

Robust portfolio optimisation: covariance estimation, machine learning forecasting, and clustering-based allocation

Sarah Alsaed

A thesis submitted for the degree of Doctor of Philosophy

School of Mathematics, Statistics and Actuarial Science

University of Essex

3rd July 2026

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Spyridon Vrontos, whose guidance has accompanied me from my MSc dissertation through to the completion of this doctoral thesis. His intellectual rigour, patience, and unwavering engagement with my work throughout these years have been invaluable. I am deeply grateful for the continuity of his supervision, which has played a fundamental role in shaping this research.

On a personal level, I owe a great deal to my family. To my mother, Rula, thank you for your constant encouragement, understanding, and support throughout this journey. To my siblings, thank you for your love and for always being there.

I owe a particular debt of gratitude to my father, Hekmat, whose support made this doctoral journey possible. From the very beginning, he believed in my abilities without reservation and encouraged me to pursue my goals with confidence. His unconditional love and support have been the constant beneath everything in this thesis, and I am more grateful to him than I know how to put into words.

Abstract

Portfolio optimisation is challenging under conditions of covariance noise, unreliable return estimates, and allocation methods that do not account for the natural grouping structure of the asset universe. This thesis addresses these three sources of instability in turn. It refines the covariance matrix, forecasts expected returns through machine learning, and allocates capital through a clustering-based structure. These components are then combined into a single, robust portfolio-construction approach.

First, under risk-based objectives, several covariance estimators are evaluated, including shrinkage methods, dynamic conditional correlation models, and a quantile-based extension of the covariance denoising framework proposed in this thesis. Dynamic estimators combined with nonlinear shrinkage, and the denoising extension, offer superior performance. Second, the original Nested Clustered Optimisation framework is extended by incorporating machine-learning forecasts of expected returns together with these refined covariance estimators, and tested under alternative clustering algorithms and distance metrics. The strongest portfolios combine Random Forest forecasts with hierarchical clustering, a rank-based distance metric, and dynamic covariance estimators. These portfolios achieve the highest Sharpe ratios recorded across all strategies tested and outperform the benchmark. The proposed framework delivers a higher level of diversification and lower weight concentration than classical mean–variance optimisation using the same inputs. Third, the framework is extended to hedge fund strategy indices to examine whether its benefits generalise beyond equity portfolios and remain effective across distinct market regimes. Results show that the clustering framework delivers its greatest advantage during periods of market stress, when correlations across strategies converge.

The empirical results show that the proposed framework produces stronger out-of-sample performance across equities and hedge funds than classical mean–variance optimisation.

Contents

Abstract	ii
List of Tables	vii
List of Figures	xii
1 Introduction	1
1.1 Background and Motivation	1
1.2 Challenges in Mean–Variance Optimisation	3
1.3 Research Contributions	5
2 Noise Reduction and Covariance Matrix Estimation	8
2.1 Portfolio optimisation	9
2.2 Noise in the Covariance Matrix	14
2.2.1 Linear Shrinkage	16
2.2.2 Nonlinear Shrinkage	18
2.2.3 Denoising	19
2.2.4 Targeted Shrinkage	24
2.3 Hierarchical Risk Parity	24
2.4 DCC GARCH	28
2.4.1 DCC-NL	29
2.5 General Framework	30
2.6 Results and Analysis	31
2.6.1 Minimum Volatility	32

2.6.2	Maximum Decorrelation	34
2.6.3	HRP	36
2.6.4	Maximum Diversification	38
2.7	Conclusion	40
3	From Return Forecasts to Optimal Portfolios: Integrating Machine Learning with Nested Clustered Optimisation	43
3.1	Return Predictability	43
3.1.1	Linear Models	45
3.1.2	Ensemble and Tree-Based Methods	47
3.1.3	Support Vector Regression	50
3.2	Nested Clustered Optimisation	50
3.2.1	Clustering Methods within the NCO Framework	53
3.3	Data	57
3.3.1	Technical Indicators	58
3.3.2	Fundamental and Macroeconomic Variables	58
3.4	Methodology	59
3.4.1	Forecasting Framework	59
3.4.2	Portfolio Optimisation Framework	60
3.5	Results and Analysis	62
3.5.1	Forecasting Performance of Machine Learning Models	62
3.5.2	Statistical Validation Using the Probabilistic Sharpe Ratio	66
3.5.3	The Added Value of Clustering Over Mean–Variance Optimisation	69
3.5.4	Impact of Distance Metrics on Portfolio Performance	71
3.5.5	Portfolio Performance under the Maximum Quadratic Utility Objective	80
3.5.6	Portfolio Performance under the Maximum Sharpe Ratio	90
3.5.7	Portfolio Weight and Diversification Analysis	100
3.6	Conclusion	102

4	Machine Learning Forecasting and Nested Clustered Optimisation Across Market Regimes: Evidence from Hedge Fund Strategy Indices	104
4.1	Introduction	105
4.2	Hedge Fund Returns	108
4.2.1	Return Characteristics	108
4.2.2	Return Predictability	109
4.3	Data	111
4.3.1	Hedge Fund Indices	111
4.3.2	Predictive Factors	112
4.4	Methodology	113
4.4.1	Return Forecasting Framework	114
4.4.2	Portfolio Construction Framework	116
4.5	Forecasting Results	119
4.6	Portfolio Performance	125
4.7	Conclusion	131
5	Conclusion	133
	Bibliography	138
	Appendix	145

List of Tables

2.1	Annualised Performance Metrics for Minimum Volatility Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)	32
2.2	Annualised Performance Metrics for Minimum Volatility Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)	34
2.3	Annualised Performance Metrics for Maximum Decorrelation Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)	35
2.4	Annualised Performance Metrics for Maximum Decorrelation Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)	36
2.5	Annualised Performance Metrics for HRP Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)	37
2.6	Annualised Performance Metrics for HRP Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)	38
2.7	Annualised Performance Metrics for Maximum Diversification Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)	39
2.8	Annualised Performance Metrics for Maximum Diversification Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)	40
3.1	Portfolio performance by decile under five forecasting models.	64
3.2	Probabilistic Sharpe Ratio results under the Maximum Quadratic Utility objective.	68
3.3	Average out-of-sample performance under the Max Quadratic Utility objective across forecasting models and clustering strategies	70
3.4	Average out-of-sample performance under the Max Sharpe Ratio objective across forecasting models and clustering strategies	71

3.5	Rank-based comparison of distance metrics under the maximum quadratic utility objective. Lower ranks indicate superior and more consistent performance across covariance estimators.	77
3.6	Rank-based comparison of distance metrics under the maximum Sharpe ratio objective.	79
3.7	Out-of-sample portfolio performance under Random Forest forecasts (ONC Algorithm) - MQ Objective — Comparison of covariance estimators under different distance metrics.	81
3.8	Out-of-sample portfolio performance under Random Forest forecasts (Hierarchical Clustering) - MQ Objective — Comparison of covariance estimators under different distance metrics.	83
3.9	Out-of-sample portfolio performance under LASSO Regression forecasts (ONC Algorithm) - MQ Objective — Comparison of covariance estimators under different distance metrics.	84
3.10	Out-of-sample portfolio performance under LASSO Regression forecasts (Hierarchical Clustering) - MQ Objective — Comparison of covariance estimators under different distance metrics.	86
3.11	Out-of-sample portfolio performance under Gradient Boosting forecasts (ONC Algorithm) - MQ Objective — Comparison of covariance estimators under different distance metrics.	87
3.12	Out-of-sample portfolio performance under Gradient Boosting forecasts (Hierarchical Clustering) - MQ Objective — Comparison of covariance estimators under different distance metrics.	88
3.13	Out-of-sample portfolio performance under Random Forest (ONC Algorithm) – MS Objective — Comparison of covariance estimators under different distance metrics	91
3.14	Out-of-sample portfolio performance under Random Forest forecasts (Hierarchical Clustering) – MS Objective — Comparison of covariance estimators under different distance metrics.	93

3.15	Out-of-sample portfolio performance under LASSO (ONC Algorithm) – MS Objective — Comparison of covariance estimators under different distance metrics	94
3.16	Out-of-sample portfolio performance under LASSO forecasts (Hierarchical Clustering) – MS Objective — Comparison of covariance estimators under different distance metrics.	96
3.17	Out-of-sample portfolio performance under Gradient Boosting (ONC Algorithm) – MS Objective — Comparison of covariance estimators under different distance metrics	97
3.18	Out-of-sample portfolio performance under Gradient Boosting forecasts (Hierarchical Clustering) – MS Objective — Comparison of covariance estimators under different distance metrics.	99
4.1	Theil’s U statistics for Random Forest, Gradient Boosting, and Support Vector Regression under expanding and rolling window schemes. Values below unity indicate superior predictive accuracy relative to the AR(1) benchmark.	121
4.2	Subperiod Statistical Evaluation — Theil’s U Statistics by Model and Period (Expanding Window)	122
4.3	Subperiod Statistical Evaluation — Theil’s U Statistics by Model and Period (Rolling Window)	123
4.4	Full-Period Portfolio Performance (January 2019 – March 2026).	126
4.5	Subperiod portfolio performance across market regimes (expanding window). All strategies impose weight bounds $(0, 0.5)$ and transaction costs $k = 0.001$. Returns and standard deviations are annualised. Max DD denotes maximum drawdown. Bold entries denote the best value in each column within each subperiod.	129
5.1	Technical indicators employed as predictor variables in the forecasting models, following Neely et al. (2014).	145

5.2	Firm-level fundamental variables sourced from the WRDS Financial Ratio used in Chapter 3.	146
5.3	Macroeconomic variables from FRED-MD used as predictors in Chapter 3. . .	149
5.4	Out-of-sample portfolio performance under Random Forest forecasts — Mean–Variance optimisation without clustering, MQ Objective.	149
5.5	Out-of-sample portfolio performance under LASSO Regression forecasts — Mean–Variance optimisation without clustering, MQ Objective.	150
5.6	Out-of-sample portfolio performance under Gradient Boosting forecasts — Mean–Variance optimisation without clustering, MQ Objective.	150
5.7	Out-of-sample portfolio performance under Support Vector Regression forecasts — Mean–Variance optimisation without clustering, MQ Objective.	150
5.8	Out-of-sample portfolio performance under Elastic Net Regression forecasts — Mean–Variance optimisation without clustering, MQ Objective.	151
5.9	Out-of-sample portfolio performance under Random Forest forecasts — Mean–Variance optimisation without clustering, MS Objective.	151
5.10	Out-of-sample portfolio performance under LASSO Regression forecasts — Mean–Variance optimisation without clustering, MS Objective.	151
5.11	Out-of-sample portfolio performance under Gradient Boosting forecasts — Mean–Variance optimisation without clustering, MS Objective.	152
5.12	Out-of-sample portfolio performance under Support Vector Regression forecasts — Mean–Variance optimisation without clustering, MS Objective.	152
5.13	Out-of-sample portfolio performance under Elastic Net Regression forecasts — Mean–Variance optimisation without clustering, MS Objective.	152
5.14	Out-of-sample portfolio performance under Support Vector Regression forecasts (ONC Algorithm) - MQ Objective — Comparison of covariance estimators under different distance metrics.	153
5.15	Out-of-sample portfolio performance under Support Vector Regression forecasts (Hierarchical Clustering) - MQ Objective — Comparison of covariance estimators under different distance metrics.	154

5.16	Out-of-sample portfolio performance under Elastic Net forecasts (ONC Algorithm) - MQ Objective — Comparison of covariance estimators under different distance metrics.	155
5.17	Out-of-sample portfolio performance under Elastic Net forecasts (Hierarchical Clustering) - MQ Objective — Comparison of covariance estimators under different distance metrics.	156
5.18	Out-of-sample portfolio performance under Support Vector Regression forecasts (ONC Algorithm) - MS Objective — Comparison of covariance estimators under different distance metrics.	157
5.19	Out-of-sample portfolio performance under Support Vector Regression forecasts (Hierarchical Clustering) - MS Objective — Comparison of covariance estimators under different distance metrics.	158
5.20	Out-of-sample portfolio performance under Elastic Net forecasts (ONC Algorithm) - MS Objective — Comparison of covariance estimators under different distance metrics.	159
5.21	Out-of-sample portfolio performance under Elastic Net forecasts (Hierarchical Clustering) - MS Objective — Comparison of covariance estimators under different distance metrics.	160

List of Figures

2.1	Comparison of Eigenvalues for Different Correlation Matrices	15
2.2	Fitting the Marchenko-Pastur distribution to the S&P500 correlation matrix eigenvalues across a range of q values.	21
2.3	Eigenvalues before and after the denoising process	23
2.4	Dow Jones: Hierarchical Clustering Dendrogram	25
2.5	Comparison of Weight Distributions: Dow Jones Stocks Using HRP and Mini- mum Variance	27
3.1	Overview of the portfolio construction and evaluation framework across fore- casting models, clustering strategies, distance metrics, and covariance estimators.	69
3.2	Comparison of Sharpe Ratios under Random Forest forecasts – MQ Ob- jective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.	82
3.3	Comparison of Sharpe Ratios under LASSO forecasts – MQ Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.	85
3.4	Comparison of Sharpe Ratios under Gradient Boosting forecasts – MQ Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.	89
3.5	Comparison of Sharpe Ratios under Random Forest forecasts – MS Ob- jective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.	92

3.6	Comparison of Sharpe Ratios under LASSO forecasts – MS Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.	95
3.7	Comparison of Sharpe Ratios under Gradient Boosting forecasts – MS Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.	98
3.8	Portfolio allocation characteristics under the Maximum Quadratic Utility objective.	100
3.9	Portfolio allocation characteristics under the Maximum Sharpe Ratio objective.	101
4.1	Cumulative portfolio performance under the expanding window scheme (January 2019 – March 2026). Shaded regions denote the Pre-COVID (January 2019 – December 2019), COVID (January 2020 – December 2021), and Post-COVID (January 2022 – March 2026) subperiods.	127
4.2	Condition Number of the Rolling 60-Month Sample Covariance Matrix (January 2019 – March 2026)	130

Introduction

1.1 Background and Motivation

Building an effective investment portfolio is a fundamental decision for both individual and institutional investors. Regardless of the context, investors consistently face the decision of how to allocate their wealth across different asset classes to achieve financial stability and meet long-term financial goals. The growing variety of asset classes and the increased accessibility of international markets mean that each investment choice carries different characteristics and risks. To that end, investors need a model they can rely on as the basis for their asset allocation decisions. Harry Markowitz in 1952 [1] addressed the fundamental dilemma of investment decisions by providing the first mathematical framework for asset allocation. His model quantified the relationship between risk and return in investment portfolios and showed how assets could be combined to achieve a desired level of return while controlling risk, laying the foundation for what later became known as Modern Portfolio Theory (MPT).

A central principle of MPT is diversification: optimal allocations depend not only on the volatilities of individual assets but also on how those assets co-move within the portfolio. The model shows that a portfolio's overall volatility can be reduced by combining assets with imperfectly correlated returns. The concept of diversification led to the development of the efficient frontier, which represents, for any given asset universe, the set of portfolios that

achieve the highest possible return for each level of risk. Markowitz subsequently framed the allocation problem within a mathematical setting through what is now known as mean–variance optimisation. In this framework, optimal portfolios are obtained by solving a quadratic optimisation problem to determine the asset weights that maximise expected return for a given level of risk, or alternatively minimise risk for a desired level of return.

Ever since its introduction, MPT has become the benchmark for investors and has shaped much of the academic work that followed in portfolio management. However, the model relies on a set of restrictive assumptions that simplify the framework but are rarely satisfied in dynamic financial markets. A central requirement of mean–variance optimisation is the estimation of expected asset returns and the covariance matrix, both of which are obtained from historical data. The model implicitly assumes that these quantities can be estimated accurately, as it relies on efficient markets and on the stationarity and normality of asset-return time series. In practice, these assumptions rarely hold. Financial data contain a substantial amount of noise and rarely exhibit stable statistical properties, which leaves both the mean and covariance estimates highly sensitive to estimation error. Because the underlying signal is weak relative to these short-term disturbances, estimates drawn directly from historical data tend to shift considerably from one sample to another. When such unrefined inputs are used in portfolio optimisation, the allocations produced are neither stable nor reliable.

Over the years, with advances in statistics and the rapid progress in machine learning, numerous methods have been proposed to improve the estimation of expected returns and the covariance matrix. This thesis contributes to this line of work by building on and extending existing statistical and machine-learning approaches to produce more reliable inputs for portfolio optimisation. These more reliable inputs translate directly into risk-adjusted performance superior to the benchmarks, evidenced by consistently higher Sharpe ratios achieved across the empirical chapters. Beyond improving these inputs, the thesis channels them into an extended clustering-based allocation framework that produces more diversified portfolios than those built under classical mean–variance optimisation. Collectively, these contributions integrate robust covariance estimation, machine-learning forecasting, and clustering-based allocation into a single framework that produces portfolios that are better diversified and more resistant to noise than the

benchmarks.

1.2 Challenges in Mean–Variance Optimisation

To this day, the mean–variance optimisation framework remains the starting point for most research in portfolio management. Despite its strong theoretical foundation, the model shows instability in practice, and its limitations need to be properly addressed before it can be applied to noisy financial markets. To estimate the mean–variance efficient portfolios, the model relies on sample means and sample covariance matrices estimated directly from historical data. In theory, the model assumes that financial returns are stationary, normally distributed, and that correlations are constant over time, so using estimates from past data would lead to accurate allocations.

However, in empirical settings, return series often fail to satisfy the stationarity assumptions on which classical models rely. Moreover, the number of assets is usually large and comparable to the amount of data available for estimation, creating a clear dimensionality problem [2]. Under these conditions, the sample mean and covariance matrix estimates will be heavily influenced by noise, and even minor errors in the data can lead to substantial changes in the resulting portfolio weights.

The literature has addressed these limitations over the years and proposed solutions to the instability of estimates in the mean–variance input variables. A substantial body of research shows that errors in the estimated mean can have a disproportionately large impact on portfolio weights, often causing far greater instability than comparable errors in the covariance matrix. According to Chopra and Ziemba [3], the sample mean is very hard to estimate and requires an unrealistically large sample size to be estimated with a level of precision. But since financial data exhibit non-stationary behaviour, relying on a long historical sample size will make the estimate a moving target and subsequently an unreliable approximation to the true expected mean return. To address this issue, Black and Litterman [4] introduced their well-known Black–Litterman model. Their model allows investors to incorporate their own subjective views on expected returns with information extracted from past data, all within a Bayesian framework. On the other hand, to overcome the numerous challenges associated with estimating expected returns,

researchers have devoted their efforts to exploring risk-based portfolio formulations that do not rely on expected return estimates. In light of this, one prominent example is the risk parity approach, which focuses on balancing the risk contribution of assets in the investment portfolio [5]. Another alternative approach is to build portfolios based on diversification by maximising the diversification ratio, where the covariance matrix is the sole input required to solve the optimisation problem [6]. However, in recent years, the interest has shifted toward the use of machine learning models for estimating expected returns, mainly because return dynamics often display non-linear behaviour that conventional models cannot fully capture. Rather than imposing a fixed functional form, machine learning methods adapt and learn from the structure of the data and can capture interactions across a wide set of predictors without prior assumptions. This makes machine learning models well suited in modern portfolio optimisation applications.

Despite these efforts, the noise in the covariance matrix remains an issue for the optimiser. Most mean–variance optimisers require the inverse of the covariance matrix to arrive at the optimal allocations. Even minor errors in the sample covariance matrix will be magnified with inversion, leading to suboptimal allocations [7]. The issue becomes more apparent in high-dimensional settings, where the sample covariance matrix approaches singularity, and its eigenvalues are heavily dominated by noise. Ledoit and Wolf [8, 9, 10] made substantial contributions over the years towards reducing the noise in the sample covariance matrix. Their well-known Ledoit-Wolf linear shrinkage estimator improves the covariance matrix by combining the sample covariance matrix with a structured target, using a single shrinkage intensity applied uniformly across the matrix [8]. They later worked to improve and overcome the limitations of their linear shrinkage estimator by proposing a nonlinear shrinkage framework [10]. The nonlinear shrinkage estimator utilises different shrinkage intensities across the eigenvalues of the covariance matrix, allowing a more tailored adjustment for noise in high-dimensional settings. Other researchers turned to Random Matrix Theory (RMT) to stabilise the covariance matrix; their approach assumes that the sample covariance matrix resembles that of a matrix generated from pure noise [7]. Using RMT allows the researchers to identify which eigenvalues mostly reflect noise rather than structure, and which contain meaningful signal. Thus, this helps stabilise the noisy part of the spectrum before using the covariance matrix in optimisation.

Beyond noise-related instability, several studies note that conventional mean–variance optimisers treat assets as independent units and allocate directly across all of them. This assumption is often unrealistic, as financial data typically exhibits clear clustering patterns driven by sectoral, geographical, or shared factor exposures. When this structure is ignored, highly correlated asset groups can dominate the optimisation, leading to unstable, highly concentrated portfolios. A growing stream of research argues that it is more effective to first identify these clusters and then perform allocation at the cluster level rather than across all individual assets. López de Prado has made a significant contribution to this field through his work on Hierarchical Risk Parity and Nested Clustered Optimisation (NCO) [11, 12]. Both methods aim to simplify the allocation problem by finding the optimal weights at the cluster level first before assigning the final allocations across clusters, producing more consistent results that better reflect the way financial markets are naturally structured.

1.3 Research Contributions

This thesis makes the following contributions to the portfolio optimisation literature:

- **Comprehensive evaluation of modern covariance estimators.** The thesis tests a wide range of covariance estimation techniques, including linear and nonlinear shrinkage (Ledoit & Wolf), RMT, DCC-GARCH, and hybrid approaches, namely DCC combined with shrinkage or denoising, and the DCC-NL estimator. These methods are evaluated across different dimensionality settings and risk-based optimisation strategies.
- **Extension of López de Prado’s denoising framework.** The analysis expands the standard RMT denoising procedure by replacing noisy eigenvalues with quantile-based values: the 75th percentile, 90th percentile, and weighted average of quantiles. This offers an extended version of the methodology not present in the original framework, and in some cases outperforms the standard approach.
- **A machine-learning forecasting framework using a rich set of predictors.** The thesis develops an expected-return forecasting framework using five machine-learning models

(LASSO, Elastic Net, Support Vector Regression (SVR), Random Forest, and Gradient Boosting) and a comprehensive set of 214 predictors covering technical, fundamental, macroeconomic, and risk-related variables, all applied in a fully rolling, stock-level out-of-sample design.

