. As the relationship between historical return and future returns with respect to futures contracts appears to be inconclusive based on our results and other research, potential future modified versions may seek to employ alternative measurements for risk and return instead of those used in this project such as Conditional Value at Risk or a minimum variance model. A number of previous works, examples including, Chang et al. (2009), Grootveld & Hallerbach (1999) and Lim et al. (2011), have explored the impact and suitability of using these and other measurements that differ from the traditional mean-variance model.

Additionally, the varying margin levels required to open a position for different futures contracts can have a tremendous impact on the true return and variance for different assets relative to the amount of capital allocated to those positions. A model which incorporates these factors would add significant complexity to the structure and likely require the use of a multi-period model with dynamic parameters. Such additional functionality would greatly increase the realism of the model when applied to this class of assets.

The impact of outlying data upon the optimisation process, and some of the procedures available to mitigate its impact have been tested and indicate that the high levels of variance that are possible within the returns of financial assets such as futures contracts present a significant challenge to any optimisation strategy. Extreme market conditions have a significant effect on results and beyond a certain level statistical tools such as the utilised shrinkage procedure may not be capable on their own of lessening their impact to a satisfactory degree. Alternative strategies for formulating the key estimates for the portfolio optimisation may need to be taken. The price movement of futures contracts, as evidenced in our analysis, can take on a level of volatility that is significantly in excess of what is typical for major publicly traded equities. Additionally, the example provided in 2020 of negative pricing provides clear warning of the fact that potential losses are not bounded by zero in the case of long positions and they face the same theoretical risk of catastrophic loss as short positions. It is clear that portfolio optimisation can be structurally applied to futures contracts as an asset class. It is, however, essential to take note of the results and data explored in this project which indicate that such a model cannot simply be applied in the absence of additional precautions which one might omit in the case of more traditional portfolio optimisation exercises that aim to produce equity portfolios for a longer term buy and hold strategy.

The most encouraging results in our tests were produced by the use of a cardinality and maximum weighting constraint in conjunction. Further exploration into procedures or modifications which could minimise the stochastic nature of results generated by these constraints under our model, and maximise the ability to generate consistent Pareto fronts without employing much more powerful computational resources or time would certainly be a valuable addition to the research we have conducted.

Evolutionary algorithms continue to be a growing area of research, drawing from the expertise of researchers in many technical disciplines. A great opportunity exists for their wider adoption among finance researchers who may employ them in optimisation projects which place significant focus upon assessing the financial implications of results as a primary focus. Multi-objective optimisation tools such as Pymoo provide tremendous versatility and opportunity to experiment and refine models which aim to be of practical utility.

List of References

Adhikari, R., Putnam, K. J., and Panta, H. (2020) “Robust Optimization-Based Commodity Portfolio Performance”, *International Journal of Financial Studies*, 8(3), p. 54. doi: 10.3390/ijfs8030054

Alander, J.T. (1992) “On Optimal Population Size of Genetic Algorithms”, In *CompEuro 1992 Proceedings computer systems and software engineering*, pp. 65-70. doi: 10.1109/CMPEUR.1992.218485

Anagnostopoulos, K.P. and Mamanis, G. (2011) “The mean–variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms”, *Expert Systems with Applications*, 38(11), pp. 14208-14217. doi: https://doi.org/10.1016/j.eswa.2011.04.233

Behr, P., Guettler, A., and Miebs, F. (2013) “On portfolio optimization: Imposing the right constraints”, *Journal of Banking & Finance*, 37(4), pp. 1232-1242. doi: https://doi.org/10.1016/j.jbankfin.2012.11.020

Bienstock, D. (1996) “Computational Study of a Family of Mixed-Integer Quadratic Programming Problems”, *Mathematical programming*, 74(2), pp.121-140. Doi: 10.1.1.47.2616

Black, F. (1976) “The Pricing of Commodity Contracts”, *Journal of Financial Economics*, 3(1-2), pp.167-179. doi: 10.1016/0304-405X(76)90024-6

Blank, J. and Deb, K. (2020) “Pymoo: Multi-Objective Optimization in Python”, *IEEE Access*, 8, pp. 89497-89509. doi: 10.1109/ACCESS.2020.2990567

Carchano, Ó. and Pardo, Á. (2009) “Rolling over stock index futures contracts”, *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 29(7), pp.684-694. doi: 10.1002/fut.20373