- **Integration of ML forecasts with advanced covariance estimators inside the NCO algorithm.** A key contribution is the combination of ML-based return forecasts with refined covariance estimators within the NCO algorithm. The thesis evaluates this setup under two clustering structures (López de Prado's Optimal Number of Clusters (ONC) algorithm and hierarchical clustering) and compares it to a non-clustered benchmark. This design produces portfolios with higher Sharpe ratios and improved diversification, outperforming the non-clustered alternative.
- **Assessment of distance metrics in clustering-based optimisation.** The thesis examines how different distance metrics, including correlation-based, covariance-based, rank-based, and mutual information-based distances, affect clustering results and portfolio performance within the NCO algorithm.
- **Extension to alternative asset classes and regime-dependent analysis.** The ML forecasting and NCO algorithm is extended to a fund-of-hedge-funds setting using Bloomberg hedge fund strategy indices and macroeconomic and financial predictors drawn from the Fung–Hsieh, Fama–French, and FRED library. The analysis evaluates whether the forecasting framework and covariance estimators generalise beyond equities and examines the performance of NCO-based portfolio construction strategies across distinct market regimes.

The remainder of this thesis is organised as follows. Chapter 2 reviews the theoretical background of portfolio optimisation, focusing on the limitations of the classical mean–variance framework and the motivation for improved estimation techniques. It discusses the sources of noise-induced instability in financial data and introduces the covariance estimators used throughout the thesis. The chapter also presents a comprehensive comparison of these estimators across different datasets, sampling frequencies, and risk-based portfolio objectives.

Chapter 3 sets out the methodological framework, combining the forecasting models with the refined covariance estimators within the NCO algorithm and examining the role of different clustering regimes and distance metrics in shaping portfolio allocations.

Chapter 4 extends the analysis by applying the proposed methods to a fund-of-hedge-funds setting, evaluating whether the ML forecasting framework and NCO-based portfolio construction strategies generalise beyond equities and whether the same techniques can improve risk management across distinct market regimes.

Finally, Chapter 5 concludes the thesis and outlines potential directions for future research.

Noise Reduction and Covariance Matrix Estimation

Traditional mean–variance optimisers that rely on raw estimates of the covariance matrix tend to be unstable because the sample covariance matrix contains a considerable amount of noise. This chapter focuses on refining and improving the covariance matrix used in portfolio optimisation by examining shrinkage techniques, Random Matrix Theory (RMT) filtering, and several hybrid approaches. It also considers the parametric DCC-GARCH model to estimate dynamic conditional correlations and proposes combining DCC-GARCH estimates with shrinkage and denoising techniques in a hybrid framework. The chapter provides a comprehensive comparison of parametric and non-parametric estimators using a rolling-window approach.

In spite of the researchers’ efforts to stabilise Markowitz optimisation using traditional approaches, the presence of noise in the sample covariance matrix remains a significant barrier to achieving optimal allocations. The sample covariance matrix of assets is estimated from historical returns, which contain considerable noise. As a result, the covariance matrix often cannot be estimated reliably. Moreover, most portfolio optimisers require the inversion of the covariance matrix; with noisy estimates, the solutions are almost guaranteed to be unstable. In other words, estimation errors are amplified when inverting the covariance matrix, leading to continuous rebalancing of portfolio weights. Ledoit and Wolf [8, 2, 10] devoted years of research to developing noise-reduction techniques aimed at improving the estimation of the covariance matrix. Their early work shrunk the sample covariance toward structured targets such as the identity matrix or a constant-correlation matrix. Building on this, they later introduced more

advanced and robust approaches, including nonlinear shrinkage methods that allow eigenvalue-specific adjustments and have proven effective in high-dimensional settings. RMT offers an alternative non-parametric approach to handling noise in the sample covariance matrix [7]. It separates noise-dominated eigenvalues from those that contain meaningful structure and applies filtering procedures to denoise the matrix accordingly. Alongside these approaches, parametric models such as the Dynamic Conditional Correlation (DCC) framework have shown considerable potential. The DCC model is widely used for its ability to capture the time-varying behaviour of asset correlations, which tend to fluctuate substantially with changing market conditions [13]. Incorporating such dynamic correlation estimates into portfolio construction can lead to more responsive and better-hedged portfolios during periods of shifting market regimes. Recent work has also explored the use of machine-learning techniques within portfolio optimisation. A central idea is to cluster assets before allocating weights, with the aim of improving diversification and reducing the dimensionality of the optimisation problem. López de Prado contributed significantly to this line of research by developing the Hierarchical Risk Parity (HRP) algorithm, which clusters assets based on their correlations or a suitable distance measure [11]. This clustering step ensures that weights are distributed across groups of similar assets rather than treating the entire universe as one homogeneous set.

2.1 Portfolio optimisation

Mean–variance optimisation is one of Harry Markowitz’s seminal contributions to the investment field [1]. It is a mathematical framework for constructing portfolios that achieve a delicate balance between risk and return. At its core, the mean–variance framework aims to allocate weights among assets, adhering to the principle of diversification while accounting for investors’ risk preferences. Optimal portfolios achieve the lowest level of risk for a given level of return or, equivalently, the highest level of return for a given level of risk. A key strategy to construct optimal portfolios is combining low-correlated assets. Hence, significant losses in one of the assets will likely be counterbalanced by returns from other investments in the same portfolio. The mean–variance can be considered a convex optimisation problem formulated using various objectives and different sets of constraints. Thus, the mean–variance offers portfolio solutions

based on a range of objective functions and constraint sets, several of which are examined in this chapter.

Minimum Variance Portfolio

The minimum volatility strategy is a portfolio construction method with the objective function of minimising the portfolio's volatility [1]. The minimum volatility allocation strategy results in a portfolio with the lowest possible risk level regardless of the portfolio's expected return, where the following optimisation problem is being solved:

$$\begin{aligned} & \text{minimise} \quad \mathbf{w}^T \Sigma \mathbf{w} \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1, \\ & \quad w_i \geq 0 \quad \text{for } i = 1, \dots, N, \end{aligned} \tag{2.1}$$

where $\Sigma \in R^{N \times N}$ is the covariance matrix of assets, and $\mathbf{w} \in R^{N \times 1}$ is the vector of weights allocated to each asset. The diversification benefits are achieved by allocating weights to assets with low volatilities and low correlations. The emphasis on risk reduction makes the strategy more desirable for risk-averse investors.

Maximum Sharpe Ratio Portfolio

The maximum Sharpe ratio optimisation problem aims to allocate weights to maximise the risk-adjusted return, known as the Sharpe ratio. The Sharpe ratio is a risk-adjusted performance measure introduced by [14]. It is used to estimate the portfolio's excess return per unit of its volatility or risk. By definition, the objective of the maximum Sharpe ratio optimisation problem is to find the optimal weights that maximise the portfolio Sharpe ratio and is formulated as

$$\begin{aligned} & \text{maximise} \quad \frac{\mathbf{w}^T \boldsymbol{\mu} - R_f}{(\mathbf{w}^T \Sigma \mathbf{w})^{\frac{1}{2}}} \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1, \\ & \quad w_i \geq 0 \quad \text{for } i = 1, \dots, N, \end{aligned} \tag{2.2}$$

where $\mathbf{w} \in R^{N \times 1}$ is the vector of portfolio weights, $\boldsymbol{\mu} \in R^{N \times 1}$ is the vector of expected asset returns, $R_f \in R$ is the risk-free rate, and $\Sigma \in R^{N \times N}$ is the covariance matrix of assets.

The strategy aims to strike a balance between risk and return by allocating weights based on assets' excess return per unit of risk. As the strategy considers both risk and return, it allows investors to construct portfolios tailored to their investment goals and risk tolerance by spreading investments across assets with low risk and high potential returns.

Black-Litterman Model

The Black-Litterman asset allocation model is a popular allocation model that builds upon modern portfolio theory [4]. The model addresses some of the shortcomings of the mean-variance optimisation by employing a Bayesian approach for portfolio allocation. The model's novelty is that it allows investors to incorporate their views about future asset returns, which might not be captured by pure reliance on historical data. The model integrates investors' subjective beliefs and objective market anticipations to estimate expected returns. In other words, the Black-Litterman model considers the expected return a parameter that needs to be estimated as

$$E(\mathbf{R}) = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1} [(\tau\Sigma)^{-1}\boldsymbol{\Pi} + P^T\Omega^{-1}\mathbf{Q}], \quad (2.3)$$

where:

- $E(\mathbf{R}) \in R^{N \times 1}$ is the vector of expected returns
- $\Sigma \in R^{N \times N}$ is the covariance matrix of asset returns
- $\mathbf{Q} \in R^{K \times 1}$ is the vector of views on expected return
- $P \in R^{K \times N}$ is a mapping matrix that maps views to the corresponding assets
- $\Omega \in R^{K \times K}$ is the views uncertainty matrix
- $\tau \in R$ is a scalar
- $\boldsymbol{\Pi} \in R^{N \times 1}$ is the vector of implied excess returns

where N is the number of assets and K is the number of investor views.

The second part of the formula $(\tau\Sigma)^{-1}\mathbf{\Pi} + P^T\Omega^{-1}\mathbf{Q}$ can be interpreted as a weighted average of the market's expectations and the investor's views regarding expected future returns. The allocation of weights is determined by the investor's confidence level about the subjective views. The more confident the investor is, the higher the weight allocated to the views vector $\mathbf{Q}$, and the less weight is allocated to the market-implied excess return $\mathbf{\Pi}$. The formula results in a posterior estimate of expected returns that can be utilised in the mean–variance optimisation problems.

Despite the model's popularity, it is important to point out that the model allows for the integration of investors' subjective beliefs, which are inherently subjective. Hence, the model carries the risk of introducing biases and incorrect assumptions into the portfolio optimisation process. Accordingly, investors need to frequently rebalance their portfolios to ensure their views align with the current market conditions.

Maximum Diversification

The maximum diversification investment strategy aims to achieve the highest level of diversification by spreading risk across different holdings [6]. This strategy is considered a risk-based allocation approach that relies solely on risk measures without directly incorporating expected returns as optimisation inputs to determine optimal portfolio weights. At its core, it aims to maximise the diversification ratio as follows:

$$\begin{aligned} & \text{maximise} \quad \frac{\mathbf{w}^T \boldsymbol{\sigma}}{\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}} \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1, \\ & \quad w_i \geq 0 \quad \text{for } i = 1, \dots, N, \end{aligned} \tag{2.4}$$

where $\boldsymbol{\sigma} \in R^{N \times 1}$ is the vector of individual asset volatilities. The term in the diversification ratio numerator, $\mathbf{w}^T \boldsymbol{\sigma}$, represents the weighted average asset volatilities, while the term in the denominator, $\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}$, represents the portfolio standard deviation. Maximising the diversification ratio ensures that risk is evenly spread across assets in the portfolio. Maximising the weighted average risk while controlling the portfolio risk leads to a more diversified portfolio comprising assets

with different risk profiles.

Maximum Decorrelation

The maximum decorrelation portfolio allocation strategy seeks to minimise the correlation between assets, aiming to achieve the highest potential diversity while reducing risk. The strategy combines assets with low historical correlations to create portfolios that are more resilient to systematic risk and dynamic market fluctuations. The optimisation problem can be mathematically expressed as follows:

$$\begin{aligned} & \text{minimise} \quad \mathbf{w}^T A \mathbf{w} \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1, \\ & \quad \quad \quad w_i \geq 0 \quad \text{for } i = 1, \dots, N, \end{aligned} \tag{2.5}$$

where $A \in \mathbb{R}^{N \times N}$ is the correlation matrix of assets. The maximum decorrelation portfolio applies in situations where investors believe that assets have similar return and volatility levels, but their correlations are heterogeneous.

Equal Risk Contribution

The equal risk contribution investment strategy allocates weights among assets to ensure that each asset contributes equally to the portfolio's overall risk [5]. The method allocates capital among portfolio constituents to balance risk evenly, achieving optimal diversification. The total risk of the portfolio can be expressed as

$$\sigma_p(\mathbf{w}) = \sum_{i=1}^N RC_i(\mathbf{w}) = \sum_{i=1}^N w_i \cdot \frac{\partial \sigma(\mathbf{w})}{\partial w_i}, \tag{2.6}$$

where $RC_i(\mathbf{w})$ is the risk contribution of asset i , and $\frac{\partial \sigma(\mathbf{w})}{\partial w_i}$ is the marginal risk contribution of asset i . The marginal risk contribution of an asset quantifies the change in portfolio overall risk associated with an increase in the weight allocated to that particular asset. The risk contribution of each asset is computed as the product of the asset's weight and its marginal risk contribution as

$$RC_i(\mathbf{w}) = w_i \frac{\partial \sigma(\mathbf{w})}{\partial w_i}. \quad (2.7)$$

Finally, the equal risk contribution ensures that the risk contribution of each asset is the same $RC_i(\mathbf{w}) = RC_j(\mathbf{w})$.

To determine the optimal weights that achieve equal risk contribution among portfolio assets, the following optimisation problem is being solved:

$$\begin{aligned} & \text{minimise} && \sqrt{\mathbf{w}^T \Sigma \mathbf{w}} \\ & \text{subject to} && \sum_{i=1}^N \ln w_i \geq c, \\ & && \sum_{i=1}^N w_i = 1, \\ & && w_i \geq 0 \quad \text{for } i = 1, \dots, N, \end{aligned} \quad (2.8)$$

where c is a constant measuring the minimum level of diversification needed to ensure that the equal risk contribution condition holds [5]. However, the choice of c is implementation-specific and is not explored further in this thesis.

The performance of all these strategies depends critically on the quality of the covariance matrix estimate, which is examined in the following section.

2.2 Noise in the Covariance Matrix

In portfolio optimisation, the covariance matrix of assets plays a crucial role in capturing the relationship between assets. It is used to identify the portfolio's possible diversification and potential risk factors. However, financial data often exhibit a low signal-to-noise ratio, making extracting meaningful patterns and accurate forecasts challenging. The noise in the input variables, particularly the noise present in asset returns, makes the estimate of the covariance matrix unstable and numerically ill-conditioned. As a result, slight deviations in expected return estimates will lead the Critical Line Algorithm (CLA) to generate different portfolio allocations.

Additionally, financial data usually encounter a curse of dimensionality, mainly when the number of assets under consideration is high and comparable to the number of available obser-

vations. In other words, the ratio $q = \frac{N}{T}$, where N is the number of assets and T is the sample size, regulates the reliability of the covariance matrix estimate. A covariance matrix of N assets consists of $\frac{N(N-1)}{2}$ entries, derived from N time series, each of length T . As $q \rightarrow 1$, traditional statistical methods will fail to accurately estimate the empirical covariance matrix. Hence, the resulting estimate will resemble a random structure due to the considerable amount of noise [7].

Noise-induced instability in covariance matrices is often regulated by the ratio $\frac{N}{T}$ and is associated with the curse of high dimensionality. However, another primary source of instability in empirical covariance matrices is induced by financial data structure and is independent of the ratio $\frac{N}{T}$. Put differently, instability may arise from the signal within financial data [12]. The numerical stability of covariance or correlation matrices is commonly assessed by the condition number, measured by the absolute ratio between the maximum and minimum eigenvalues. A low condition number implies a well-behaved correlation matrix with minimal numerical instability. As the condition number increases, the correlation matrix will be less stable and more sensitive to slight variations in input variables.

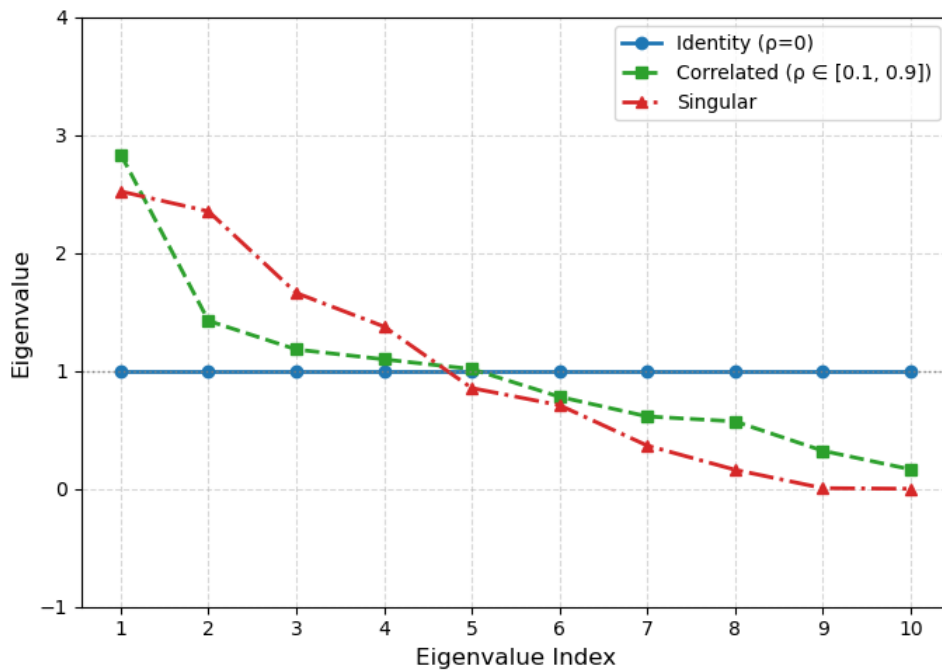


Figure 2.1: Comparison of Eigenvalues for Different Correlation Matrices

Figure 2.1 displays the eigenvalues of a diagonal correlation matrix with $\rho = 0$, a correlation matrix of highly correlated investments with $\rho \in [0.1, 0.9]$, and a singular correlation matrix. The identity matrix has a condition number of 1, resembling a stable correlation matrix. As correlated

assets are included, the difference between the maximum and minimum eigenvalues widens quickly, causing the condition number to rise. The inherent instability in the covariance matrix arising as ρ deviates from the uncorrelated case has significant implications on Markowitz's solution. The inverse of the covariance matrix plays a crucial role in finding the optimal weights under Markowitz optimisation problem. However, unless $\rho \approx 0$, the inverse of the covariance matrix is guaranteed to be unstable, resulting in unreliable solutions [12]. This source of instability exists regardless of the ratio $\frac{N}{T}$; it depends on the structure of the data and is referred to as signal-induced instability.

In summary, employing empirical covariance matrices in portfolio optimisation problems poses many challenges. Hence, using the raw sample covariance matrix is not advisable. Instead, investors should address the covariance matrix's instability to ensure proper robust allocations. Subsequent sections explore prevalent statistical and machine learning approaches used to tackle noise in covariance matrices.

2.2.1 Linear Shrinkage

Ledoit and Wolf [8, 2] dedicated years of research to addressing the noise-induced instability in the covariance matrix. They introduced the famous Ledoit and Wolf shrinkage estimator and continued to build upon it over the years. The starting point was the seminal work of Stein [15] on estimating the mean vector; Ledoit and Wolf extended Stein's model to the covariance matrix estimation, retaining the Mean Squared Error (MSE) as a risk function. The goal is to minimise the MSE between the true and the estimated covariance matrices. To that end, Ledoit and Wolf formulated their shrinkage estimator as follows.

The Frobenius norm of an $N \times N$ matrix A is defined as

$$\|A\|_F = \sqrt{\frac{\text{Tr}(AA^T)}{N}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N a_{ij}^2}, \quad (2.9)$$

where a_{ij} denotes the (i, j) -th entry of the matrix A , and $\text{Tr}(A)$ is the trace of matrix A .

The (scaled) Frobenius loss function is defined as

$$L_F(\hat{\Sigma}_T, \Sigma_T) := \|\hat{\Sigma}_T - \Sigma_T\|_F^2, \quad (2.10)$$

where $\hat{\Sigma}_T$ is an estimator of the true covariance matrix Σ_T , and $\|\cdot\|_F$ denotes the Frobenius norm.

Finally, the (scaled) MSE is the expected value of the Frobenius loss, and is given by

$$E[L_F(\hat{\Sigma}_T, \Sigma_T)] = E\left[\|\hat{\Sigma}_T - \Sigma_T\|_F^2\right]. \quad (2.11)$$

The core concept of Ledoit and Wolf's work is to strike a balance between the sample covariance matrix and a structured estimator known as the shrinkage target. The shrinkage model aims to minimise the expected Frobenius loss of the difference between the shrinkage estimator and the true covariance matrix. In their classical model, they used the identity matrix $I_T \in R^{N \times N}$ as a shrinkage target and expressed the optimisation problem as follows:

$$\min_{q_1, q_2} E\left[\|\hat{\Sigma}_T - \Sigma_T\|_F^2\right], \quad (2.12)$$

subject to:

$$\hat{\Sigma}_T = q_1 I_T + q_2 S_T, \quad (2.13)$$

where $q_1, q_2 \in R$ are the coefficients to be optimised, and $S_T \in R^{N \times N}$ is the sample covariance matrix. Setting $q_1 = \gamma_T^* \mu_T$ and $q_2 = 1 - \gamma_T^*$, the solution to the optimisation problem can be viewed as a convex linear combination given by

$$\Sigma_T^* = \gamma_T^* \mu_T I_T + (1 - \gamma_T^*) S_T, \quad (2.14)$$

where

$$0 \leq \gamma_T^* \leq 1.$$

which implies $q_2 \in [0, 1]$ and $q_1 \in [0, \mu_T]$. The shrinkage intensity parameter γ_T^* controls the level of structure added to the covariance matrix. A higher shrinkage intensity imposes more structure, shrinking the estimator more towards a multiple of the identity matrix $\mu_T I_T$, where $\mu_T = \frac{1}{N} \text{Tr}(S_T)$ is the average sample variance.

Ledoit and Wolf extended their work in subsequent papers by introducing various shrinkage

targets tailored to the characteristics of financial data. The choice of an optimal shrinkage target aims to balance simplicity with accuracy. In other words, the shrinkage target should involve as few parameters as possible while closely representing the structure of the true covariance matrix. The selection of a shrinkage target is primarily based on prior knowledge of the data's structure. Ledoit and Wolf in [8] proposed the constant correlation model as a shrinkage target. The model implies that all pairwise correlations are equal and constant over time; the mean value of all correlations serves as the shared invariant correlation. Alternatively, in [9][2][8], the single-index matrix of Sharpe, the identity matrix, and the constant variance matrix were used as shrinkage targets to preserve further robustness.

Ledoit and Wolf's shrinkage estimator has been widely tested using empirical market data. Tuan's [16] study on real estate portfolios showcases the advantages of using the shrinkage estimator in enhancing portfolio performance compared to classical methods. Chen et al. [17] sought to enhance Ledoit and Wolf's shrinkage estimator. They introduced the Oracle approximating shrinkage (OAS); in their study, results demonstrated that the OAS estimator dominates Ledoit and Wolf's estimator under Gaussian sample distribution. Although shrinkage techniques have been widely utilised in the literature to tackle the presence of noise in the covariance matrix, authors have criticised this approach for its potential over-simplification. Shrinkage mitigates the noise in the covariance matrix without discriminating between noise and signal in the data. Thus, shrinkage might lead to a diluted signal alongside the noise [12]. For this reason, linear shrinkage techniques might not be the most favourable noise reduction approach, especially when dealing with financial data with an inherently weak signal-to-noise ratio.

2.2.2 Nonlinear Shrinkage

Despite the popularity and successful application of linear shrinkage in estimating robust financial covariance matrices, it has some limitations that can be improved upon. Linear shrinkage applies a uniform shrinkage intensity γ_T^* to all elements of the covariance matrix despite their sampling error. A more intuitive approach would be to use different shrinkage intensities for different covariance matrix elements based on their respective sampling error. However, estimating

N^2 different shrinkage intensities would be computationally challenging, so Ledoit and Wolf proposed a nonlinear approach that shrinks the eigenvalues of the covariance matrix (of order N) rather than adjusting individual elements. This method enhances computational efficiency and ensures that the resulting shrunk covariance matrix is positive semidefinite.

The theoretical foundation for nonlinear shrinkage proposed by Ledoit and Wolf [10] is to reduce overfitting by adjusting the sample eigenvalues to better approximate the population eigenvalues. To recover population eigenvalues, Ledoit and Wolf's nonlinear estimator inverts the Marčenko-Pastur equation [18], which defines the asymptotic distribution of the eigenvalues of the sample covariance matrix as the dimensions of the data matrix grow.

However, inverting the Marčenko-Pastur equation directly can be challenging, mainly when dealing with high-dimensional data. To tackle this, Ledoit and Wolf [19] introduced the "Quantized Eigenvalues Sampling Transform," known as the QuEST function. The QuEST function is a numeric deterministic function that facilitates the inversion of the Marčenko-Pastur equation; by discretising the equation, the function allows a practical estimation of population eigenvalues even in high dimensional settings.

Building upon the foundation of nonlinear shrinkage, Ledoit and Wolf [20] introduced advanced techniques under different loss functions to further improve the estimation of the covariance matrix. For instance, the Quadratic-Inverse Shrinkage (QIS) estimator, derived under the Frobenius loss and its two cousins, Inverse Stein's loss and Minimum Variance loss, provides robustness and estimation accuracy in high-dimensional settings. Additionally, the Geometric Inverse Shrinkage (GIS) estimator, derived under the Symmetrized Kullback-Leibler loss, combines linear and quadratic shrinkage elements to offer greater flexibility. The Linear-Inverse Shrinkage (LIS) estimator, developed under Stein's loss, remains one of the most computationally efficient approaches. For further details, see [20].

2.2.3 Denoising

The reliance on the use of empirical correlation matrices in portfolio allocation raises many doubts for several reasons. Firstly, the nondeterministic nature and the randomness of historical asset returns introduce noise in the estimated correlation matrix values.

Moreover, correlation matrices are guaranteed to be noisy with high condition numbers when the number of observations used (T) is relatively small compared to the time series dimensions (N). Increasing the number of observations is not a practical solution to addressing this matter due to the trade-off. The time series length should strike a balance between having enough data to ensure the reliability of the correlation estimator while simultaneously capturing current market dynamics.

According to Laloux et al. [7], about 94% of the spectrum of a correlation matrix derived from the returns of S&P500 closely resembles the spectrum of a correlation matrix observed under a purely random scenario. Laloux et al.'s insights highlighted the role of RMT in reducing the impact of noise in empirical correlation matrices through a process known as noise filtering or denoising [12].

Convergence to Marchenko–Pastur Distribution

The Marchenko-Pastur theory stands as one of the elemental outcomes of RMT. The theory defines a theoretical distribution of eigenvalues derived from random matrices [18].

Assume a matrix X with dimensions $T \times N$, where T is the number of observations and N is the number of assets, with observations derived from a process with a mean of zero and a variance of σ^2 . In the limit ($N \rightarrow +\infty$ and $T \rightarrow +\infty$) with ($1 < \frac{T}{N} < \infty$), the eigenvalues (λ) of matrix $C = T^{-1}X'X$ converge to the Marchenko-Pastur probability density function:

$$f(\lambda) = \begin{cases} \frac{T}{N} \sqrt{\frac{(\lambda_+ - \lambda)(\lambda - \lambda_-)}{2\pi\sigma^2}}, & \text{if } \lambda \in (\lambda_+, \lambda_-) \\ 0, & \text{if } \lambda \notin (\lambda_+, \lambda_-) \end{cases}, \quad (2.15)$$

where:

λ_+ : the maximum expected eigenvalue, $\lambda_+ = \sigma^2(1 + \sqrt{\frac{N}{T}})^2$.

λ_- : the minimum expected eigenvalue, $\lambda_- = \sigma^2(1 - \sqrt{\frac{N}{T}})^2$.

Fitting the Marchenko–Pastur distribution to an empirical correlation matrix helps separate noise-related components from signal-related components. Eigenvalues linked to noise will follow the Marchenko–Pastur distribution. However, distinct eigenvalues that are too sound and clear to be associated with noise will lie outside the confines of the Marchenko–Pastur

distribution. In other words, eigenvalues $\lambda \in [\lambda_-, \lambda_+]$ are associated with noise, whereas eigenvalues $\lambda \notin [\lambda_-, \lambda_+]$ indicate a meaningful signal in the data.

Figure 2.2 illustrates the fitting of the Marchenko–Pastur distribution to the empirical distribution of eigenvalues observed in S&P500 correlation matrices. The fitting is performed across different values of $q = \frac{N}{T}$ ratios ($q = 1.2$, $q = 0.4$, $q = 0.2$, and $q = 0.1$, respectively).

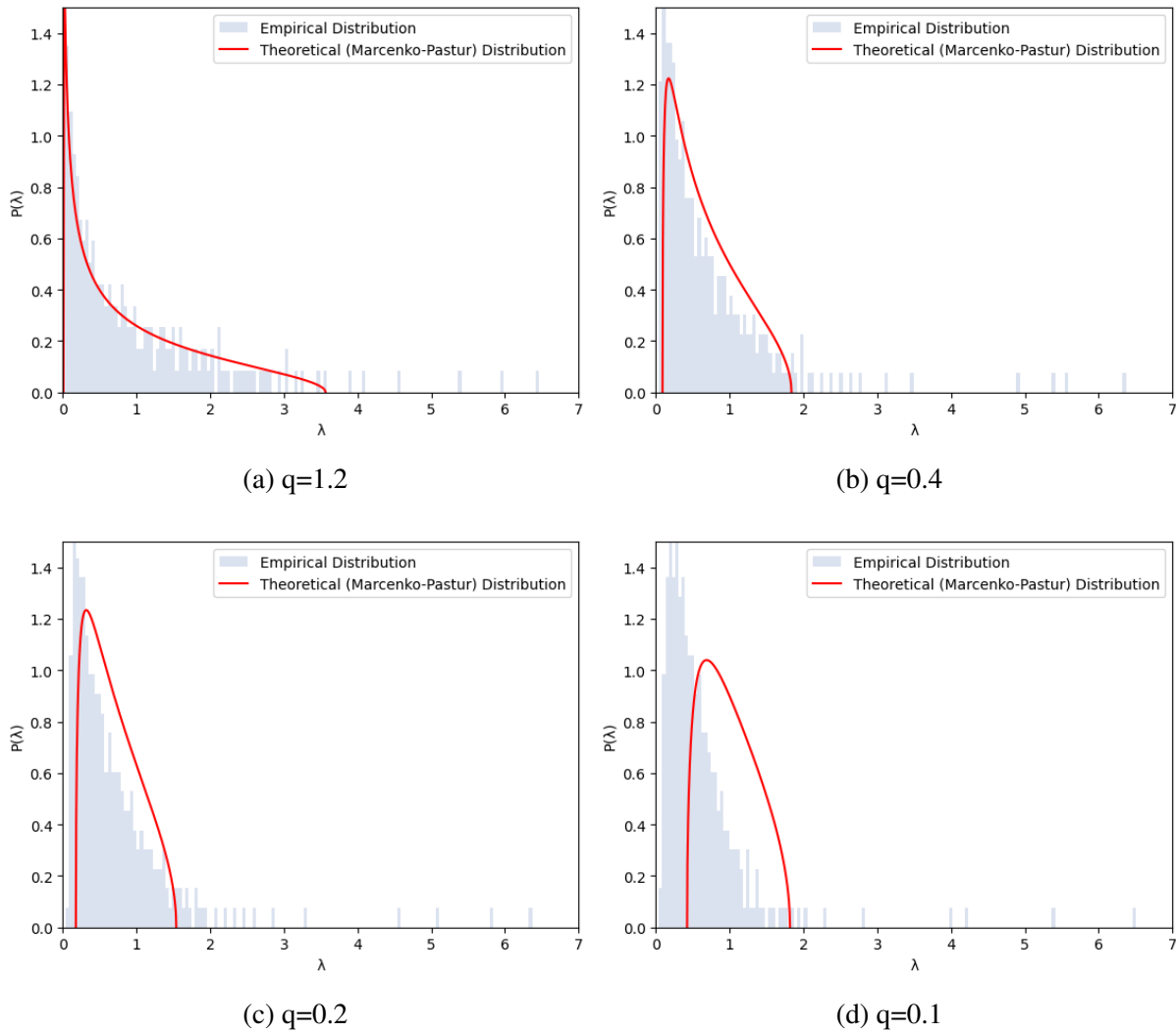


Figure 2.2: Fitting the Marchenko-Pastur distribution to the S&P500 correlation matrix eigenvalues across a range of q values.

In Figure 2.2, the Marchenko-Pastur distribution effectively fits the empirical distribution of eigenvalues in the case where $q=1.2$. This result is because when the number of observations T is insufficient to estimate a robust correlation matrix, the correlation matrix is guaranteed to contain significant noise and resemble a random structure. Put differently, when $\frac{N}{T} \rightarrow 1$, λ_- will converge to 0, resulting in a determinant of C that approaches 0. Unless the noise is addressed,

the correlation matrix C should not be used as an input for portfolio optimisation problems or further analysis.

Constant Residual Eigenvalue Method

The discrimination between eigenvalues linked to noise and those linked to signal paves the way to the denoising process. Marcos López de Prado [12] proposed the constant residual eigenvalues method to denoise the correlation matrix. In his methodology, eigenvalues associated with noise, particularly $\lambda \in [\lambda_-, \lambda_+]$, are replaced by their average value while preserving $\lambda \notin [\lambda_-, \lambda_+]$. Consequently, the new set of eigenvalues is used to reconstruct the denoised correlation matrix C_1 through eigendecomposition, thus maintaining the trace of the matrix as follows:

$$\tilde{C}_1 = W\tilde{\Lambda}W^T, \quad (2.16)$$

$$C_1 = \tilde{C}_1 \left[\left(\text{diag}[\tilde{C}_1] \right)^{1/2} \left(\text{diag}[\tilde{C}_1] \right)^{1/2^T} \right]^{-1}, \quad (2.17)$$

where W is the matrix of eigenvectors, $\tilde{\Lambda}$ is a diagonal matrix with the corrected eigenvalues positioned on the diagonal. The second transformation is to ensure a proper structure of C_1 with ones on the diagonal.

Figure 2.3 compares the eigenvalues of the S&P500 correlation matrix before and after the denoising process using constant residual eigenvalue method.

As depicted in Figure 2.3, noisy eigenvalues are replaced by their average value. This will ensure that a higher value replaces small eigenvalues approaching 0. As a result, the determinant of the matrix will increase, ensuring more numerical stability. Furthermore, the condition number $\kappa(C) = \lambda_{\max}/\lambda_{\min}$ measures the numerical stability of the matrix. A higher value of $\lambda_{\min}$ achieved by denoising will ensure a lower condition number, indicating a well-behaved correlation matrix compared to the empirical estimator.

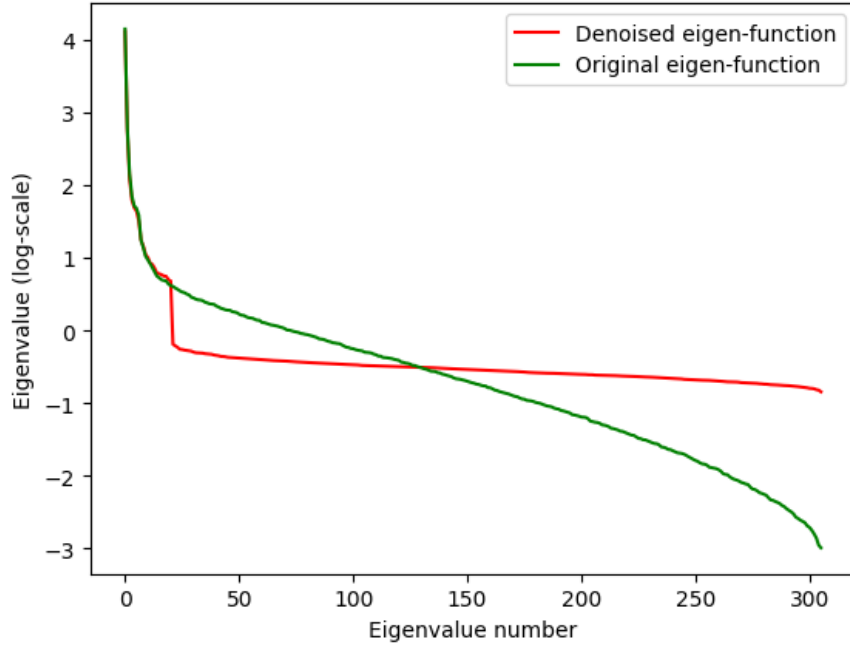


Figure 2.3: Eigenvalues before and after the denoising process

Quantile-Based Eigenvalue Correction

In this thesis, the constant residual eigenvalue method is extended by correcting noisy eigenvalues using quantile-based values computed from the noise-band eigenvalues themselves, rather than their simple average. As established in Equation (2.16), the diagonal matrix $\tilde{\Lambda}$ contains the corrected eigenvalues, where eigenvalues associated with noise, $\lambda \in [\lambda_-, \lambda_+]$, are corrected while eigenvalues associated with signal, $\lambda \notin [\lambda_-, \lambda_+]$, are preserved unchanged. Let

$$\Lambda_{\text{noise}} = \{\lambda : \lambda \in [\lambda_-, \lambda_+]\} \quad (2.18)$$

denote the set of noisy eigenvalues. Three quantile-based correction rules are proposed for populating the corresponding diagonal entries of $\tilde{\Lambda}$. The first substitutes each noisy eigenvalue with the 75th percentile of Λ_{noise} :

$$\lambda^* = Q_{75}(\Lambda_{\text{noise}}), \quad (2.19)$$

while the second uses the 90th percentile:

$$\lambda^* = Q_{90}(\Lambda_{\text{noise}}). \quad (2.20)$$

The third combines multiple quantile levels into a weighted average:

$$\lambda^* = 0.25 Q_{25}(\Lambda_{\text{noise}}) + 0.50 Q_{50}(\Lambda_{\text{noise}}) + 0.25 Q_{75}(\Lambda_{\text{noise}}), \quad (2.21)$$

where $Q_p(\cdot)$ denotes the p -th percentile of the noisy eigenvalue set Λ_{noise} . In each case, every noisy eigenvalue in Λ_{noise} is corrected to the same value λ^* , analogous to how the original method sets every noisy eigenvalue to their shared average, and the denoised correlation matrix is reconstructed using the same eigendecomposition procedure as in Equations (2.16) and (2.17).

2.2.4 Targeted Shrinkage

Targeted shrinkage is another practical strategy that tackles noise in empirical correlation matrices while retaining signal [12]. In contrast to Ledoit and Wolf's shrinkage method, targeted shrinkage shrinks the noisy components of the correlation matrix while preserving signal components. This approach ensures that the noise reduction does not compromise the clarity of the signal.

Targeted shrinkage utilises the Marchenko-Pastur distribution to separate noise and signal, as explained in the previous section. Subsequently, eigenvalues linked to the signal component, $\lambda \notin [\lambda_-, \lambda_+]$, are used to create an information correlation matrix. On the other hand, noisy eigenvalues, $\lambda \in [\lambda_-, \lambda_+]$, and eigenvectors undergo shrinkage as follows:

$$C_1 = W_I \Lambda_I W_I^T + \alpha W_N \Lambda_N W_N^T + (1 - \alpha) \text{diag}[W_N \Lambda_N W_N^T], \quad (2.22)$$

where W_I and λ_I are eigenvectors and eigenvalues associated with signal, W_N and λ_N are eigenvectors and their corresponding eigenvalues associated with noise. The variable α controls the shrinkage intensity. Higher values of α correspond to less shrinkage, allowing eigenpairs to retain most of their magnitude, while a value of 0 is used for exhaustive shrinkage.

2.3 Hierarchical Risk Parity

HRP is a machine-learning-based algorithm introduced by López de Prado [11]. HRP algorithm addresses some of the caveats in quadratic portfolio optimisation problems that lack the hierarchy

structure. Traditional Mean–variance optimisation algorithms use unstructured correlation matrices. As a result, weights are allowed to move freely between investments without distinguishing between assets’ substitute and complementary nature. However, the HRP algorithm considers the relationships and the hierarchies among assets, which enhance the construction of more diversified, less concentrated portfolios.

The algorithm starts by clustering assets with similar characteristics based on the data. Accordingly, weights are allocated among clusters in a top-down manner. One of the notable advantages of the HRP algorithm is that it utilises the inverse-variance allocation strategy to distribute weights among clusters. The inverse-variance allocation strategy does not require the inverse of the covariance matrix. Thus, the algorithm is less sensitive to estimation errors and can tolerate the presence of ill-conditioned or even singular covariance matrices. Moreover, inverse-variance allocation seeks to balance the risk evenly among clusters, ensuring equalised risk contribution from assets in the portfolio. The algorithm operates through three steps as follows:

- **Tree Clustering**

In the tree-clustering step, the algorithm uses agglomerative hierarchical clustering. It begins with the $N \times N$ correlation matrix, treating each asset as its own cluster. At each iteration, the algorithm merges the two most similar clusters according to a chosen linkage criterion. This bottom-up merging continues until all assets are combined into a single root cluster, producing the full dendrogram that represents the hierarchy as shown in Figure 2.4.

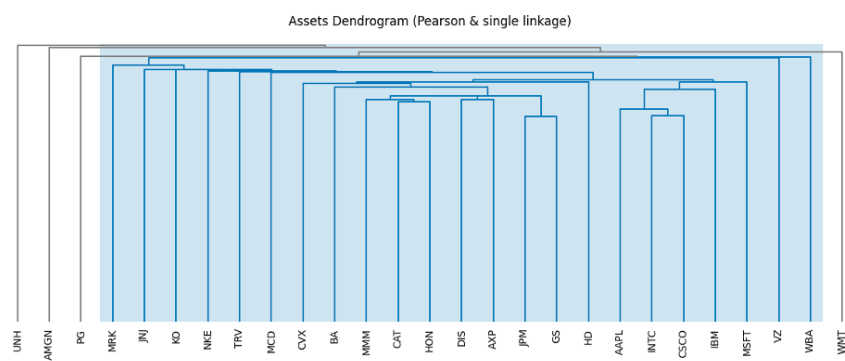


Figure 2.4: Dow Jones: Hierarchical Clustering Dendrogram

In the dendrogram, the choice of the linkage criteria determines how clusters are combined;

examples of commonly used linkage criteria are single linkage, average linkage, and complete linkage.

- **Quasi-diagonalisation**

In this step of the HRP algorithm, the covariance matrix is rearranged in a way that enhances the clustering structure. In particular, rows and columns are reordered without changing the basis. Placing similar assets together and dissimilar assets far apart with large covariances along the diagonal.

- **Recursive bisection**

Recursive bisection is the most crucial step of the algorithm. In this step, actual weights are assigned to the assets in the portfolio. It utilises the Quasi-diagonalised matrix and the inverse-variance allocation as a method to allocate weights. At the end of the tree clustering method, one giant cluster containing all the stocks is at the top of the tree with nested subclusters. Weights are allocated in a top-down manner where each cluster has two child clusters, v_i and v_j . The variance of each cluster is calculated as

$$\tilde{V} = \mathbf{w}^T V \mathbf{w}, \quad (2.23)$$

where

$$\mathbf{w} = \frac{\text{diag}[V]^{-1}}{\text{trace}(\text{diag}[V]^{-1})}. \quad (2.24)$$

Subsequently, the new covariance matrix $\tilde{V}$ is used to compute a weighting factor a :

$$a_i = 1 - \frac{\tilde{V}_i}{\tilde{V}_i + \tilde{V}_j}; a_j = 1 - a_i. \quad (2.25)$$

Finally, the weight of each subcluster is adjusted using the weighting factor a as follows:

$$\mathbf{W}_i = \mathbf{w}_i * a_i, \quad (2.26)$$

$$\mathbf{W}_j = \mathbf{w}_j * a_j. \quad (2.27)$$

The recursive bisection step iterates until each cluster contains one stock. Weights are allocated to stocks using the method described earlier, ensuring risk parity throughout.

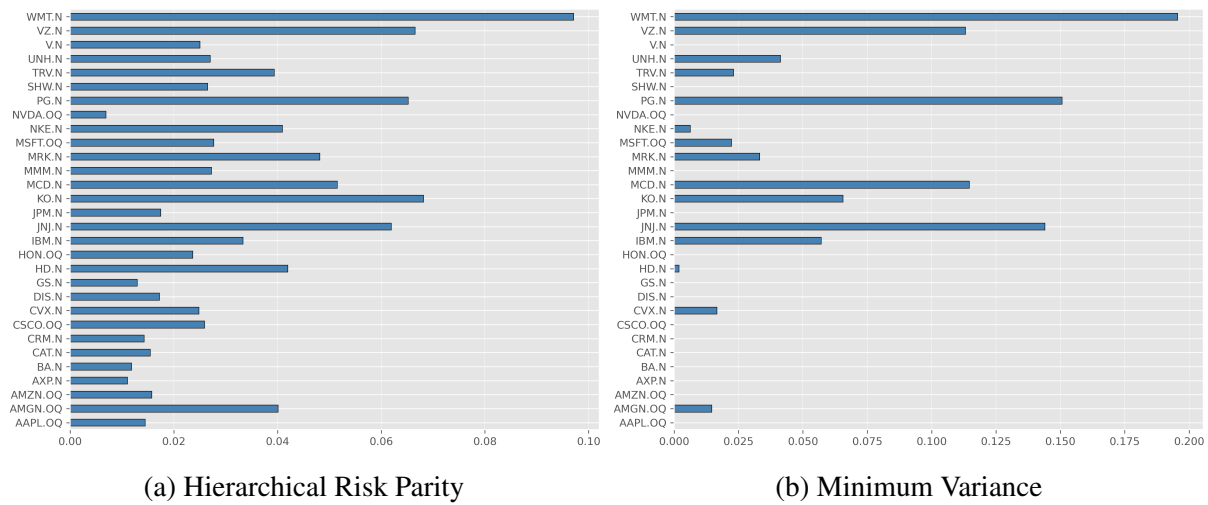


Figure 2.5: Comparison of Weight Distributions: Dow Jones Stocks Using HRP and Minimum Variance

The results of Monte Carlo experiments showed that HRP portfolios achieve lower out-of-sample risk than portfolios constructed using the CLA, and they outperform other risk parity portfolios in terms of risk, as demonstrated by López de Prado[11]. Furthermore, HRP portfolios are more stable and less sensitive to estimation errors because the algorithm does not require the invertibility of the covariance matrix to find the optimal allocations. Finally, utilising the hierarchical clustering drives the algorithm to build more diversified portfolios. The Hierarchical structure guarantees that weights are allocated in a top-down manner, starting at the sector and industry level and progressing down to individual asset level. Figure 2.5 illustrates the weight allocations of Dow Jones stocks using Minimum Variance allocation and HRP allocation.