Chang, T.J., Meade, N., Beasley, J.E. and Sharaiha, Y.M. (2000) “Heuristics for cardinality constrained portfolio optimisation”, *Computers & Operations Research*, 27(13), pp.1271-1302. doi: https://doi.org/10.1016/S0305-0548(99)00074-X

Chang, T.J., Yang, S.C. and Chang, K.J. (2009) “Portfolio optimization problems in different risk measures using genetic algorithm”, *Expert Systems with Applications*, *36*(7), pp. 10529-10537. doi: https://doi.org/10.1016/j.eswa.2009.02.062

Chiam, S.C., Tan, K.C. and Al Mamum, A. (2008) “Evolutionary Multi-objective Portfolio Optimization in Practical Context”, *International Journal of Automation and Computing*, 5(1), pp.67-80. doi: 10.1007/s11633-008-0067-2

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002) "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, 6(2), pp. 182-197. doi: 10.1109/4235.996017

Erb, C.B. and Harvey, C.R. (2016) “The Strategic and Tactical Value of Commodity Futures”, *Financial Analysts Journal*, 62(2), pp. 125-178. Available at: https://www.jstor.org/stable/4480745 (Accessed 8 August 2022)

DeMiguel, V., Garlappi, L. and Uppal, R. (2009) “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?”, *The Review of Financial Studies*, *22*(5), pp. 1915-1953. Available at: http://www.jstor.com/stable/30226017 (Accessed 15 August 2022)

Fuertes, A.M., Miffre, J. and Rallis, G. (2010) “Tactical allocation in commodity futures markets: Combining momentum and term structure signals”, *Journal of Banking & Finance*, *34*(10), pp. 2530-2548. doi: https://doi.org/10.1016/j.jbankfin.2010.04.009

Gorton, G. and Rouwenhorst, K. G. (2006) “Facts and Fantasies about Commodity Futures”, *Financial Analysts Journal*, 62(2), pp. 47-68. Available at: http://www.jstor.org/stable/4480744 (Accessed 8 August 2022).

Gorton, G.B., Hayashi, F. and Rouwenhorst, K.G. (2013) “The Fundamentals of Commodity Futures Returns”, *Review of Finance*, 17(1), pp. 35-105. doi: https://doi.org/10.1093/rof/rfs019

Grootveld, H. and Hallerbach, W. (1999) “Variance vs downside risk: Is there really that much difference?”, *European Journal of Operational Research*, 114(2), pp. 304-319. doi: https://doi.org/10.1016/S0377-2217(98)00258-6

Guerreiro, A.P., Fonseca, C.M. and Paquete, L. (2021) “The hypervolume indicator: Computational problems and algorithms”, *ACM Computing Surveys (CSUR)*, 54(6), pp. 1-42. doi: https://doi-org.plymouth.idm.oclc.org/10.1145/3453474

Huo, L., Kim, T.H. and Kim, Y. (2012) “Robust estimation of covariance and its application to portfolio optimization”, *Finance Research Letters*, 9(3), pp. 121-134. doi: https://doi.org/10.1016/j.frl.2012.06.001

Jacobs, B.I., Levy, K.N. and Markowitz, H.M. (2005) “Portfolio optimization with factors, scenarios, and realistic short positions”, *Operations Research*, 53(4), pp. 586-599. Available at: https://www.jstor.org/stable/25146895 (Accessed 14 May 2022)

Jacobs, B. I., Levy, K. N., and Markowitz, H. M. (2006) “Trimability and Fast Optimization of Long-Short Portfolios”, *Financial Analysts Journal*, 62(2), pp. 36-46. doi: 10.2469/faj.v62.n2.4082

Jagannathan, R. and Ma, T. (2003) “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps”, *The Journal of Finance*, 58(4), pp. 1651-1683. Available at: https://www.jstor.org/stable/3648224 (Accessed 26 June 2022)

Kalayci, C.B., Ertenlice, O. and Akbay, M.A. (2019) “A comprehensive review of deterministic models and applications for mean-variance portfolio optimization”, *Expert Systems with Applications*, 125, pp. 345-368. doi: https://doi.org/10.1016/j.eswa.2019.02.011

Kolm, P.N., Tütüncü, R. and Fabozzi, F.J. (2014) "60 Years of Portfolio Optimization: Practical Challenges and Current Trends", *European Journal of Operational Research*, 234(2), pp. 356-371. doi: https://doi.org/10.1016/j.ejor.2013.10.060

Kritzman, M. (2006) "Are optimizers error maximizers?", *The Journal of Portfolio Management*, 32(4), pp. 66-69. doi: https://doi.org/10.3905/jpm.2006.644197