The portfolio created using the HRP algorithm is more diversified than the Minimum Variance portfolio, as shown in Figure 2.5. Fewer stocks have zero weight, and the allocations are less concentrated. They are more evenly spread out across all stocks in the portfolio. Hence, the HRP distribution boosts diversity and robustness, indicating a more promising investment strategy.

2.4 DCC GARCH

The Dynamic Conditional Correlation GARCH Model (DCC GARCH) is an extension of the Constant Conditional Correlation GARCH model (CCC GARCH), which assumes correlations are fixed over time, and a keystone in modelling financial time series [13]. Unlike static models, the DCC GARCH model is a parametric model that allows conditional volatilities and correlations to vary over time, better capturing the dynamic nature of financial assets. The model strikes a balance between the univariate GARCH model's flexibility and the complexity of multivariate GARCH, allowing the model to estimate large correlation matrices effectively. The conditional covariance matrix under the DCC GARCH model is estimated as follows:

$$H_t = D_t R_t D_t, \quad (2.28)$$

where $H_t \in R^{N \times N}$ is the conditional covariance matrix, $D_t \in R^{N \times N}$ is a diagonal covariance matrix with the root of time varying volatilities on the diagonal ($\sqrt{h_{it}}$), and $R_t \in R^{N \times N}$ is the conditional correlation matrix. The model operates in two steps; the first step involves fitting the univariate GARCH model for individual assets to estimate their conditional volatilities:

$$h_{it} = \omega_i + \sum_{p=1}^{P_i} \alpha_{ip} r_{it-p}^2 + \sum_{q=1}^{Q_i} \beta_{iq} h_{it-q}. \quad (2.29)$$

In the second step, the dynamic correlation matrix R_t is structured using the unconditional covariance matrix of the standardised residuals $\boldsymbol{\varepsilon}_t \in R^{N \times 1}$, denoted $\bar{Q}$, resulting from the first step as follows:

$$R_t = Q_t^{*-1} \cdot Q_t \cdot Q_t^{*-1}, \quad (2.30)$$

$$Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^T + \beta Q_{t-1}, \quad (2.31)$$

where α and β are scalar parameters, and $\bar{Q} = Cov(\boldsymbol{\varepsilon}_t)$ representing the unconditional covariance matrix of residuals. This formulation of Q_t allows for incorporating the nature of residuals into

the calculations of the time-varying correlation matrix R_t . Thus, it accurately represents the relationships between financial assets as observed through their residuals. Finally, the correlation matrix R_t is structured using Q_t and its inverse transpose, Q_t^{*-1} on both sides. This structure ensures proper properties of the correlation matrix, such as unity on the diagonal and positive semidefiniteness.

The DCC GARCH model and its extensions have been widely tested in the literature and received significant attention, especially in estimating the covariance matrix in portfolio management framework. Lee et al. [21] conducted a study to test the efficiency of time-varying correlation in estimating large-dimensional matrices against other models. The results demonstrate the superiority of the DCC GARCH model in measuring value-at-risk (VaR) over other models. Furthermore, the model showed strong performance and applicability when empirically tested for S&P500 and Dow Jones stocks [22].

2.4.1 DCC-NL

Despite the wide applications and popularity of the DCC-GARCH model and its extensions in time series modelling, the model still suffers from computational challenges in high-dimensional settings, particularly when the number of assets is large or when working with high-frequency data. To overcome this, Pakel et al. [23] proposed the composite likelihood as an alternative to the full likelihood approach in the DCC-GARCH model. Instead of maximising the full likelihood function, composite likelihood decomposes the likelihood into smaller, pairwise likelihoods, allowing for more efficient model estimation even in large dimensions. Building upon the composite likelihood approach, Engle et al. [24] further enhanced the DCC-GARCH model by introducing the DCC-NL model. In this framework, they incorporated nonlinear shrinkage in estimating the correlation targeting matrix to address the overfitting and in-sample bias. In their paper, they applied the nonlinear shrinkage estimator introduced in [10] to adjust the eigenvalues of the correlation targeting matrix in the DCC-GARCH model and shrink them toward their true population values. Theoretically, the DCC-NL model constitutes a substantial advancement, effectively addressing overfitting and computational challenges.

2.5 General Framework

The study utilises three distinct datasets containing the total returns of stocks, obtained from LSEG Datastream, from Dow Jones Industrial Average, S&P100, and S&P500, allowing for a comparative analysis of method efficiency across different datasets and stock counts. The data spans January 1990 to December 2024 and is available on both daily and monthly frequencies.

Total return expresses a stock's overall financial gain or loss over a period, incorporating price changes and dividends received. This measure provides a more accurate assessment of portfolio performance than price changes alone, reflecting the actual returns available to an investor.

The analysis involves a rolling window approach to test portfolio performance under different investment strategies and covariance matrix estimations. The optimisation input parameters and the optimal weights are determined for each window using the data within the window. Subsequently, the portfolio is constructed out-of-sample for the following month, incorporating a proportional transaction cost of 0.1% (10 basis points) applied at each rebalancing, deducted from the portfolio return. Finally, the process is repeated by rolling the window forward. For instance, with a 10-year window size of monthly data, the first window's optimisation input parameters and the optimal portfolio weights are determined based on the estimation period of 10 years (120 months). The following month's returns are used to calculate the first out-of-sample portfolio return. Subsequently, the estimation period is rolled forward by a month. After generating the out-of-sample returns for each month, the monthly returns are averaged, and the average is annualised to obtain the final performance measure.

The analysis focuses on risk-based optimisation objectives, which depend solely on the covariance matrix estimate, allowing for a direct assessment of each estimator's effect on portfolio performance. The objectives considered are:

- Minimum Variance
- Maximum Decorrelation
- Maximum Diversification Ratio

- The application of HRP.

For each optimisation objective and rolling window, portfolios are constructed using different covariance matrix estimators employing denoising, shrinkage, and DCC techniques. The portfolio performance is then evaluated under each estimator, thoroughly comparing each estimator's effect on the portfolio return and risk. The empirical covariance matrix is initially used as input for the optimisation problem to serve as a reference. Then, the covariance matrix is refined and estimated using parametric and nonparametric techniques. For denoising, López de Prado's method is applied by replacing noisy eigenvalues with their mean, and extended by using alternative replacements, such as the 75th percentile, the 90th percentile, and the average weight of the quantiles. Shrinkage techniques include linear shrinkage with different shrinkage targets and nonlinear shrinkage. Specifically, nonlinear shrinkage is implemented using the QIS and analytical nonlinear shrinkage methods. In addition, the DCC-GARCH model is employed to estimate time-varying covariance matrices. The results also incorporate combinations where the DCC-GARCH covariance matrix is further refined through linear shrinkage and denoising techniques. Finally, the DCC-NL model is also used to estimate the covariance matrix. Portfolio performance under each estimate is evaluated on the basis of risk, return, and Sharpe ratio.

2.6 Results and Analysis

This section presents the results of the portfolio performance evaluation under different covariance matrix estimators and optimisation objectives. The results are organised into subsections, where each portfolio optimisation strategy is analysed separately.

The performance of each strategy is evaluated using two tables, one for monthly data and one for daily data, reporting results across different dataset sizes ($N = 30$, $N = 100$, and $N = 500$) and under various covariance matrix estimation techniques.

Portfolio performance is assessed using annualised return, annualised standard deviation, and annualised Sharpe ratio, providing insights into the efficacy of different covariance matrix estimates under each investment strategy.

2.6.1 Minimum Volatility

Tables 2.1 and 2.2 present the annualised performance metrics for the minimum volatility portfolio using monthly and daily data respectively. Across all estimators and dataset sizes, the monthly specifications consistently produce lower portfolio volatility than their daily counterparts, reflecting the higher levels of microstructure noise and short-run price fluctuations embedded in daily return data.

Table 2.1: Annualised Performance Metrics for Minimum Volatility Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	10.40	12.97	0.8024	11.31	12.48	0.9058	10.48	11.85	0.8844
Denoising									
Mean of Eigenvalues	9.81	12.64	0.7760	10.79	12.46	0.8657	10.89	11.12	0.9793
75th Percentile of Eigenvalues	9.76	12.54	0.7784	10.73	12.25	0.8760	10.58	11.25	0.9398
90th Percentile of Eigenvalues	9.82	12.49	0.7867	10.72	12.10	0.8860	11.05	11.31	0.9765
Average Weight of Quantiles	9.82	12.67	0.7754	10.80	12.54	0.8606	10.58	11.38	0.9302
Linear Shrinkage									
Identity Matrix	10.95	12.81	0.8550	11.40	12.18	0.9358	11.20	11.52	0.9727
Constant Variance	10.19	12.79	0.7964	11.47	12.21	0.9396	10.83	11.54	0.9387
Single Factor	9.74	12.66	0.7696	11.04	12.20	0.9049	10.55	11.35	0.9290
Constant Correlation	9.22	12.59	0.7317	10.22	12.25	0.8338	10.51	11.69	0.8990
Non-Linear Shrinkage									
QIS	10.11	12.71	0.7952	11.34	12.14	0.9342	10.97	11.56	0.9492
Analytical Nonlinear Shrinkage	10.10	12.71	0.7945	11.35	12.15	0.9337	11.08	11.39	0.9720
Dynamic Estimates									
DCC	10.98	13.24	0.8289	11.73	12.00	0.9774	10.45	11.61	0.8999
DCC Shrinkage	10.66	12.90	0.8263	11.18	11.87	0.9415	11.12	11.29	0.9850
DCC Denoising	10.59	13.12	0.8074	11.50	11.98	0.9601	10.91	11.47	0.9508
DCC-NL	10.67	12.92	0.8255	11.39	12.17	0.9366	10.83	10.79	1.0032

The strongest result in Table 2.1 is DCC-NL at $N = 500$ monthly, which achieves a Sharpe ratio of 1.0032 and a standard deviation of 10.79%. This is the only Sharpe ratio above one in both tables and the lowest standard deviation across all daily and monthly specifications. Relative to the raw sample covariance at the same dataset size, $SR = 0.8844$ and $SD = 11.85\%$, DCC-NL delivers a 13.4% improvement in risk-adjusted performance and an 8.9% reduction in annualised risk. The estimator combines time-varying correlation dynamics through the DCC

model, estimated via composite likelihood to remain feasible at large N , with nonlinear shrinkage applied to the eigenvalues of the correlation targeting matrix to correct for in-sample bias and overfitting. This makes the dynamic estimates both computationally tractable across 500 assets and robust to noise in the long-run correlation structure, which is why the gains are concentrated at large N where both challenges are most critical.

The denoising estimators display a pattern that maps directly onto the ratio $q = N/T$. At $N = 500$ monthly, where approximately 500 assets are estimated from 120 monthly observations, all denoising variants improve on the raw sample covariance: the mean eigenvalue replacement achieves $SR = 0.9793$ and the 90th percentile reaches $SR = 0.9765$, both well above raw $SR = 0.8844$. Replacing noise-dominated eigenvalues lifts $\lambda_{\min}$, reduces the condition number $\kappa(C) = \lambda_{\max}/\lambda_{\min}$, and produces a matrix that behaves more reliably under inversion. The benefits of eigenvalue replacement are greatest in this setting, where estimation noise dominates the covariance matrix and the asset universe is larger than the number of available observations. For $N = 100$ and $N = 30$, where $q \approx 0.83$ and $q \approx 0.25$ respectively, the covariance matrix contains less estimation noise relative to genuine structure and a larger fraction of the eigenvalue spectrum may carry economically meaningful information. In these regimes, replacing eigenvalues classified as noise introduces estimation bias that outweighs the variance reduction, and denoising does not improve on the raw sample covariance.

At daily frequency Table 2.2, the benefits of refinement largely disappear. The rolling estimation window of 120 months corresponds to approximately $T = 2520$ daily observations, assuming 21 trading days per month, giving $q \approx 0.20$, 0.04, and 0.01 for $N = 500$, 100, and 30 respectively. In this low- q regime the covariance matrix is already reasonably well-conditioned and estimation noise constitutes a small fraction of the spectrum. Denoising and shrinkage techniques have little genuine noise to correct, and risk discarding real economic signal instead. The DCC family is the most affected: DCC at $N = 30$ daily yields $SR = 0.5587$ and DCC Denoising reaches only $SR = 0.5402$, against a raw estimate of $SR = 0.6705$, with similar deterioration observed at $N = 100$ and $N = 500$. With abundant observations relative to the asset universe, the model fits noise dynamics rather than genuine correlation variation, and the resulting covariance matrices destabilise rather than stabilise the optimised weights.

Table 2.2: Annualised Performance Metrics for Minimum Volatility Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	10.13	15.11	0.6705	12.25	15.08	0.8122	12.69	14.08	0.9017
Denoising									
Mean of eigenvalues	9.73	15.04	0.6465	12.06	15.28	0.7890	12.02	14.53	0.8274
75th percentile of eigenvalues	9.70	15.05	0.6446	12.04	15.19	0.7925	12.18	14.46	0.8424
90th percentile of eigenvalues	9.76	15.05	0.6483	11.97	15.13	0.7910	12.39	14.39	0.8613
Average Weight of Quantiles	9.72	15.05	0.6458	12.06	15.28	0.7890	11.96	14.57	0.8213
Linear Shrinkage									
Identity matrix	11.45	15.78	0.7255	12.10	15.18	0.7973	12.58	14.31	0.8789
Constant Variance	10.12	15.11	0.6699	12.23	15.07	0.8119	12.69	14.07	0.9018
Single Factor	10.09	15.09	0.6686	12.21	15.07	0.8098	12.54	14.07	0.8912
Constant Correlation	9.99	15.09	0.6618	12.15	15.09	0.8051	12.47	14.12	0.8832
Non-Linear Shrinkage									
QIS	10.12	15.11	0.6696	12.24	15.05	0.8136	12.63	14.05	0.8989
Analytical Nonlinear Shrinkage	10.03	15.23	0.6590	12.73	15.45	0.8239	12.63	14.04	0.8995
Dynamic Estimates									
DCC	8.17	14.62	0.5587	8.73	13.98	0.6246	11.06	14.52	0.7613
DCC Shrinkage	10.36	15.27	0.6784	10.75	14.40	0.7465	11.64	14.59	0.7980
DCC Denoising	7.89	14.61	0.5402	8.60	14.04	0.6129	10.98	14.44	0.7601
DCC-NL	8.560	14.55	0.5881	9.21	14.11	0.6527	10.43	12.40	0.8410

2.6.2 Maximum Decorrelation

Tables 2.3 and 2.4 present the annualised performance metrics for the maximum decorrelation portfolio. At monthly frequency,

Table 2.3, consistent with the minimum volatility results, the best performing estimator is DCC-NL at $N = 500$, achieving the highest Sharpe ratio in the table at $SR = 0.9890$ and the lowest portfolio volatility across both maximum decorrelation tables at $SD = 10.23\%$, confirming its ability to produce well-conditioned correlation estimates at large dimensions. Notably, a key departure from the minimum volatility setting is that denoising now improves on the raw sample covariance across all three dataset sizes, with the 75th and 90th percentile variants delivering the most consistent gains. The improvement is particularly striking at $N = 100$, where the raw sample covariance yields $SR = 0.5755$ and $SD = 17.74\%$; the 90th percentile denoising reduces portfolio volatility to 11.84% , a reduction of 33.3% , and raises SR to 0.9035 . At $N = 100$,

Table 2.3: Annualised Performance Metrics for Maximum Decorrelation Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	9.36	13.80	0.6784	10.21	17.74	0.5755	8.93	11.71	0.7623
Denoising									
Mean of Eigenvalues	9.47	12.81	0.7393	10.88	12.65	0.8595	8.71	10.90	0.7989
75th Percentile of Eigenvalues	9.53	12.57	0.7581	10.84	12.22	0.8872	8.81	10.95	0.8040
90th Percentile of Eigenvalues	9.61	12.47	0.7701	10.69	11.84	0.9035	10.01	10.60	0.9444
Average Weight of Quantiles	9.39	12.85	0.7308	10.96	12.86	0.8524	7.75	11.35	0.6827
Linear Shrinkage									
Identity Matrix	10.84	12.80	0.8468	11.43	11.98	0.9543	10.01	10.77	0.9295
Constant Variance	9.74	12.94	0.7525	11.99	12.46	0.9624	9.57	11.12	0.8604
Single Factor	9.06	12.92	0.7014	10.90	12.17	0.8954	9.48	10.78	0.8795
Constant Correlation	7.84	12.89	0.6085	8.84	13.08	0.6756	6.89	12.27	0.5616
Non-Linear Shrinkage									
QIS	9.62	12.87	0.7478	11.46	12.34	0.9294	9.59	10.87	0.8822
Analytical Nonlinear Shrinkage	9.68	12.89	0.7513	11.66	12.35	0.9443	9.42	10.81	0.8714
Dynamic Estimates									
DCC	9.08	14.24	0.6378	9.05	14.67	0.6173	7.97	10.82	0.7365
DCC Shrinkage	10.82	12.76	0.8056	10.62	11.72	0.9065	9.81	10.58	0.9264
DCC Denoising	9.43	13.68	0.6892	10.83	11.86	0.9132	9.57	10.37	0.9231
DCC-NL	10.17	13.38	0.7600	10.25	12.14	0.8440	10.12	10.23	0.9890

all estimators improve on the raw sample covariance, including denoising, linear shrinkage, nonlinear shrinkage, and the DCC family, suggesting that at this dimension the raw correlation matrix is sufficiently noisy that any form of regularisation delivers meaningful gains.

At daily frequency, Table 2.4, the results reflect a different estimation environment. Since maximum decorrelation is constructed from the correlation matrix alone and is therefore insensitive to the scale of individual asset volatilities, what matters is the reliability of the estimated correlation structure. With approximately $T = 2,520$ daily observations, the correlation matrix is estimated with considerable precision, and the raw sample covariance at $N = 500$ daily achieves $SR = 1.0272$. As in the minimum volatility setting, once the number of observations is sufficiently large relative to the asset universe, the marginal gains from refinement shrink. The low- q regime leaves little noise to correct, and any further adjustment to an already well-estimated correlation matrix risks discarding genuine structure rather than noise, which is precisely what the table shows. However, in this low-noise setting, nonlinear shrinkage still offers a small benefit:

Table 2.4: Annualised Performance Metrics for Maximum Decorrelation Portfolio Using Daily Data (Datasets: N = 30, 100, 500)

	N = 30			N = 100			N = 500		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	9.14	14.84	0.6158	12.65	14.38	0.8798	12.53	12.20	1.0272
Denoising									
Mean of Eigenvalues	8.67	14.86	0.5836	11.23	14.89	0.7541	11.13	13.12	0.8478
75th Percentile of Eigenvalues	8.79	14.82	0.5930	11.26	14.61	0.7703	11.62	12.92	0.8988
90th Percentile of Eigenvalues	8.97	14.79	0.6066	11.32	14.45	0.7835	12.01	12.78	0.9397
Average Weight of Quantiles	8.61	14.87	0.5791	11.21	14.91	0.7521	11.13	13.27	0.8390
Linear Shrinkage									
Identity Matrix	11.17	15.62	0.7150	12.00	14.54	0.8248	12.90	12.72	1.0148
Constant Variance	9.18	14.84	0.6191	12.62	14.34	0.8800	12.54	12.12	1.0343
Single Factor	9.03	14.82	0.6095	12.53	14.32	0.8748	12.71	12.06	1.0538
Constant Correlation	8.66	14.84	0.5832	12.12	14.34	0.8449	11.92	12.17	0.9798
Non-Linear Shrinkage									
QIS	9.16	14.84	0.6176	12.58	14.31	0.8790	12.78	12.02	1.0634
Analytical Nonlinear Shrinkage	9.07	14.97	0.6054	13.62	14.72	0.9250	12.80	12.02	1.0652
Dynamic Estimates									
DCC	6.43	14.43	0.4458	8.57	13.71	0.6253	7.86	15.19	0.5177
DCC Shrinkage	9.70	14.98	0.6473	10.43	13.73	0.7592	10.26	14.17	0.7243
DCC Denoising	6.05	14.30	0.4232	10.61	13.69	0.7417	9.57	13.72	0.6978
DCC-NL	6.83	14.24	0.4796	9.03	13.69	0.6598	8.10	11.95	0.6772

the Analytical Nonlinear Shrinkage estimator achieves the highest Sharpe ratio in the table (SR = 1.0652), with QIS close behind (SR = 1.0634), suggesting that adaptive, eigenvalue-specific shrinkage can extract modest gains even when little noise remains to correct.

2.6.3 HRP

Tables 2.5 and 2.6 present the annualised performance metrics for the HRP portfolio. The HRP algorithm utilises hierarchical clustering to group assets based on their similarities and does not require the inversion of the covariance matrix, which significantly reduces its sensitivity to estimation errors and to the choice of covariance matrix estimator. The immediate consequence of this robustness is visible in the tables: Sharpe ratio ranges are notably tight across all estimators and dataset sizes, both monthly and daily, and no single estimator dominates by a significant margin.

Nevertheless, a consistent pattern does emerge: the Single Factor estimator delivers the

Table 2.5: Annualised Performance Metrics for HRP Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	11.58	13.42	0.8631	11.95	13.11	0.9116	13.09	13.35	0.9805
Denoising									
Mean of eigenvalues	11.45	13.40	0.8545	11.76	13.22	0.8893	13.07	13.37	0.9774
75th percentile of eigenvalues	11.42	13.32	0.8573	11.83	13.15	0.8991	12.96	13.33	0.9723
90th percentile of eigenvalues	11.40	13.31	0.8565	11.81	13.02	0.9064	13.13	13.29	0.9882
Average Weight of Quantiles	11.40	13.41	0.8503	11.86	13.14	0.9021	13.01	13.31	0.9770
Linear Shrinkage									
Identity matrix	11.85	13.61	0.8709	11.74	13.43	0.8741	13.03	13.63	0.9558
Constant Variance	11.42	13.41	0.8513	11.73	13.17	0.8906	12.99	13.49	0.9630
Single Factor	11.40	13.11	0.8694	11.64	12.72	0.9156	12.77	12.69	1.0067
Constant Correlation	11.53	13.48	0.8550	11.72	13.26	0.8841	13.03	13.65	0.9550
Non-Linear Shrinkage									
QIS	11.16	13.37	0.8349	11.49	13.12	0.8759	12.94	13.30	0.9728
Analytical Nonlinear Shrinkage	11.12	13.42	0.8286	11.58	13.03	0.8889	12.91	13.20	0.9779
Dynamic Estimates									
DCC	13.78	14.72	0.9362	11.41	12.67	0.9004	12.67	12.88	0.9836
DCC Shrinkage	13.73	14.73	0.9325	11.59	13.07	0.8867	12.68	13.16	0.9635
DCC Denoising	14.07	14.71	0.9561	11.51	12.69	0.9070	12.66	12.86	0.9847
DCC-NL	11.20	13.13	0.8533	11.46	12.67	0.9046	12.76	12.77	0.9993

strongest risk-adjusted performance at $N = 100$, and $N = 500$ across both monthly and daily frequencies.

Since the HRP algorithm constructs portfolios using a hierarchical clustering procedure based on the estimated correlation matrix, the structure of that matrix can influence the resulting dendrogram. The Single Factor estimator, based on Sharpe's single-index model, shrinks the covariance matrix towards a target in which all pairwise correlations are expressed through each asset's beta to the market, reinforcing the dominant source of systematic co-movement across the asset universe. This simple yet structured market-driven representation yields more stable asset groupings and a more reliable hierarchical tree, which may explain the marginal but consistent outperformance of the Single Factor estimator at $N = 100$ and $N = 500$ across both monthly and daily frequencies.

Table 2.6: Annualised Performance Metrics for HRP Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	12.83	16.75	0.7661	13.22	16.87	0.7839	14.62	17.23	0.8482
Denoising									
Mean of eigenvalues	12.56	16.78	0.7484	12.97	17.12	0.7577	14.55	17.36	0.8380
75th percentile of eigenvalues	12.76	16.78	0.7602	13.11	17.11	0.7661	14.63	17.34	0.8435
90th percentile of eigenvalues	12.64	16.73	0.7554	13.08	17.04	0.7674	14.67	17.28	0.8489
Average Weight of Quantiles	12.71	16.79	0.7572	13.04	17.15	0.7602	14.41	17.37	0.8295
Linear Shrinkage									
Identity matrix	13.10	17.24	0.7602	13.22	17.23	0.7672	14.67	17.51	0.8376
Constant Variance	12.89	16.75	0.7692	13.29	16.90	0.7866	14.65	17.22	0.8508
Single Factor	12.79	16.75	0.7635	13.31	16.88	0.7884	14.66	17.17	0.8542
Constant Correlation	12.81	16.77	0.7641	13.19	16.90	0.7810	14.58	17.27	0.8445
Non-Linear Shrinkage									
QIS	12.85	16.75	0.7675	13.09	16.89	0.7750	14.68	17.22	0.8525
Analytical Nonlinear Shrinkage	12.58	16.75	0.7508	13.31	16.89	0.7883	14.67	17.21	0.8526
Dynamic Estimates									
DCC	11.61	16.11	0.7208	12.30	16.01	0.7684	13.49	16.29	0.8284
DCC Shrinkage	12.27	16.68	0.7354	12.36	16.49	0.7497	13.44	16.67	0.8062
DCC Denoising	11.59	16.29	0.7154	11.92	16.11	0.7400	16.00	19.90	0.8041
DCC-NL	11.71	16.07	0.7285	12.34	16.05	0.7687	13.44	16.05	0.8371

2.6.4 Maximum Diversification

Tables 2.7 and 2.8 present the annualised performance metrics for the maximum diversification portfolio. Unlike the previous strategies, maximum diversification maximises the diversification ratio, defined as the weighted average of individual asset volatilities divided by portfolio volatility. Consequently, the optimisation depends simultaneously on estimates of individual asset volatilities and the full covariance matrix. Estimation errors can therefore affect both the numerator and denominator of the objective function, making portfolio performance particularly dependent on the quality of covariance estimation. As a result, improvements in covariance matrix estimation through denoising or shrinkage have the potential to translate more directly into improved portfolio allocations than in strategies that rely on a more limited subset of covariance information.

This sensitivity is reflected directly in Table 2.7 across all dataset sizes and estimator groups.

Table 2.7: Annualised Performance Metrics for Maximum Diversification Portfolio Using Monthly Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	14.51	15.72	0.9235	15.09	23.19	0.6509	13.61	13.63	0.998
Denoising									
Mean of eigenvalues	14.59	13.61	1.0725	16.98	13.50	1.2575	14.05	13.05	1.0811
75th percentile of eigenvalues	13.93	13.31	1.0464	16.17	12.72	1.2719	14.21	13.01	1.0901
90th percentile of eigenvalues	13.33	13.24	1.0065	15.08	12.27	1.2291	14.91	13.01	1.1511
Average Weight of Quantiles	14.53	13.59	1.0696	17.29	13.93	1.2418	16.21	14.11	1.1521
Linear Shrinkage									
Identity matrix	12.53	13.18	0.9512	13.38	11.98	1.1161	11.91	10.80	1.1032
Constant Variance	13.74	13.91	0.9876	15.91	12.86	1.2368	12.65	12.44	1.0233
Single Factor	13.84	14.01	0.9879	16.10	12.71	1.2663	13.09	12.11	1.0871
Constant Correlation	13.32	13.38	0.9955	15.80	12.57	1.2565	12.98	12.75	1.0241
Non-Linear Shrinkage									
QIS	13.37	13.60	0.9831	15.08	12.38	1.2183	13.42	12.56	1.0731
Analytical Nonlinear Shrinkage	13.46	13.60	0.9896	15.40	12.46	1.2361	13.74	13.01	1.0621
Dynamic Estimates									
DCC	13.99	15.31	0.9141	14.13	21.18	0.6670	11.26	10.73	1.0497
DCC Shrinkage	12.09	13.13	0.9209	12.56	11.99	1.0473	11.77	10.46	1.1065
DCC Denoising	13.95	13.88	1.0050	16.16	13.33	1.2122	12.76	11.71	1.0896
DCC-NL	14.29	13.89	1.0281	16.28	13.09	1.2433	12.12	12.04	1.0112

All refinements to the raw sample covariance improve risk-adjusted performance, a pattern that does not hold as consistently in any of the previous strategies. The denoising family is the strongest group overall: the 75th percentile of eigenvalues achieves the highest Sharpe ratio in the monthly table at $N = 100$ with $SR = 1.2719$, the mean of eigenvalues leads at $N = 30$ with $SR = 1.0725$, and the average weight of quantiles and the 90th percentile deliver the best performance at $N = 500$. Within the dynamic estimator group, plain DCC on its own struggles at $N = 100$, yielding $SR = 0.6670$, and $SD = 21.18\%$, close to the raw sample covariance at $SR = 0.6509$. However, combining DCC with shrinkage or denoising restores performance substantially: DCC Shrinkage reaches $SR = 1.0473$, DCC Denoising $SR = 1.2122$, and DCC-NL $SR = 1.2433$ at $N = 100$.

Taken together, these results reinforce how sensitive the maximum diversification objective is to the quality of covariance estimation, with covariance refinement being most rewarded here relative to the other strategies in this chapter, translating directly into improved portfolio

outcomes.

Table 2.8: Annualised Performance Metrics for Maximum Diversification Portfolio Using Daily Data (Datasets: $N = 30, 100, 500$)

	$N = 30$			$N = 100$			$N = 500$		
	R	SD	SR	R	SD	SR	R	SD	SR
Raw Estimates									
Raw Covariance Matrix	16.78	17.25	0.9730	20.69	17.63	1.1736	21.40	17.16	1.2466
Denoising									
Mean of eigenvalues	16.59	16.82	0.9860	22.05	19.48	1.1320	20.26	19.85	1.0206
75th percentile of eigenvalues	16.14	16.75	0.9634	20.77	18.20	1.1411	20.40	18.15	1.1239
90th percentile of eigenvalues	15.63	16.69	0.9362	19.53	17.22	1.1344	20.16	16.59	1.2151
Average Weight of Quantiles	16.47	16.82	0.9792	22.17	19.71	1.1251	20.26	20.64	0.9816
Linear Shrinkage									
Identity matrix	13.41	16.33	0.8210	14.57	15.20	0.9587	15.97	13.16	1.2134
Constant Variance	16.61	17.13	0.9693	20.43	17.39	1.1750	21.19	16.61	1.2759
Single Factor	16.69	17.16	0.9730	20.67	17.50	1.1817	21.58	16.49	1.3088
Constant Correlation	16.45	17.04	0.9652	20.39	17.38	1.1732	21.29	16.44	1.2951
Non-Linear Shrinkage									
QIS	16.62	17.11	0.9715	20.34	17.26	1.1781	20.50	15.31	1.3396
Analytical Nonlinear Shrinkage	16.43	17.21	0.9545	20.65	17.50	1.1798	20.54	15.31	1.3416
Dynamic Estimates									
DCC	14.17	17.11	0.8281	16.45	18.80	0.8750	11.65	14.49	0.8041
DCC Shrinkage	12.07	15.75	0.7664	12.67	14.44	0.8776	12.35	14.51	0.8514
DCC Denoising	14.37	16.58	0.8867	15.99	18.97	0.8428	13.54	12.25	0.8880
DCC-NL	14.52	16.98	0.8551	15.40	19.10	0.8064	15.11	16.28	0.9281

At daily frequency Table 2.8, the results follow the same logic as the previous strategies. With approximately $T = 2520$ daily observations, the raw sample covariance is already well-estimated and most refinements either add little or risk removing genuine signal, with non-linear shrinkage providing modest gains at $N = 500$. The DCC family deteriorates substantially across all dataset sizes, as fitting time-varying dynamics to daily returns causes the model to absorb microstructure noise rather than genuine co-movement, destabilising the diversification ratio rather than improving it.

2.7 Conclusion

This chapter comprehensively analyses the effect of different parametric and nonparametric covariance matrix estimates on portfolio optimisation results. We evaluated each estimator's

performance under various risk-based investment strategies using a rolling window approach and out-of-sample testing. We tested various covariance matrix estimates, including denoising, linear and nonlinear shrinkage, and DCC estimates. We extended De Prado's denoising framework by considering different quantile levels for eigenvalue substitution. Moreover, we combined the standard DCC estimate with denoising and linear shrinkage to explore the potential for improved performance. Finally, we employed the DCC-NL estimate, which is designed to estimate covariance matrices robustly in high-dimensional settings.

The results reveal that the value of covariance refinement depends fundamentally on the estimation environment. Monthly data, which pairs a large asset universe with a relatively short observation window, produces a sample covariance matrix that is heavily contaminated by noise. In this setting, refinement techniques consistently improve portfolio outcomes regardless of strategy. Daily data operates under opposite conditions: with a substantially larger number of observations relative to assets, the sample covariance is already well-estimated and further adjustment tends to eliminate genuine signal. The DCC family is particularly sensitive to this, as fitting time-varying correlation dynamics to daily returns captures transient price fluctuations rather than structural changes in correlation.

Within the monthly setting, estimator performance varies by investment objective. DCC-NL delivers the strongest results under minimum volatility and maximum decorrelation, as it simultaneously accounts for time-varying correlation structure and corrects eigenvalue distortion. The maximum diversification objective responds more strongly to denoising, reflecting the fact that estimation error enters through both the asset volatilities and the full covariance matrix, making eigenvalue correction especially consequential. In particular, the quantile-based extension of the denoising framework proves most effective here. At every dataset size in the monthly setting, an eigenvalue-quantile variant produces the top Sharpe ratio: the mean of eigenvalues performs best at $N = 30$ ($SR = 1.0725$), with the average weight of quantiles close behind ($SR = 1.0696$); at $N = 100$, the 75th percentile produces the highest Sharpe ratio recorded across all monthly specifications in this chapter ($SR = 1.2719$); and at $N = 500$, the average weight of quantiles ($SR = 1.1521$) and the 90th percentile ($SR = 1.1511$) are nearly identical at the top.

Linear shrinkage produces reliable improvements across objectives without emerging as the

dominant approach in any particular case. HRP occupies a separate position in this analysis: its allocation mechanism operates through hierarchical clustering rather than matrix inversion, making it largely insensitive to the choice of estimator.

The identified estimators are carried forward into the next chapter, where they are embedded within a clustering-based allocation framework that combines robust covariance inputs with machine learning return forecasts. This integration allows the portfolio construction process to draw simultaneously on well-estimated covariance structure and predictive return signals, with the aim of building portfolios that are both statistically robust and economically informed.

From Return Forecasts to Optimal Portfolios: Integrating Machine Learning with Nested Clustered Optimisation

Chapter 2 established that refining the covariance matrix substantially improves portfolio stability, yet an accurately estimated covariance matrix controls portfolio risk but cannot substitute for a reliable estimate of expected returns. Estimating the expected return vector with sufficient precision to improve investment outcomes has proven to be the more persistent and less tractable challenge in portfolio construction. This chapter addresses that gap directly. Machine-learning return forecasts and refined covariance estimators are jointly embedded within the Nested Clustered Optimisation (NCO) algorithm and evaluated across alternative clustering regimes and distance metrics, with the aim of establishing how each component contributes to portfolio performance.

3.1 Return Predictability

Return predictability sits at the heart of quantitative portfolio management. Markowitz [1] established expected returns as a central input to mean–variance optimisation, yet estimating them with sufficient precision to improve investment outcomes has proven one of the most

challenging problems in empirical finance. Financial market data are dominated by noise. The remaining signal is not only weak but also structurally unstable, shifting in ways that make any fixed model a poor approximation of the underlying data-generating process. Part of this instability is self-reinforcing: as investors identify and act on predictive signals, those signals become priced in and lose their forecasting power [25]. The process is further complicated by continuous external perturbations, including shifts in policy, macroeconomic shocks, and the arrival of new investor types, all of which constantly reshape the environment in which returns are generated. The result is an estimation problem characterised by a weak signal, high dimensionality, and nonlinear dynamics. These are precisely the conditions that the conventional approaches dominant in the early literature were not designed to meet.

The earliest systematic attempts to forecast equity returns were grounded in the time-series tradition, most notably through Autoregressive Integrated Moving Average (ARIMA) models. These models characterise returns as a linear function of their own lagged values and past forecast errors, and assume that the underlying series is stationary, with constant mean, variance, and autocorrelation structure [26]. In practice, neither condition holds reliably in financial time series data. Returns exhibit structural breaks, regime shifts, and persistent responsiveness to macroeconomic developments that a fixed autoregressive structure cannot fully capture. Extensions such as ARCH and GARCH models partially address this by allowing the conditional variance to vary over time, reflecting the well-documented tendency of volatility to cluster in financial markets [27][28]. However, these models target the second moment of the return distribution rather than the first. The conditional mean, which is the central object of interest in return forecasting, remains a linear function of past returns and is typically assumed to be stable. Beyond this, these frameworks do not incorporate the macroeconomic and financial predictors that later research identifies as carrying meaningful predictive information about future returns ([29], [30], and [31]). Therefore, when taken alone, these time-series models provide only a limited foundation for forecasting equity returns.

The limitations of pure time-series approaches prompted a shift toward predictor-based linear regression models, in which macroeconomic and financial variables serve as conditioning information for return forecasts. Variables such as the dividend yield, price-earnings ratio,

interest rate spreads, and default premia were shown to carry meaningful in-sample predictive power, following the frameworks established by Campbell and Shiller [32] and Fama and French [30, 31]. This generated considerable optimism about the prospects for return forecasting. Out-of-sample performance, however, told a different story. Many of these predictors broke down during the 1990s bull market, and studies of real-time forecasting found predictability either confined to particular sub-samples or largely absent. Goyal and Welch [33] delivered the definitive assessment: none of the conventional predictors appears capable of systematically beating a simple historical mean forecast out-of-sample when used in linear regression models. Two responses emerged in the literature. Campbell and Thompson [34] showed that imposing simple economically motivated restrictions on forecasting models, such as constraining coefficients to have the correct sign, meaningfully improves out-of-sample performance. Rapach et al. [35] demonstrated that combining forecasts across multiple models recovers further predictive signal that individual regressions cannot reliably capture. These limitations, namely fragile predictive relationships, overfitting in small samples, and the inability to capture nonlinear dynamics, motivate the application of machine learning methods, which are better suited to the high-dimensional, noisy environment of equity return forecasting.

3.1.1 Linear Models

Moving beyond univariate time-series models, linear regression frameworks allow a broader set of macroeconomic and financial predictors to enter the forecast directly. The simplest such specification is Ordinary Least Squares (OLS), which models expected returns as a linear combination of predictor variables:

$$r_{t+1} = \mathbf{x}_t^\top \mathbf{w} + \varepsilon_{t+1}, \quad (3.1)$$

where $\mathbf{x}_t \in R^{P \times 1}$ is the vector of predictor variables at time t , $\mathbf{w} \in R^{P \times 1}$ is the vector of regression coefficients, P is the number of predictors, and $\mathbf{x}_t^\top \mathbf{w}$ corresponds to the model's predicted return, obtained as a linear combination of the predictors weighted by the coefficient vector $\mathbf{w}$, and ε_{t+1} is the error term. The model aims to find the set of coefficients that best estimate the target variable by minimising the residual sum of squares between the observed target and the

forecasted values, by solving the following optimisation problem:

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{r}\|_2^2. \quad (3.2)$$

The simple linear regression model encounters challenges when the number of predictors is high and approaches the number of observations. In this case, the model becomes insufficient and struggles to estimate a large number of coefficients given limited observations. As a result, the model tends to overfit noise rather than structure, leading to unreliable coefficients that are bound to fail in out-of-sample forecasting. One way to address overfitting is to impose regularisation on the model by adding a penalty term to the loss function. Ridge Regression extends the simple linear model by adjusting the OLS to include an L_2 penalty term that aims to stabilise the coefficients and reduce their overall variance [36]. The L_2 regulariser controls the magnitude of coefficients by shrinking them toward zero, which is particularly useful when predictors are highly collinear. As a result, Ridge Regression tends to produce more precise and lower variance estimates than the standard OLS estimator. A Ridge regression estimator is obtained by solving the following optimisation problem:

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{r}\|_2^2 + \lambda \|\mathbf{w}\|_2^2, \quad (3.3)$$

where $\|\mathbf{w}\|_2^2$ is the L_2 regularisation term and $\lambda \geq 0$ represents the regularisation parameter that controls the strength of the penalty applied to coefficients.

LASSO Regression [37] is another extension of the simple linear model, but instead of using the L_2 penalty, it introduces an L_1 penalty term in the loss function. Unlike the L_2 regulariser, which shrinks all coefficients toward zero, the L_1 penalty term performs variable selection by setting weak and irrelevant coefficients exactly to zero. In practice, this means LASSO often keeps only the more relevant variables and drops the rest, giving a simpler model that is easier to interpret and less likely to overfit. LASSO estimates the model parameters by minimising the objective below:

$$\min_{\mathbf{w}} \|\mathbf{X}\mathbf{w} - \mathbf{r}\|_2^2 + \lambda \|\mathbf{w}\|_1, \quad (3.4)$$

where $\|\mathbf{w}\|_1$ denotes the L_1 norm, obtained by summing the absolute values of the coefficients, and $\lambda \geq 0$ is a tuning parameter that controls the strength of the L_1 penalty.

Since Ridge and LASSO offer complementary advantages, namely Ridge stabilising coefficients when predictors are highly correlated and LASSO selecting variables by shrinking some to zero, it is often useful to have a model that combines both. Elastic Net regression was designed for this purpose. It blends the two regularisers into a single framework by minimising a loss function that includes the usual least-squares term along with both L_1 and L_2 penalties:

$$\min_{\mathbf{w}} \|X\mathbf{w} - \mathbf{r}\|_2^2 + \lambda (\alpha \|\mathbf{w}\|_1 + (1 - \alpha) \|\mathbf{w}\|_2^2). \quad (3.5)$$

The α parameter controls the balance between LASSO and Ridge penalties, where $0 \leq \alpha \leq 1$.

3.1.2 Ensemble and Tree-Based Methods

Even though linear models and their regularised extensions work well in high-dimensional settings, they can only capture linear relationships and assume that the target variable is a linear combination of the predictors. In most financial applications, relationships between variables are complex, predictors interact in ways linear models cannot fully capture, making the linear assumption overly restrictive. Decision trees offer one of the simplest ways to move beyond the linear structure. Regression trees [38] allow interactions among predictors to emerge naturally from the data. Instead of fitting a single global equation, regression trees predict the target variable by learning decision rules from the data. The tree identifies groups of observations that behave similarly and then splits the data step by step, choosing at each stage the predictor that best separates the remaining observations. The final prediction for a new observation is the average of the training observations of the target value that fall into the same leaf. In standard regression trees, each node is split based on the predictor and the threshold that minimises the total mean squared error across the two resulting child nodes:

$$\text{MSE}_{\text{split}} = \frac{1}{n_L} \sum_{i \in L} (r_i - \bar{r}_L)^2 + \frac{1}{n_R} \sum_{i \in R} (r_i - \bar{r}_R)^2, \quad (3.6)$$

where $\text{MSE}_{\text{split}}$ represents the total mean squared error resulting from a candidate split, computed

as the sum of the mean squared errors in the left (L) and right (R) child nodes.

Splitting continues until the tree reaches its predefined maximum depth, or until an additional split no longer produces a meaningful reduction in error. Splitting also stops when a node becomes too small to divide any further.

While regression trees can learn nonlinear patterns and identify dependencies among predictors, a single tree is often unreliable. Regression trees operate through a hierarchical, top-down set of binary splits. So, even a small perturbation in the data might alter an early split, changing the entire downstream structure of the tree. Moreover, each split is based on a strict cut-off; observations near a split point can switch sides with minimal noise, changing the composition of subsequent nodes. When combined with the fact that leaves often contain only a limited number of observations, particularly in deeper trees, the result is a high-variance model with unstable predictions.

Random Forest addresses these issues by fitting an ensemble of decision trees rather than relying on a single tree. The algorithm draws bootstrap samples from the training data, fits a separate decision tree to each sample, and obtains the final prediction by averaging the outputs of all individual trees [39]. Formally, the Random Forest prediction is given by

$$\hat{r}_{t+1} = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}_t), \quad (3.7)$$

where $T_b(\mathbf{x}_t)$ is the prediction of the b -th tree and B is the total number of trees in the ensemble. Random Forests also introduce another layer of randomness beyond bagging through random feature selection. At each split, the algorithm draws a random subset of predictors, which reduces correlation among trees and prevents a single dominant feature from shaping every tree in the ensemble.