Kritzman, M., Page, S. and Turkington, D. (2010) "In Defense of Optimization: The Fallacy of 1/N", *Financial Analysts Journal*, *66*(2), pp. 31-39. Available at: https://www.jstor.org/stable/27809177 (Accessed 15 August 2022)

Ledoit, O. and Wolf, M. (2003) "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection", *Journal of Empirical Finance*, 10(5), pp. 603-621. doi: https://doi.org/10.1016/S0927-5398(03)00007-0

Ledoit, O. and Wolf, M. (2004) "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis*, 88(2), pp. 365-411. doi: https://doi.org/10.1016/S0047-259X(03)00096-4

Lim, A.E., Shanthikumar, J.G. and Vahn, G.Y. (2011) "Conditional value-at-risk in portfolio optimization: Coherent but fragile", *Operations Research Letters*, *39*(3), pp. 163-171. doi: https://doi.org/10.1016/j.orl.2011.03.004

Mansini, R. and Speranza, M.G. (1999) "Heuristic algorithms for the portfolio selection problem with minimum transaction lots", *European Journal of Operational Research*, *114*(2), pp. 219-233. doi: https://doi.org/10.1016/S0377-2217(98)00252-5

Markowitz, H. M. (1952) "Portfolio Selection", *The Journal of Finance*, 7(1), pp. 77-91. Available at: https://www.jstor.org/stable/2975974 (Accessed 8 August 2022)

Markowitz, H. M. (2010) "Portfolio Theory: As I Still See It.", *Annual Review of Financial Economics*, 2, pp. 1-23. Available at: http://www.jstor.org/stable/42940208 (Accessed 7 August 2022)

Metaxiotis, K. and Liagkouras, K. (2012) “Multiobjective Evolutionary Algorithms for Portfolio Management: A comprehensive literature review”, *Expert Systems with Applications*, 39(14), pp. 11685-11698. doi: https://doi.org/10.1016/j.eswa.2012.04.053

Michaud, R.O. (1989), “The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal?”, *Financial Analysts Journal*, 45(1), pp. 31-42. doi: https://doi.org/10.2469/faj.v45.n1.31

Miffre, J. and Rallis, G. (2007) “Momentum strategies in commodity futures markets”, *Journal of Banking & Finance*, 31(6), pp. 1863-1886. doi: https://doi.org/10.1016/j.jbankfin.2006.12.005

Pai, G.V. and Michel, T. (2014) “Metaheuristic multi-objective optimization of constrained futures portfolios for effective risk management”, *Swarm and Evolutionary Computation*, 19, pp. 1-14. doi: https://doi.org/10.1016/j.swevo.2014.08.002

Pearson, K. (1905) “Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson. A Rejoinder”, *Biometrika*, 4(1/2), pp. 169–212. doi: https://doi.org/10.2307/2331536

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J. (2011) “Scikit-learn: Machine learning in Python”, *The Journal of Machine Learning Research*, 12, pp. 2825-2830. doi: https://doi.org/10.48550/arXiv.1201.0490

Pogue, G. A. (1970) “An Extension of the Markowitz Portfolio Selection Model to Include Variable Transactions’ Costs, Short Sales, Leverage Policies and Taxes”, *The Journal of Finance*, 25(5), pp. 1005-1027. Available at: https://www.jstor.org/stable/2325576 (Accessed 17 July 2022)

Stevens Analytics. (2022) Continuous Futures Dataset. *Nasdaq Data Link.* Available from: https://data.nasdaq.com/databases/SCF/documentation (Accessed 30 June 2022)

Verma, S., Pant, M. and Snasel, V. (2021) “A Comprehensive Review on NSGA-II for Multi-objective Combinatorial Optimization Problems”, *IEEE Access*, 9, pp. 57757-57791. doi: 10.1109/ACCESS.2021.3070634

Westfall, P.H. (2014) “Kurtosis as peakedness, 1905–2014. RIP”, *The American Statistician*, 68(3), pp. 191-195. Available at: https://www.jstor.org/stable/24591697 (Accessed 14 August 2022)

Zitzler, E. and Thiele, L. (1998) “Multiobjective optimization using evolutionary algorithms —a comparative case study”, In *International conference on parallel problem solving from nature*, pp. 292-301. doi: https://doi.org/10.1007/BFb0056872

Zitzler, E. and Thiele, L. (1999) “Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach”, *IEEE Transactions on Evolutionary Computation*, 3(4), pp. 257-271. doi: 10.1109/4235.797969