Gradient Boosting, or Gradient Boosted Regression Trees (GBRT) [40], is another ensemble technique that extends decision trees through boosting rather than bagging. The theory behind boosting is to first fit a weak, shallow tree, then iteratively compute the negative gradient of a chosen loss function and fit a new tree to these pseudo-residuals. The core idea is to repeatedly correct the errors of the previous ensemble to arrive at an optimal model that captures the complex relationships in the data. Mathematically, Gradient Boosting seeks to solve an optimisation

problem that minimises the expected loss of a target variable by approximating the solution as an additive model of weak learners. The optimisation problem can be expressed as

$$\hat{F} = \arg \min_F E[L(r, F(\mathbf{x}))]. \quad (3.8)$$

Rather than optimising this problem directly, Gradient Boosting approximates the solution using an additive model of weak learners:

$$\hat{F}(\mathbf{x}) = F_0(\mathbf{x}) + \sum_{m=1}^M \nu \gamma_m h_m(\mathbf{x}), \quad (3.9)$$

where $F_0(\mathbf{x})$ represents the starting prediction, and each $h_m(\mathbf{x})$ is a weak learner (typically a shallow regression tree) whose contribution is scaled by $\nu \gamma_m$.

To determine the contribution of the next learner $h_m(\mathbf{x})$, Gradient Boosting fits this tree to the pseudo-residuals produced by the current model. The pseudo-residuals indicate the direction in which the model should adjust its predictions and are defined as

$$e_i^{(m)} = - \left. \frac{\partial L(r_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right|_{F=F_{m-1}}. \quad (3.10)$$

At each iteration, the new model F_m is obtained by updating the previous model F_{m-1} in the direction that most reduces the loss. This update takes the form

$$F_m(\mathbf{x}_t) = F_{m-1}(\mathbf{x}_t) + \nu \gamma_m h_m(\mathbf{x}_t), \quad (3.11)$$

where $\nu \in (0, 1)$ is the learning rate controlling the shrinkage of each tree's contribution and γ_m is the step size. In practice, a new regression tree $h_m(\mathbf{x})$ is trained to approximate these pseudo-residuals, and the resulting tree becomes the next additive component of the model. These steps are repeated until either no additional trees provide meaningful value or a prespecified number of trees is reached. The final model is therefore the sum of many small corrective steps contributed by shallow regression trees trained on the pseudo-residuals of the current ensemble.

3.1.3 Support Vector Regression

Unlike the ensemble methods discussed above, which build predictive models by aggregating multiple decision trees, Support Vector Regression (SVR) [41] originates from a fundamentally different framework. Rather than constructing an ensemble, SVR seeks a single function that fits the observed returns within a specified tolerance band ε , penalising only predictions that deviate beyond this threshold. This makes it robust to small perturbations in the training data and less sensitive to outliers than least-squares approaches, which is particularly relevant in financial return series where extreme observations are common. The primal optimisation problem is

$$\min_{\mathbf{w}, b, \xi, \xi^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{t=1}^T (\xi_t + \xi_t^*), \quad (3.12)$$

$$\text{subject to } r_{t+1} - \mathbf{w}^\top \phi(\mathbf{x}_t) - b \leq \varepsilon + \xi_t, \quad (3.13)$$

$$\mathbf{w}^\top \phi(\mathbf{x}_t) + b - r_{t+1} \leq \varepsilon + \xi_t^*, \quad (3.14)$$

where $\mathbf{w}$ is the weight vector, C controls the penalty on deviations beyond the ε -tube, b is the bias term, and $\xi_t, \xi_t^* \geq 0$ are slack variables. A radial basis function kernel is employed to capture nonlinear relationships between predictors and returns without requiring an explicit mapping into the high-dimensional feature space.

3.2 Nested Clustered Optimisation

The use of clustering methods in finance traces back to Mantegna [42], who showed that the correlation structure of financial assets is not arbitrary but reflects a genuine hierarchical organisation. By converting the correlation matrix into a distance metric and applying hierarchical clustering, assets were found to group naturally into clusters that broadly correspond to economic sectors and shared risk exposures. This was an important observation; it established that the correlation matrix carries useful structure beyond its role as an input to mean–variance optimisation. The idea that this structure could be exploited for portfolio construction has

since motivated a range of clustering-based allocation approaches, differing in how clusters are formed and how capital is allocated across them. López de Prado [11] formalised this through the HRP algorithm, which uses hierarchical clustering to allocate risk across assets without inverting the covariance matrix, as discussed in Chapter 2. The NCO algorithm, introduced by López de Prado [12], takes this a step further by embedding clustering within a mean–variance framework, enabling the incorporation of return forecasts while retaining the stability benefits of the two-stage decomposition.

The instability of mean–variance optimisation in high-dimensional settings arises primarily from the structure of the covariance matrix. When N/T is close to one, $\hat{\Sigma}$ becomes poorly conditioned: eigenvalues absorb estimation noise, and even small changes in the input data can cause large swings in the optimal portfolio weights. Even after applying the denoising and shrinkage techniques discussed in Chapter 2, a further source of instability remains. In standard mean–variance optimisation, the optimiser treats the entire asset universe as a single high-dimensional problem and allocates simultaneously across all N assets, meaning that noise in any pairwise covariance estimate can propagate freely across the entire portfolio, amplifying estimation errors rather than containing them.

López De Prado [12] introduced the NCO algorithm to address this structural weakness directly. The core idea is to reduce the effective dimensionality of the optimisation problem by first partitioning the asset universe into clusters of internally correlated assets, and then solving two smaller, better-conditioned optimisation problems: one *within* each cluster to determine intra-cluster weights, and one *across* clusters to determine how capital is allocated between them. The final portfolio weights are the product of these two allocation layers. By restricting the optimiser to work within clusters rather than across the full asset universe, the NCO framework limits the propagation of noise and produces portfolios that are considerably more stable than those obtained from a single unrestricted optimisation.

Mathematical Framework

Let $\mathbf{r}_t \in \mathbb{R}^{N \times 1}$ denote the vector of asset returns at time t , with expected return vector $\boldsymbol{\mu} \in \mathbb{R}^{N \times 1}$ and covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$. Standard mean–variance optimisation seeks portfolio weights

$\mathbf{w} \in R^{N \times 1}$ solving

$$\begin{aligned} & \text{Maximise} \quad \boldsymbol{\mu}^T \mathbf{w} - \frac{\gamma}{2} \mathbf{w}^T \Sigma \mathbf{w} \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1, \quad w_i \geq 0 \quad \text{for } i = 1, \dots, N, \end{aligned} \quad (3.15)$$

where $\gamma > 0$ is the risk aversion parameter.

The NCO algorithm replaces this single optimisation with a structured two-stage procedure. Suppose the clustering step partitions the N assets into K clusters $C_1, C_2, \dots, C_K$, where each cluster C_k contains n_k assets and $\sum_{k=1}^K n_k = N$.

Stage 1 — Intra-cluster optimisation . For each cluster C_k , let $\boldsymbol{\mu}^{(k)} \in R^{n_k \times 1}$ and $\Sigma^{(k)} \in R^{n_k \times n_k}$ denote the sub-vector of expected returns and sub-matrix of the covariance matrix restricted to the assets in C_k . The intra-cluster weights $\mathbf{w}^{(k)} \in R^{n_k \times 1}$ are obtained by solving

$$\begin{aligned} & \text{Maximise} \quad \left(\boldsymbol{\mu}^{(k)} \right)^T \mathbf{w}^{(k)} - \frac{\gamma}{2} \left(\mathbf{w}^{(k)} \right)^T \Sigma^{(k)} \mathbf{w}^{(k)} \\ & \text{subject to} \quad \sum_{i \in C_k} w_i^{(k)} = 1, \quad w_i^{(k)} \geq 0 \quad \text{for all } i \in C_k. \end{aligned} \quad (3.16)$$

Stage 2 — Inter-cluster optimisation . Once intra-cluster weights $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(K)}$ are obtained, each cluster is treated as a single synthetic asset. The expected return of cluster k is

$$\tilde{\mu}_k = \left(\mathbf{w}^{(k)} \right)^T \boldsymbol{\mu}^{(k)}, \quad (3.17)$$

and the covariance between clusters k and l is

$$\tilde{\Sigma}_{kl} = \left(\mathbf{w}^{(k)} \right)^T \Sigma^{(k,l)} \mathbf{w}^{(l)}, \quad (3.18)$$

where $\Sigma^{(k,l)}$ is the $n_k \times n_l$ sub-matrix of Σ with rows indexed by the assets in C_k and columns indexed by the assets in C_l . The inter-cluster weights $\tilde{\mathbf{w}}$ are then obtained by solving

$$\text{Maximise} \quad \tilde{\boldsymbol{\mu}}^T \tilde{\mathbf{w}} - \frac{\gamma}{2} \tilde{\mathbf{w}}^T \tilde{\Sigma} \tilde{\mathbf{w}} \quad (3.19)$$

$$\text{subject to } \sum_{k=1}^K \tilde{w}_k = 1, \quad \tilde{w}_k \geq 0 \quad \text{for } k = 1, \dots, K.$$

Final weights. The weight assigned to asset i in cluster C_k is

$$w_i = \tilde{w}_k \cdot w_i^{(k)}, \quad (3.20)$$

where $\tilde{w}_k$ is the inter-cluster weight for cluster k and $w_i^{(k)}$ is the intra-cluster weight for asset i within that cluster.

The key structural advantage of this decomposition is that the inter-cluster covariance matrix $\tilde{\Sigma}$ has dimension $K \times K$ with $K \ll N$, and its off-diagonal elements are substantially smaller than those of the full covariance matrix Σ . This follows because clustering groups assets with high mutual correlation into the same cluster, so the cross-cluster covariances, which become the off-diagonal elements of $\tilde{\Sigma}$, tend to be small. The optimiser therefore operates on a matrix that is closer to diagonal, and the condition number of $\tilde{\Sigma}$ is much lower than that of Σ . In this sense, the NCO algorithm does not simply reduce dimensionality; it actively transforms an ill-conditioned estimation problem into a more tractable one.

3.2.1 Clustering Methods within the NCO Framework

A central component of the NCO algorithm is the clustering step, which determines how the asset universe is partitioned before the two-stage optimisation is carried out. Clustering algorithms generally fall into two broad classes. Partitional methods assign each asset to exactly one cluster, producing a flat, non-nested partition of the asset universe. K-means is the most widely used example of this class. Hierarchical methods take a different approach; rather than producing a single partition, they generate a sequence of nested groupings represented as a tree, with a single all-inclusive cluster at the top and individual assets at the bottom. The tree can be built either bottom-up, by successively merging the closest pairs of clusters, or top-down, by recursively splitting them. One practical advantage of hierarchical clustering is that any flat partition can be recovered from it simply by cutting the tree at a chosen level, whereas deriving a hierarchical structure from a flat partition is not generally possible. The original NCO formulation of López de Prado [12] employs a modified k-means algorithm to partition the asset universe. This study

also considers hierarchical clustering as an alternative, given that it does not require the number of clusters to be fixed in advance and naturally reflects the dependency structure among assets. The two approaches are implemented and compared within the NCO framework across all portfolio objectives considered in this chapter.

Optimal Number of Clusters

In this chapter, the first clustering method employed within the NCO framework is the Optimal Number of Clusters (ONC) algorithm, introduced by López de Prado and Lewis [43]. ONC builds directly on k-means but modifies it in two important ways to make it suitable for financial applications. The first modification introduces an objective function that enables the algorithm to determine the optimal number of clusters K from the data, rather than requiring it to be fixed in advance. The second modification addresses the sensitivity of k-means to random initialisation by repeating the clustering across multiple initialisations and selecting the best result. Together, these two modifications form the base clustering stage of ONC. A subsequent higher-level refinement step then re-examines clusters of below-average quality and re-clusters them where doing so improves the overall partition.

Before applying k-means, ONC constructs an observations matrix from the pairwise asset correlations. Each entry of this matrix is defined as

$$X_{i,j} = \sqrt{\frac{1}{2} (1 - \rho_{i,j})}, \quad (3.21)$$

where $\rho_{i,j}$ is the Pearson correlation between assets i and j .

This transformation maps correlations into a proper distance metric satisfying non-negativity, symmetry, and the triangle inequality. A distance of zero corresponds to perfect positive correlation and a distance of one to perfect negative correlation. Clustering on this distance matrix rather than on the raw correlation matrix makes the algorithm more robust to outliers and nonlinear co-movement patterns among assets.

Base Clustering. The first modification ONC makes to standard k-means is to introduce an objective function that identifies the optimal K . For this purpose, the algorithm adopts the silhouette score introduced by Rousseeuw [44]. For a given asset i and a given partition, the

silhouette coefficient is defined as

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}; \quad i = 1, \dots, N, \quad (3.22)$$

where a_i is the average distance between asset i and all other members of its own cluster, and b_i is the average distance from asset i to all assets in the nearest cluster of which it is not a member. A coefficient of one indicates that the asset is well-placed within its cluster and clearly separated from others, while a coefficient of minus one signals a poor assignment. Rather than simply averaging the silhouette scores, ONC evaluates partition quality through a quality score q , defined as

$$q = \frac{E[\{S_i\}]}{\sqrt{V[\{S_i\}]}} \quad (3.23)$$

where $E[\{S_i\}]$ and $V[\{S_i\}]$ denote the mean and variance of the silhouette scores across all N assets. This formulation rewards high average quality while penalising inconsistency across assets, ensuring that the selected partition is coherent throughout rather than driven by a few well-separated clusters.

The second modification addresses the initialisation problem. The base clustering stage operates through a double loop. In the first loop, the algorithm tries different values of $k = 2, \dots, N$, applying k-means for one given initialisation at each value and evaluating the quality q of the resulting partition. The second loop repeats the first loop multiple times with different random initialisations, so that a range of starting points is explored at each k . Across both loops, the algorithm retains the clustering that achieves the highest q , thereby jointly selecting the optimal number of clusters and the best initialisation in a single coherent procedure.

Higher-level clustering The third modification to k-means concerns the uneven quality that can emerge across clusters after the base stage. Even when the overall partition achieves a high q , individual clusters may vary considerably in how well-defined they are. Some might be tightly grouped, whereas others might be loosely assembled. To identify and correct this, the algorithm computes a per-cluster quality score q_k for each cluster $k = 1, \dots, K$ and derives the average $\bar{q}$ across all clusters. Any cluster falling below this average is identified for re-clustering, and the

count of such clusters is denoted $K_1 < K$.

When $K_1 \leq 1$, the base partition is kept unchanged, as a single poorly-performing cluster does not provide enough structure to re-cluster meaningfully. When $K_1 \geq 2$, the assets within those flagged clusters are isolated, a reduced observations matrix is built from that subset alone, and the base clustering algorithm is re-applied to it. The quality of this new sub-partition is then compared against the original. If it represents an improvement, the revised clustering replaces the flagged clusters in the final partition; if not, the original groupings are preserved. The result is a clustering that is not only globally coherent but locally refined where the base stage fell short.

Hierarchical clustering

Hierarchical clustering builds a nested sequence of partitions among assets, represented as a tree structure known as a dendrogram. The tree can be grown in two directions: agglomeratively by starting from individual assets and merging upward, or divisively, starting from a single group and splitting downward. Agglomerative approaches are by far the most common choice in financial applications, given their computational tractability and ease of interpretation [42, 45, 11] and this study follows that convention. Beginning with each asset as its own singleton cluster, the algorithm performs $N-1$ successive merges, at each step combining the pair of clusters deemed closest according to both the chosen distance metric and the linkage criterion, until all assets are joined into a single group.

The linkage criterion determines which pair of clusters is merged at each step of the algorithm, and its choice directly affects the structure of the resulting partition. Among the linkage methods commonly used in hierarchical clustering are single linkage, complete linkage, average linkage, and Ward's linkage. Single linkage identifies the two clusters whose nearest individual elements are closest and merges them, regardless of how far apart the rest of their members may be. While simple in construction, this approach is prone to a chaining effect, where assets that are only loosely related end up grouped together because a single pair of close elements triggered the merge. Complete linkage takes the opposite approach, basing the merge decision on the most distant pair of elements across two clusters, which tends to produce tighter groupings but can be

sensitive to outliers. Average linkage computes the mean distance across all pairs of elements between two clusters, one from each, and merges the pair of clusters with the smallest mean distance, offering a more stable measure of inter-cluster proximity that sits between the two extremes. Ward’s criterion [46], on the other hand, relies on variance minimisation. At each step, rather than measuring distances between individual elements, it identifies the merge that produces the smallest increase in total within-cluster variance across the entire partition and proceeds accordingly. In this study, Ward linkage is employed within the NCO framework, as it produced the most stable and coherent cluster partitions across the rolling estimation windows considered.

Once the dendrogram is constructed, the optimal number of clusters is recovered by cutting the tree at a chosen height, where the number of intersecting branches corresponds to the number of clusters formed. The choice of where to cut directly determines the resulting partition and is therefore non-trivial. Fixing k arbitrarily risks imposing a structure that does not reflect the genuine cluster organization present in the data, with no principled basis for preferring one value over another.

To avoid imposing an arbitrary cluster count, the optimal number of clusters is estimated from the data using the Two-Order Difference to Gap Statistic of Yue et al. [47]. The method compares the observed within-cluster dispersion across candidate values of k against the expected dispersion under a null reference distribution. The optimal k is identified at the point at which additional clusters no longer yield a meaningful reduction in dispersion. This provides a stable, data-driven estimate of the cluster structure in the asset universe.

3.3 Data

In our study, we use monthly data of the S&P 500, covering the period from January 1990 to December 2024. The dataset includes information from multiple sources available through WRDS. We include stock-level data from the Center for Research in Security Prices (CRSP), firm-level fundamentals from Compustat, and macroeconomic variables sourced from the FRED-MD database, compiled by McCracken and Ng [48]. In addition, we compute monthly firm-specific risk and performance measures. We use a total of 214 predictors, all aligned to the end-of-month

date and lagged by one month to ensure proper out-of-sample forecasts.

The target variable in our study is the Monthly Holding Period Return ($RET_{i,t}$) obtained from CRSP. It represents the change in the total value of an investment in a common stock over some period of time per dollar of initial investment, and can be defined as follows:

$$r_t = \left(\frac{p_t f_t + d_t}{p_{t'}} \right) - 1, \quad (3.24)$$

where p_t is the last sale price or closing bid/ask average at time t , $p_{t'}$ is the last available price prior to t , f_t is the price adjustment factor that accounts for stock splits and similar events, and d_t is the cash adjustment, which includes dividends.

3.3.1 Technical Indicators

We use fourteen technical indicators derived from monthly price and volume data, following the methodology developed by Neely et al. [49]. These indicators are designed to represent three main categories of trend-following strategies: moving averages, momentum, and volume-based signals, each aiming to identify recurring market patterns that may help predict future price and return behaviour. The moving-average rules span both short- and long-term horizons, including MA(1,9), MA(1,12), MA(2,9), MA(2,12), MA(3,9), and MA(3,12). Momentum indicators include MOM(9) and MOM(12), while the volume-based measures consist of VOL(1,9), VOL(1,12), VOL(2,9), VOL(2,12), VOL(3,9), and VOL(3,12). (see Table 5.1 in the Appendix for full variable definitions).

3.3.2 Fundamental and Macroeconomic Variables

We obtain the firm-level fundamental variables from the Compustat database via WRDS. We include sixty-nine monthly financial ratios that capture the most critical aspects of a company's performance, including valuation, profitability, capitalization, financial soundness, solvency, liquidity, and efficiency ratios. More details can be found in the appendix.

We also include a comprehensive set of 126 macroeconomic variables, sourced from the FRED-MD database, which provides a standardised set of monthly U.S. macroeconomic factors.

These variables are widely used in empirical asset pricing and forecasting research [48]. The macroeconomic variables represent systematic, broad economic conditions that directly affect the prediction of stock returns. The variables are grouped into the following main categories: output and income, labor market, consumption, orders and inventories, money and credit, interest rate and exchange rates, prices, and stock market. Finally, each variable was transformed according to the transformation recommended by FRED-MD [48].

In addition to the technical, fundamental and macroeconomic variables, we include a set of firm-specific risk and performance measures, including the Sharpe ratio, Sortino ratio, Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), and Calmar ratio. Each of these variables is computed on a rolling 12-month window basis to capture the evolving risk-adjusted and downside performance of each firm's returns over time.

3.4 Methodology

3.4.1 Forecasting Framework

The main objective of this study is to develop and evaluate a comprehensive framework for portfolio optimisation. The framework combines machine-learning-based predictions of stock returns with alternative covariance matrix estimates, all within the NCO approach for asset allocation. This framework uses forecasted returns and refined covariance matrix estimators as joint inputs into the two-step NCO algorithm to determine the optimal intra-cluster weights and portfolio level weights.

The forecasts of one-month-ahead stock returns are estimated using a fully out-of-sample rolling window design. Each forecasting model is trained on the most recent 120 monthly observations, corresponding to 10 years of data, to predict the return at time $t + 1$. The first training window spans January 1990 to December 1999, and the first forecast is produced for January 2000. After each prediction, the training window advances by one month, and the model parameters are re-estimated until the final forecast for December 2024.

The target variable is the Monthly Holding Period Return ($RET_{i,t}$), as defined in Section 3.3. The predictor set includes 214 variables, which cover a range of areas such as technical indicators,

firm-level fundamental variables, macroeconomic variables, and firm-specific measures of risk and performance. Given that each forecasting model is estimated using a rolling window of 120 monthly observations, the number of predictors is large relative to the number of observations available for estimation, raising a clear risk of overfitting. A model with too many free parameters relative to the sample size can fit the noise in the training window rather than the underlying signal, producing forecasts that perform well in-sample but fail to generalise out-of-sample. This risk is addressed in several ways throughout the forecasting framework. LASSO and Elastic Net directly penalise model complexity through L1 and L2 regularisation, shrinking or eliminating coefficients on uninformative predictors. Random Forest and Gradient Boosting mitigate overfitting through ensemble methods and controls on model complexity, such as tree depth, which limit the capacity of individual trees to fit noise. Additionally, the fully out-of-sample rolling window design, where each forecast is generated using only data available prior to the forecast date and the model is re-estimated at every step, ensures that predictive performance is assessed strictly out-of-sample rather than reflecting in-sample fit.

All predictors are aligned to the month-end date and are lagged by one month to prevent look-ahead bias and ensure that forecasts at time $t+1$ are estimated using only the information available up to time t .

Each forecasting model is estimated individually for each stock and every window rather than pooling all firms into a single cross-sectional model. The firm-level approach enables the algorithm to identify unique trends and patterns for each stock over time, trends that can often be masked when stocks are analysed together. To further avoid overfitting, the parameters of each model are tailored specifically to our case and the provided dataset. These forecasts are then directly fed into the portfolio optimisation step.

3.4.2 Portfolio Optimisation Framework

This subsection outlines the method used to construct portfolios based on the forecasted returns and covariance matrix estimators. The forecasted returns from five forecasting models, namely LASSO Regression, Elastic Net Regression, SVR, Random Forest, and Gradient Boosting, along with various covariance matrix estimators, serve as joint inputs to the portfolio optimisation

process. The covariance matrices are estimated from historical data and refined using different noise reduction techniques to address the inherent instability of high-dimensional covariance matrices. The raw covariance matrix is used as a baseline estimator. To improve its stability, the covariance matrix is estimated using several methods. First, the covariance matrix is denoised using Random Matrix Theory (RMT), where the noisy eigenvalues in the spectrum are replaced in three different ways: by their mean value, by the 75th percentile, and by the 90th percentile of the eigenvalue distribution. In addition, Ledoit and Wolf's linear shrinkage and analytical nonlinear shrinkage-based estimators are employed [50, 51].

Beyond static estimators, this study also incorporates the Dynamic Conditional Correlations (DCC-GARCH) model to estimate time-varying correlations between assets. The DCC-GARCH estimators are further refined using denoising and linear shrinkage methods. Finally, we also utilise the DCC-NL model [24]. This model replaces the full likelihood in the DCC-GARCH with a composite likelihood and further refines the covariance matrix through nonlinear shrinkage.

The forecasted returns and the covariance matrix estimators are used as inputs to the portfolio optimisation process. Portfolio optimal weights are determined using the NCO algorithm. The NCO algorithm involves two optimisation levels; the first level determines the intra-cluster weights — the optimal weights within each cluster. The second level optimises the inter-cluster weights based on the covariance matrix between clusters and the weighted average of expected returns within each cluster.

The portfolio's out-of-sample returns are estimated on a rolling basis with a rolling window of 120 monthly observations. The covariance matrix used in each rebalancing period is based on the past 120 months of data, while the forecasted returns correspond to the following month, $t + 1$. After the portfolio weights are optimised, the realised returns for month $t + 1$ are used to compute the out-of-sample portfolio performance, adjusted for transaction costs of $k = 0.001$ per rebalancing period. The rolling process is repeated over 300 periods, producing 300 out-of-sample monthly portfolio net returns.

The NCO algorithm is performed under two objective functions: The Maximum Quadratic Utility and the Maximum Sharpe Ratio. Both optimisation objectives rely directly on the forecasted vector of returns (μ_t) and the covariance matrix (Σ_t) to arrive at optimal weights. In this

study, the original NCO framework, based on López de Prado's ONC algorithm and the Pearson correlation distance metric, is compared to the same NCO framework but using hierarchical clustering. Additionally, the study examines the effect of using different distance metrics on the clustering process and the overall portfolio performance. Specifically, the following distance metrics are explored: Spearman correlation, Kendall correlation, covariance distance, correlation distance, and mutual information. Finally, to evaluate the added value of the NCO algorithm, a no-clustering benchmark portfolio is constructed using the same inputs.

3.5 Results and Analysis

This section presents the key findings of the study. We begin by evaluating the out-of-sample forecasting performance of the machine-learning models and assessing the statistical significance of the results using the Probabilistic Sharpe Ratio. We then examine the incremental contribution of the NCO clustering structure by comparing clustered portfolios against a non-clustered benchmark using identical inputs. The role of distance metrics in shaping clustering outcomes and portfolio performance is assessed next, followed by a full evaluation of portfolio performance across covariance estimators under both the Maximum Quadratic Utility and Maximum Sharpe Ratio objectives. The section concludes with a portfolio weight and diversification analysis.

3.5.1 Forecasting Performance of Machine Learning Models

This section evaluates the economic performance of the machine learning forecasts by forming equally weighted decile portfolios sorted based on predicted stock returns. At the end of each month, from January 2000 to December 2024, stocks are sorted into ten deciles based on their predicted end-of-month returns. Decile 1 (Low) contains the stocks with the lowest forecasted returns, whereas Decile 10 (High) includes those with the highest. Each portfolio is equally weighted and rebalanced monthly, allowing its composition to evolve dynamically in response to updated model forecasts. The evaluation also includes an H-L (High minus Low) portfolio, which is formed by taking a long position in the high decile and a short position in the low decile, creating a zero-cost investment strategy that captures the spread between predicted winners and

losers, following the methodology in Gu et al. [52].

Table 3.1 summarises the performance of equally weighted decile portfolios. At the end of each month, stocks are ranked and sorted into ten deciles based on their predicted returns from five forecasting models. The table reports the average results over all months: “Pred” shows the mean predicted return for each decile, “Avg” shows the mean realised monthly return, “SD” is the standard deviation of those monthly returns, and “SR” is the corresponding annualised Sharpe ratio. The key diagnostic is whether Avg rises consistently with Pred: if it does, the model is genuinely ranking stocks rather than fitting noise.

The realised average returns (Avg) generally rise with predicted returns for most models; the increase is consistent and strong for LASSO Regression, and Random Forest, with the increase being monotonic, except for a slight dip between deciles 7 and 8 observed under Random Forest. This confirms genuine cross-sectional ranking ability: these models are correctly ordering stocks, not fitting noise.

The annualised Sharpe ratios (SR) in Table [3.1] are positive for the H–L strategy across all forecasting models, indicating that each model successfully identified profitable return spreads. LASSO Regression delivers the strongest performance, with an annualised Sharpe ratio of 1.60, an annualised mean return of 24.12%, and an annualised standard deviation of 15.07%. Its L_1 penalty enforces sparsity across the 214 predictors, concentrating the signal in the most persistently informative variables and producing cleaner separation between predicted winners and losers.

Random Forest, and Gradient Boosting also perform well, each producing H–L Sharpe ratios above 1.1. Notably, Gradient Boosting achieves one of the highest H–L returns and Sharpe ratios despite some irregularities in the monotonic ordering of middle deciles. This suggests that its predictive strength is concentrated in the extreme deciles rather than the full cross-section.

On the other hand, SVR and Elastic Net exhibit much lower Sharpe ratios, as the realised Avg spread between the lowest and highest decile is substantially narrower and less consistent than that of the stronger models, indicating that these models fail to reliably separate high-return stocks from low-return stocks across the cross-section.

Finally, the average standard deviation of monthly returns (SD) is relatively stable across all

Table 3.1: Portfolio performance by decile under five forecasting models.

(a) Random Forest					(b) Gradient Boosting				
Decile	Pred	Avg	SD	SR	Decile	Pred	Avg	SD	SR
Low (1)	-3.08	0.61	6.85	0.31	Low (1)	-5.82	0.71	6.47	0.38
2	-0.92	0.67	5.39	0.43	2	-2.37	0.81	5.24	0.54
3	-0.10	0.79	4.96	0.55	3	-1.00	0.78	4.78	0.56
4	0.51	0.84	4.60	0.63	4	0.00	1.00	4.69	0.73
5	1.05	1.10	4.47	0.85	5	0.87	1.21	4.66	0.90
6	1.57	1.18	4.67	0.88	6	1.71	1.14	4.82	0.82
7	2.12	1.48	4.62	1.11	7	2.60	1.07	4.48	0.83
8	2.79	1.35	4.72	0.99	8	3.65	1.32	4.75	0.96
9	3.74	1.59	5.00	1.11	9	5.14	1.57	4.87	1.12
High (10)	6.68	1.94	5.59	1.20	High (10)	9.51	1.93	5.83	1.15
H-L	9.76	1.33	4.10	1.12	H-L	15.33	1.22	3.68	1.15
(c) LASSO Regression					(d) Support Vector Regression				
Decile	Pred	Avg	SD	SR	Decile	Pred	Avg	SD	SR
Low (1)	-5.40	0.32	6.58	0.17	Low (1)	-2.61	0.99	7.31	0.47
2	-1.23	0.36	5.57	0.22	2	-0.88	1.16	5.32	0.75
3	-0.21	0.68	4.95	0.48	3	-0.20	0.97	4.86	0.69
4	0.46	0.95	4.82	0.69	4	0.32	1.07	4.84	0.77
5	1.01	1.08	4.60	0.81	5	0.79	1.11	4.67	0.83
6	1.53	1.14	4.54	0.87	6	1.25	1.18	4.44	0.92
7	2.08	1.37	4.53	1.05	7	1.74	1.14	4.51	0.88
8	2.76	1.57	4.88	1.11	8	2.32	1.09	4.72	0.80
9	3.86	1.76	5.04	1.21	9	3.16	1.23	4.87	0.87
High (10)	8.28	2.33	6.20	1.30	High (10)	5.55	1.61	5.69	0.98
H-L	13.68	2.01	4.35	1.60	H-L	8.16	0.62	5.26	0.41
(e) Elastic Net									
Decile	Pred	Avg	SD	SR					
Low (1)	-0.12	1.10	7.17	0.53					
2	0.67	1.07	5.07	0.73					
3	0.88	1.04	4.77	0.76					
4	1.05	1.03	4.51	0.79					
5	1.21	1.08	4.58	0.82					
6	1.37	1.16	4.46	0.90					
7	1.55	1.11	4.55	0.84					
8	1.78	1.01	4.94	0.71					
9	2.13	1.28	5.01	0.89					
High (10)	3.41	1.68	6.76	0.86					
H-L	3.52	0.59	5.75	0.35					

deciles and models, suggesting that portfolios with higher realised returns are not necessarily associated with greater volatility.

3.5.2 Statistical Validation Using the Probabilistic Sharpe Ratio

In this section, we assess the statistical significance of the performance improvements generated by our proposed framework using the Probabilistic Sharpe Ratio (PSR). We also aim to assess the incremental added value of each component of our proposed methodology. The standard Sharpe ratio provides a single estimate of risk-adjusted returns; however, the PSR extends the analysis by quantifying statistical confidence within a probabilistic framework.

Bailey and López de Prado [53] introduced the PSR to quantify the probability that a given Sharpe ratio exceeds a benchmark in a one-sided, right-tailed hypothesis-testing framework. In each case, the PSR accounts for the uncertainty in the estimate and reports the probability that the strategy outperforms the benchmark at a given significance level. One advantage of PSR is that the ratio assumes that the return series is stationary but not necessarily normally distributed, which better aligns with the nature of financial time series. Moreover, the ratio accounts for higher-order moments of returns, such as skewness, kurtosis, and finite-sample bias. It adjusts the denominator of the PSR test statistic to account for both skewness and kurtosis of the return time series as follows:

$$\hat{z}[SR^*] = \frac{(\widehat{SR} - SR^*)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4-1}{4}\widehat{SR}^2}}, \quad (3.25)$$

where:

- $\widehat{SR}$ = estimated (sample) Sharpe ratio (non-annualised)
- SR^* = benchmark ratio
- T = number of observations
- $\hat{\gamma}_3$ = skewness of returns
- $\hat{\gamma}_4$ = kurtosis of returns

The corresponding PSR, representing the probability that the true Sharpe ratio exceeds the benchmark SR^* , is given by

$$PSR(SR^*) = \Phi(\hat{z}[SR^*]),$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution.

In our analysis, the benchmark ratio SR^* is defined as the performance of a portfolio built from mean historical returns and the sample covariance matrix, evaluated separately for each optimisation objective. We first compare the historical-mean benchmark portfolio to forecast-based portfolios constructed without clustering to assess whether predictive modelling alone leads to statistically meaningful improvements in risk-adjusted performance, as measured by the Probabilistic Sharpe Ratio. We then evaluate the incremental contribution of the NCO framework by comparing clustered forecast-based portfolios, under both the ONC algorithm and hierarchical clustering structures, against the same historical-mean benchmark.

The PSR analysis is conducted under the Maximum Quadratic Utility (MQ) objective. A corresponding analysis under the Maximum Sharpe Ratio (MS) objective is not presented. This is because the MS benchmark monthly Sharpe ratio of 0.2733, substantially higher than the MQ benchmark of 0.1772, sets a target that none of the strategies statistically exceed, reflecting the inherent efficiency of an objective that directly optimises the Sharpe ratio. The out-of-sample portfolio performance under the MS objective is documented in Section 3.5.6.

Table 3.2 reports the PSR results for all strategies considered under the Maximum Quadratic Utility optimisation objective. The table presents the monthly average Sharpe ratio for each strategy, the corresponding PSR relative to the benchmark monthly Sharpe ratio of 0.1772, the associated test z-statistic, and the skewness and kurtosis computed from the time series of realised monthly portfolio returns.

The Sharpe ratios presented in Table 3.2 indicate that all forecast-based strategies, regardless of clustering application, outperform the benchmark portfolio, which has a monthly Sharpe ratio SR^* of 0.1772. Specifically, the Sharpe ratios for the competing strategies range from approximately 0.205 to 0.295. These results demonstrate that incorporating return forecasts yields economically significant improvements in risk-adjusted performance compared to the historical-mean benchmark.

Analysis of the PSR reveals a consistent pattern of incremental added value: implementing

Table 3.2: Probabilistic Sharpe Ratio results under the Maximum Quadratic Utility objective.

Model	Clustering	Sharpe	PSR (%)	Skew	Kurt	z
RF	None	0.240	81.6	0.19	6.06	0.90
RF	ONC	0.295	97.5	-0.01	4.01	1.97
RF	Hierarchical	0.268	93.7	0.03	3.62	1.53
LASSO	None	0.214	67.7	0.24	4.74	0.46
LASSO	ONC	0.271	95.0	0.55	7.32	1.64
LASSO	Hierarchical	0.296	98.7	0.99	7.84	2.22
GB	None	0.244	82.9	0.12	6.97	0.95
GB	ONC	0.266	93.4	0.11	5.53	1.50
GB	H	0.273	94.5	-0.02	4.80	1.60
EN	None	0.205	62.3	0.31	4.95	0.31
EN	ONC	0.220	76.8	0.06	4.43	0.73
EN	Hierarchical	0.266	94.4	0.56	5.93	1.59
SVR	None	0.246	84.3	0.18	4.00	1.01
SVR	ONC	0.265	93.7	0.27	3.90	1.53
SVR	Hierarchical	0.267	93.8	0.14	4.26	1.54
<i>Benchmark Sharpe SR*</i>						0.1772

the NCO algorithm with either the ONC algorithm or hierarchical clustering results in a higher PSR than the corresponding forecast-based portfolios without clustering. Under all forecasting models, portfolios created using the NCO algorithm yield higher PSR, as shown in column 4 of the table. These results emphasise the statistical robustness achieved by utilising a multi-step optimisation framework beyond that provided by forecasting alone in a Markowitz optimisation context. At the 5% significance level ($z > 1.645$), two strategies reject the null hypothesis of no outperformance: the Random Forest strategy with the ONC algorithm ($z = 1.97$, PSR = 97.5%) and the LASSO strategy with hierarchical clustering ($z = 2.22$, PSR = 98.7%). The LASSO strategy with the ONC algorithm approaches the threshold ($z = 1.64$, PSR = 95.0%). This ordering is consistent with the detailed out-of-sample results in Section 3.5, where Random Forest and LASSO emerge as the strongest performing models across covariance estimators and clustering configurations.

The reported skewness values range from -0.01 to 0.99, which is generally moderate and indicates limited asymmetry in monthly portfolio returns. However, the reported kurtosis values often exceed the Gaussian benchmark of 3, reaching values of 7 in some cases. This pattern

reflects the presence of fat-tailed returns, which is expected and common in financial monthly return series, and further justifies the use of PSR to account for non-normality and validate performance.

3.5.3 The Added Value of Clustering Over Mean–Variance Optimisation

Section 3.5.2 tested the statistical significance of the proposed framework against a naive historical-mean benchmark. The natural next question is whether the NCO clustering structure itself adds value, or whether the gains are driven entirely by the ML forecasts. To answer this, we compare classical mean–variance optimisation and NCO-based portfolios directly, both using the same ML return forecasts, under the MQ and MS objectives. Results are averaged across all 10 covariance estimators for all strategies and additionally across all 6 distance metrics for NCO portfolios, ensuring that any observed improvement is attributable to clustering alone rather than covariance or distance metric choice. Portfolio weights under the MQ objective use a risk aversion parameter of $\lambda = 3$. Figure 3.1 illustrates the hierarchical structure of the experimental design and the workflow of our methodology. The results presented here reflect aggregate averages; the full breakdown across individual covariance estimators and distance metrics is presented in Sections 3.5.5 and 3.5.6.

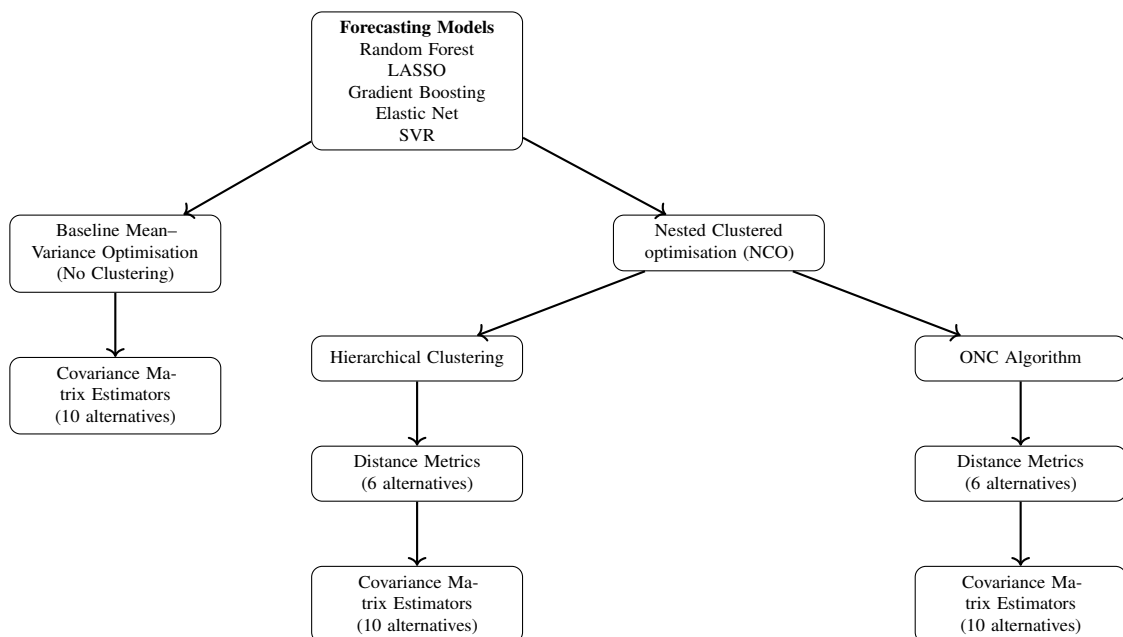


Figure 3.1: Overview of the portfolio construction and evaluation framework across forecasting models, clustering strategies, distance metrics, and covariance estimators.

Table 3.3: Average out-of-sample performance under the **Max Quadratic Utility** objective across forecasting models and clustering strategies

Model	Portfolio Strategy	Return (%)	Std (%)	Sharpe Ratio
Random Forest				
RF	No clustering	24.43	29.71	0.8226
RF	NCO – ONC	24.72	23.31	1.0619
RF	NCO – Hierarchical	23.32	24.24	0.9641
LASSO				
LASSO	No clustering	26.47	35.45	0.7471
LASSO	NCO – ONC	25.82	27.82	0.9285
LASSO	NCO – Hierarchical	29.30	28.15	1.0413
Gradient Boosting				
GB	No clustering	27.47	32.94	0.8339
GB	NCO – ONC	24.70	27.03	0.9132
GB	NCO – Hierarchical	23.60	26.16	0.9025
Support Vector Regression				
SVR	No clustering	27.06	32.23	0.8392
SVR	NCO – ONC	22.78	25.15	0.9060
SVR	NCO – Hierarchical	23.71	25.87	0.9188
Elastic Net				
EN	No clustering	21.35	31.14	0.6844
EN	NCO – ONC	20.53	25.00	0.8206
EN	NCO – Hierarchical	19.74	24.69	0.7965

Tables 3.3 and 3.4 compare the annualised out-of-sample performance of classical mean–variance portfolios against NCO-based portfolios under the MQ and MS objectives.

Results show that, at the aggregate level, the introduction of the NCO framework substantially improves portfolio performance, as measured by risk-adjusted returns. In all cases, the Sharpe ratio of the NCO portfolio, whether under the ONC algorithm or hierarchical clustering, outperforms the baseline portfolio. The magnitude of improvement varies across forecasting models and objectives. Notably, Random Forest and LASSO consistently achieve the strongest NCO-based Sharpe ratios across both objectives, a pattern that is consistent with their superior forecasting performance. Finally, the preferred clustering regime varies by forecasting model and objective, suggesting that the interaction between the signal structure of each model and the

Table 3.4: Average out-of-sample performance under the **Max Sharpe Ratio** objective across forecasting models and clustering strategies

Model	Portfolio Strategy	Return (%)	Std (%)	Sharpe Ratio
Random Forest				
RF	No clustering	14.84	14.25	1.0441
RF	NCO – ONC	15.75	14.23	1.1069
RF	NCO – Hierarchical	16.03	14.18	1.1327
LASSO				
LASSO	No clustering	17.54	19.03	0.9256
LASSO	NCO – ONC	19.59	17.69	1.1113
LASSO	NCO – Hierarchical	19.11	18.00	1.0648
Gradient Boosting				
GB	No clustering	15.18	15.79	0.9605
GB	NCO – ONC	15.99	16.19	0.9859
GB	NCO – Hierarchical	15.73	15.55	1.0115
Support Vector Regression				
SVR	No clustering	12.36	13.52	0.9138
SVR	NCO – ONC	13.84	14.24	0.9715
SVR	NCO – Hierarchical	12.92	13.62	0.9484
Elastic Net				
EN	No clustering	13.68	15.89	0.8681
EN	NCO – ONC	15.11	15.34	0.9881
EN	NCO – Hierarchical	13.23	14.73	0.9047

clustering algorithm warrants the detailed examination presented in Sections 3.5.5 and 3.5.6

3.5.4 Impact of Distance Metrics on Portfolio Performance

In this section, we study how the choice of distance metric affects NCO portfolio construction under both the ONC algorithm and hierarchical clustering. Within the NCO framework, clustering defines the structure in which optimisation takes place: assets are first grouped into clusters, and portfolio weights are then optimised both within and across these groups. For this procedure to be effective, clusters must have a meaningful economic interpretation in terms of co-movements and potential diversification.

The role of clustering is to enable risk decomposition by grouping assets into coherent risk

blocks, while diversification arises naturally through allocation across these blocks rather than by treating the optimisation problem as a single high-dimensional task. To ensure proper clustering, the choice of distance metric is therefore important, as it defines the space in which clustering occurs. We first discuss the characteristics of each distance metric and then analyse how these choices affect overall portfolio performance.

Pearson Distance

The **Pearson distance** [54] is the first and most intuitive distance measure considered in this study and is also employed in López de Prado's original NCO framework. It is a dependence-based distance derived from the Pearson correlation coefficient, which captures linear co-movement between asset returns. In practice, this means it measures how closely two assets move together and, therefore, how much they tend to contribute jointly to portfolio variance.

Let $\mathbf{r}_i = (r_{i,1}, \dots, r_{i,T})$ and $\mathbf{r}_j = (r_{j,1}, \dots, r_{j,T})$ denote the return series of assets i and j , respectively. The Pearson correlation coefficient between the two assets is given by

$$\rho_{ij} = \frac{\sum_{t=1}^T (r_{i,t} - \bar{r}_i)(r_{j,t} - \bar{r}_j)}{\sqrt{\sum_{t=1}^T (r_{i,t} - \bar{r}_i)^2} \sqrt{\sum_{t=1}^T (r_{j,t} - \bar{r}_j)^2}}, \quad (3.26)$$

where $\bar{r}_i$ and $\bar{r}_j$ denote the sample means of the return series.

To transform correlation into a distance measure suitable for clustering, Pearson distance is defined as

$$d_{ij} = \sqrt{\frac{1 - \rho_{ij}}{2}}. \quad (3.27)$$

The resulting distance metric d_{ij} converts the Pearson correlation coefficient into a dissimilarity measure that satisfies the properties of a proper distance metric for clustering purposes. The metric, therefore, is always non-negative, assigns a zero distance to perfectly correlated assets, and increases as the degree of linear dependence decreases.

Pearson distance is based on linear correlations and, therefore, represents a metric that reflects the type of co-movement that directly enters portfolio risk calculations. Portfolio risk is computed as $\mathbf{w}^\top \Sigma \mathbf{w}$, where the covariance matrix depends on correlations and individual asset volatilities. Since Pearson distance is constructed from these linear correlations, it reflects

the same dependence structure that drives how assets jointly contribute to overall portfolio risk. In that sense, the Pearson distance measure is consistent with the way risk is evaluated in the optimisation step and supports a coherent risk decomposition within the NCO clustering framework.

While the Pearson distance can provide a stable partition of the asset universe that reflects the dominant covariance structure, it still relies solely on linear interactions derived from second-moment information. As a result, it may overlook nonlinear or tail-specific dependence patterns that become more relevant during periods of market stress or regime shifts.

Spearman Distance

Spearman distance [55] is the second metric employed in this study and is derived from Spearman's rank correlation coefficient. Spearman correlation measures the strength of a monotonic relationship between asset returns. While Pearson correlation is designed to capture the linear relationships between assets based on their return magnitudes, Spearman distance instead assesses whether assets' returns move together in a consistent direction regardless of the magnitude. Spearman correlation can be defined as the Pearson correlation applied to ranked observations rather than raw return values, as follows:

$$\rho_{ij}^{(S)} = \frac{\sum_{t=1}^T (R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)}{\sqrt{\sum_{t=1}^T (R_{i,t} - \bar{R}_i)^2} \sqrt{\sum_{t=1}^T (R_{j,t} - \bar{R}_j)^2}}, \quad (3.28)$$

where $R_{i,t}$ and $R_{j,t}$ denote the ranks of returns at time t , $\bar{R}_i$ and $\bar{R}_j$ are the average ranks, and T is the number of observations.

The corresponding Spearman distance is then constructed as

$$d_{ij}^{(S)} = \sqrt{\frac{1 - \rho_{ij}^{(S)}}{2}}, \quad (3.29)$$

where $d_{ij}^{(S)}$ denotes the pairwise distance between assets i and j derived from their rank-based dependence structure.

Unlike Pearson correlation, Spearman correlation mitigates the noise introduced by differences in scale and volatility by focusing on the ranking of returns rather than their raw

magnitudes. This helps reduce the influence of large return values and extreme outliers that could otherwise distort the dependence measure. By relying on ranks, the metric can detect monotonic nonlinear relationships and identify assets that exhibit persistent directional co-movement over time. From a clustering standpoint, this leads to a more robust grouping of assets in environments characterised by skewed return distributions and heteroskedasticity, as the emphasis is placed on consistency in ordering rather than on the size of fluctuations.

Kendall Distance

In addition to Spearman distance, Kendall distance [56] is employed in this study as an alternative rank-based, non-parametric metric. Although both measures rely on ordinal data, they differ in how dependence is constructed. The Spearman distance is obtained from the Pearson correlation applied to ranked returns and therefore reflects the strength of the linear association between the rank values. By contrast, Kendall distance evaluates dependence through direct pairwise comparisons rather than rank magnitudes.

Kendall distance is derived from Kendall's tau coefficient. The construction of Kendall's tau begins by considering all possible pairs of observations within two return series. For each pair of time points, the method examines whether the relative movement of the two assets is consistent. A pair is classified as concordant if both assets move in the same direction between those two observations, and discordant if they move in opposite directions. The coefficient is then computed as the normalised difference between the number of concordant and discordant pairs relative to the total number of comparisons, expressed as

$$\tau_{ij} = \frac{C_{ij} - D_{ij}}{\binom{T}{2}}, \quad (3.30)$$

where C_{ij} denotes the number of concordant pairs between assets i and j , D_{ij} denotes the number of discordant pairs and T is the total number of observations.

The corresponding Kendall distance metric is then constructed as

$$d_{ij}^{(K)} = \sqrt{\frac{1 - \tau_{ij}}{2}}. \quad (3.31)$$

By focusing on ordinal co-movements between return series, Kendall distance offers a different way to characterise dependence. Since it is based on pairwise directional agreement rather than return magnitudes or rank differences, it is less influenced by extreme observations and the heavy-tailed behaviour often observed in financial data. The measure, therefore, emphasises whether assets move consistently in the same relative direction over time, rather than how large those movements are.

Mutual Information

Grounded in information theory [57], the mutual information distance is constructed from a mutual information estimator that quantifies the amount of information two return series share. The estimator does not assume any specific functional form; therefore, it can identify complex patterns of dependence, including linear, nonlinear, and cyclical relationships that correlation-based measures may overlook. By measuring the reduction in uncertainty about one asset's returns given knowledge of another, the approach provides a broader representation of inter-asset dependence.

The mutual information between two return series R_i and R_j is defined as

$$I(R_i, R_j) = H(R_i) + H(R_j) - H(R_i, R_j), \quad (3.32)$$

where $H(R_i)$ and $H(R_j)$ denote the marginal entropies of the return series R_i and R_j , respectively, and $H(R_i, R_j)$ denotes their joint entropy. The corresponding distance metric, referred to as the variation of information [58], is defined as

$$VI(R_i, R_j) = H(R_i, R_j) - I(R_i, R_j) = H(R_i) + H(R_j) - 2I(R_i, R_j), \quad (3.33)$$

which transforms the mutual information measure into a proper distance. To avoid scale effects and ensure comparable distances across all pairs, the measure is normalised. This normalisation bounds the metric and mitigates the influence of entropy magnitude, ensuring that the distance captures relative dependence rather than scale differences. The pairwise mutual information distance between assets i and j is therefore given by

$$d_{ij}^{(MI)} = \frac{VI(R_i, R_j)}{H(R_i, R_j)}. \quad (3.34)$$

In practice, the mutual information distance between two assets reflects how much uncertainty persists in one asset's return once the behaviour of the other asset is known. If the assets share a strong dependence structure, the remaining uncertainty decreases and so does the mutual information distance. The metric provides an alternative perspective for clustering, as it can capture broader dependence structures beyond those identified by conventional correlation measures.

Covariance Distance

Covariance-based distance can be viewed as a practical extension of the traditional angular correlation metric. Standard distance metrics often rely on the empirical covariance matrix, which is unstable and often captures estimation noise rather than true economic signals. To address this, the covariance distance framework replaces the raw sample matrix with the Gerber covariance estimator [59], which produces more stable co-movement estimates by reducing sensitivity to noise and extreme observations. By deriving the distance from a stabilised covariance estimator rather than a noisy raw matrix, the resulting correlations reflect more stable asset interdependencies. This grounding makes the distance measure far less sensitive to market volatility and outliers, which ultimately leads to more consistent and reliable clustering results in portfolio construction.

Correlation Distance

The correlation distance employed in this study is based on the distance correlation measure introduced by Székely et al. [60], which captures both linear and nonlinear dependencies between asset return series. Unlike Pearson-based measures, distance correlation is equal to zero if and only if the variables are statistically independent, making it a more general dependence metric.

In this study, a threshold is applied to the distance correlation matrix to filter weak dependencies before constructing the angular distance. This induces a sparse dependence structure that focuses on stronger and more economically meaningful relationships between assets.

Table 3.5: Rank-based comparison of distance metrics under the maximum quadratic utility objective. Lower ranks indicate superior and more consistent performance across covariance estimators.

Distance Metric	NCO + ONC Algorithm		NCO + Hierarchical Clustering	
	Avg Rank	Rank	Avg Rank	Rank
Pearson	2.30	1	3.40	3
Spearman	3.36	3	2.31	1
Kendall	3.97	5	2.98	2
Mutual Information	3.64	4	3.94	4
Covariance	2.34	2	4.16	5
Correlation	4.54	6	4.16	6

Tables 3.5 and 3.6 presents the rank-based performance of the distance metrics considered in this study across covariance matrix estimators and forecasting models. The ranking ranges from 1 to 6, where 1 indicates the best-performing distance based on the highest out-of-sample Sharpe ratio. The ranking is calculated separately for the ONC algorithm and hierarchical clustering. By ranking performance across specifications, the table reflects the relative consistency of each distance metric rather than performance driven by a single model or covariance estimator.

To understand why certain distance metrics perform differently across clustering regimes, it is important to highlight the structural differences between the two clustering methods. The ONC algorithm begins with a k-means procedure; it then extends the traditional k-means by computing silhouette scores to assess cluster quality. Clusters with below-average quality are recursively re-clustered, and the final structure is retained only if it improves the overall quality measure. In this sense, the method is adaptive, recursive, and explicitly quality-controlled. It allows for the correction of poorly formed clusters and does not rely solely on the initial partition.

In contrast, hierarchical clustering follows a fundamentally different logic. It begins by treating each asset as its own cluster and iteratively merging the closest clusters based on the selected distance measure. Each merge is irreversible; early decisions in the tree directly influence all subsequent merges and the overall tree structure. In other words, hierarchical clustering requires a stable distance metric since any distortion in pairwise distances can result in a weak overall dendrogram structure.

Looking at Table 3.5 results, under the ONC algorithm, the Pearson distance achieves the highest performance, with an average rank of 2.30, closely followed by the covariance

distance at 2.34. This result is largely explained by the internal mechanics of the ONC algorithm framework. The Pearson correlation serves as an efficient estimator of linear dependence, directly capturing dominant co-movement patterns among assets. In equity markets, linear relationships typically account for a significant share of cross-asset dynamics, as asset returns are shaped by common market and risk exposures that are often well approximated by linear models. For this reason, the Pearson distance is well-suited to capturing this fundamental structure. While the metric is known to be sensitive to outliers and heavy tails, the recursive framework of the ONC algorithm mitigates these shortcomings by identifying poorly formed clusters (characterised by low silhouette t-statistics) and re-clustering them. The re-clustering step reduces the influence of noisy pairwise distances and allows for adjustments. Consequently, the efficiency of the Pearson metric outweighs concerns about its possible instability. Moreover, Pearson distance -being derived directly from the covariance matrix- retains the magnitude of linear co-movement across assets, which is consistent with the variance-minimising nature of k-means clustering.

The covariance distance also demonstrates strong performance, ranking second overall. In this study, we employ the Gerber covariance estimator within the covariance-based distance to reduce the impact of noise and extreme observations. By stabilising the covariance matrix prior to its transformation into correlations, the covariance distance reduces noise in the eigenstructure and consequently enhances cluster stability.

Under hierarchical clustering, the ranking changes because the algorithm's structure is different. Rank-based distance metrics - Spearman and Kendall - show strong outperformance with average ranks of 2.31 and 2.98, respectively. Unlike the ONC algorithm, hierarchical clustering builds the tree sequentially and does not allow early clustering decisions to be revisited, and so the stability of the dependence measure becomes more crucial. Local distortions and noise in pairwise distances, therefore, propagate and affect the overall structure of the tree. Spearman and Kendall distance measures rely on rank values and monotonic relationships, making them less sensitive to magnitude-driven noise, which explains their strong performance under hierarchical clustering.

Table 3.6 presents the same results but under the Maximum Sharpe Ratio optimisation objective. It is important to note that under the Maximum Sharpe Ratio optimisation objective,

Table 3.6: Rank-based comparison of distance metrics under the maximum Sharpe ratio objective.

Distance Metric	NCO + ONC Algorithm		NCO + Hierarchical Clustering	
	Avg Rank	Rank	Avg Rank	Rank
Pearson	3.44	4	3.36	3
Spearman	3.22	3	3.06	2
Kendall	2.62	2	3.72	4
Mutual Information	3.48	5	4.14	5
Covariance	2.60	1	2.16	1
Correlation	5.56	6	4.36	6

the optimiser is more sensitive to estimation errors than under the Maximum Quadratic Utility. Maximum Quadratic Utility penalises variance in a linear-quadratic form. However, volatility enters the denominator of the Sharpe ratio, and so any errors in covariance estimation or distortions introduced through unstable clustering structures are magnified by the optimiser. For this reason, the covariance distance ranks first under both clustering regimes, as shown in Table 3.6. The covariance distance metric stabilises the covariance matrix before computing distances, and since the Maximum Sharpe Ratio objective values robustness, the two align naturally. It is then followed by the rank-based distances, namely Spearman and Kendall, which also perform strongly under the Sharpe objective. Since these measures rely on rank information rather than magnitude, they are inherently more stable and less sensitive to distributional distortions.

Interestingly, both mutual information and distance correlation yield weaker results in our setting, as shown in Tables 3.5 and 3.6. Although these measures are designed to capture broad forms of dependence, including nonlinear ones, their flexibility does not translate into stronger portfolio performance here. The optimisation step in our framework is explicitly variance-based and therefore grounded in second-moment risk. Pearson and covariance distances are directly derived from this structure and naturally align with the objective function. Even the rank-based distances preserve monotonic relationships that remain economically relevant for variance-driven allocation. In contrast, mutual information and distance correlation may identify broader dependence patterns whose magnitude does not directly inform variance minimisation. In this framework, structural alignment appears to matter more than maximal dependence generality.

3.5.5 Portfolio Performance under the Maximum Quadratic Utility Objective

Tables 3.7–3.12 summarise the out-of-sample MQ performance of the NCO portfolios across Random Forest, LASSO, and Gradient Boosting forecasts under both the ONC algorithm and hierarchical clustering — the three models selected based on their superior cross-sectional forecasting performance in Section 3.5.1, and to which all analysis in this section corresponds. Results for SVR and Elastic Net are provided in the Appendix. The tables’ results provide a structured evaluation of the full proposed framework, integrating machine-learning forecasts, advanced covariance estimators, and clustering-based allocation structures.

The results reveal systematic structural effects across three key dimensions: covariance refinement, clustering design, and the interaction between clustering and the choice of distance metric.

The classical mean–variance portfolio, created under the MQ objective using historical mean returns and the sample covariance matrix, serves as a benchmark. Its out-of-sample performance is:

- annualised Return: 16.43%
- annualised Volatility: 26.76%
- Sharpe ratio: 0.6138

Before the NCO clustering structure is introduced, the portfolios created using forecasted returns yielded Sharpe ratios ranging from 0.7231 to 0.8874, representing an improvement in risk-adjusted performance of approximately 17%–44% relative to the benchmark¹. The highest Sharpe ratio of 0.8874 is observed under the Random Forest forecasts and the DCC-NL estimator.

Tables 3.7–3.12 add the NCO clustering structure to the analysis; the Sharpe ratio now ranges from 0.8091 to 1.2305, indicating that in many specifications it is close to or above 1.0, corresponding to improvements of roughly **32%–101%** over the benchmark. These initial results

¹The full breakdown of no-clustering portfolio performance across all covariance estimators and forecasting models is reported in the Appendix.

suggest a layered improvement mechanism. Machine-learning forecasts alone enhance performance relative to historical mean returns, but the introduction of cluster-structured allocation further amplifies these gains. The increase in Sharpe ratios within the NCO framework indicates that performance improvements are not driven by forecasting alone, but by the interaction between forecasting and structural portfolio construction.

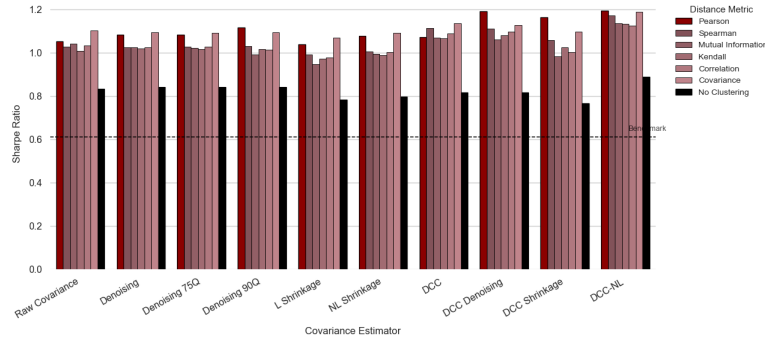
Table 3.7: Out-of-sample portfolio performance under Random Forest forecasts (ONC Algorithm) - MQ Objective — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	25.71	24.40	1.0537	Raw Covariance	23.74	23.12	1.0269
Denoising (Mean)	26.27	24.26	1.0830	Denoising (Mean)	23.63	23.08	1.0240
Denoising (75th)	26.05	24.03	1.0839	Denoising (75th)	23.72	23.09	1.0273
Denoising (90th)	26.60	23.84	1.1155	Denoising (90th)	23.91	23.20	1.0306
Shrinkage	25.29	24.33	1.0393	Shrinkage	23.61	23.83	0.9905
NL Shrinkage	26.06	24.19	1.0774	NL Shrinkage	23.60	23.50	1.0043
DCC	23.61	22.03	1.0717	DCC	23.48	21.08	1.1140
DCC Denoising	25.77	21.62	1.1920	DCC Denoising	23.42	21.11	1.1099
DCC Shrinkage	26.91	23.14	1.1631	DCC Shrinkage	23.55	22.27	1.0573
DCC-NL	27.73	23.23	1.1937	DCC-NL	25.32	21.58	1.1731
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	25.59	24.59	1.0407	Raw Covariance	23.17	22.98	1.0082
Denoising (Mean)	25.38	24.73	1.0259	Denoising (Mean)	23.24	22.82	1.0183
Denoising (75th)	25.34	24.77	1.0231	Denoising (75th)	23.22	22.82	1.0177
Denoising (90th)	24.83	25.04	0.9917	Denoising (90th)	23.26	22.90	1.0156
Shrinkage	24.42	25.82	0.9461	Shrinkage	23.16	23.79	0.9733
NL Shrinkage	25.16	25.30	0.9946	NL Shrinkage	23.17	23.46	0.9876
DCC	24.21	22.67	1.0681	DCC	22.45	21.03	1.0678
DCC Denoising	24.08	22.68	1.0620	DCC Denoising	22.82	21.14	1.0794
DCC Shrinkage	23.80	24.20	0.9835	DCC Shrinkage	22.93	22.35	1.0260
DCC-NL	26.15	23.01	1.1365	DCC-NL	24.46	21.59	1.1330
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	24.49	23.69	1.0337	Raw Covariance	26.22	23.75	1.1039
Denoising (Mean)	24.35	23.74	1.0257	Denoising (Mean)	26.05	23.83	1.0931
Denoising (75th)	24.38	23.74	1.0266	Denoising (75th)	26.03	23.83	1.0923
Denoising (90th)	24.16	23.85	1.0129	Denoising (90th)	26.15	23.91	1.0935
Shrinkage	24.25	24.80	0.9776	Shrinkage	26.34	24.62	1.0701
NL Shrinkage	24.36	24.27	1.0034	NL Shrinkage	26.27	24.09	1.0903
DCC	24.20	22.25	1.0877	DCC	25.44	22.38	1.1371
DCC Denoising	24.49	22.29	1.0985	DCC Denoising	25.36	22.48	1.1284
DCC Shrinkage	23.81	23.74	1.0028	DCC Shrinkage	25.74	23.46	1.0971
DCC-NL	25.49	22.68	1.1237	DCC-NL	26.98	22.72	1.1876

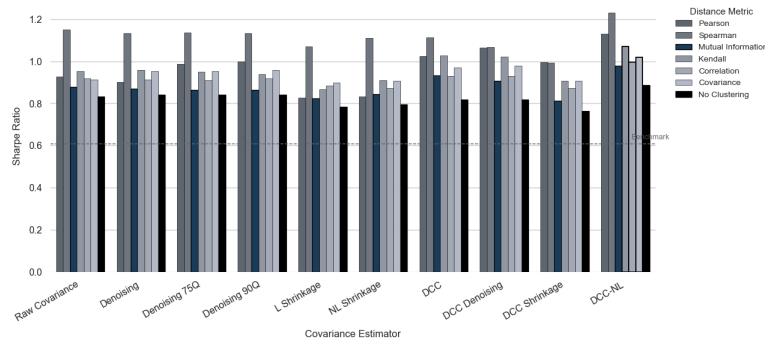
Random Forest

Within the NCO framework, Random Forest forecasts exhibit the strongest overall performance relative to the alternative forecasting models. The MQ global maximum portfolio combines Random Forest forecasts, hierarchical clustering, and the DCC-NL estimator under the Spearman

distance with a Sharpe ratio of **1.2305**. This portfolio achieves an annualised return of **28.14%** and a standard deviation of **22.87%**, representing the most efficient risk-return trade-off across all specifications, as highlighted in Table 3.8.



(a) Sharpe ratios under Random Forest forecasts — ONC Algorithm.



(b) Sharpe ratios under Random Forest forecasts — Hierarchical Clustering (Ward Linkage).

Figure 3.2: Comparison of Sharpe Ratios under Random Forest forecasts – MQ Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.

Very similar performance is observed under the ONC algorithm (Table 3.7), where DCC-NL and DCC-Denoising paired with Pearson distance achieve Sharpe ratios of 1.1937 and 1.1920 respectively. These results are consistent with the distance metric findings in Section 3.5.4: hierarchical clustering pairs best with Spearman distance, while the ONC algorithm pairs best with Pearson distance. Figure 3.2 illustrates the Sharpe ratios across all covariance estimators and distance metrics under Random Forest forecasts. In both subfigures, every NCO portfolio, regardless of distance metric or covariance estimator, clears the benchmark line (dashed) and exceeds the no-clustering portfolios (black bars), confirming that clustering adds value universally. Under the ONC algorithm, Pearson distance (dark red) consistently produces the tallest bars, while under hierarchical clustering, Spearman distance dominates. The DCC-NL column achieves the highest bars in both regimes.

Table 3.8: **Out-of-sample portfolio performance under Random Forest forecasts (Hierarchical Clustering) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	23.01	24.78	0.9288	Raw Covariance	27.27	23.72	1.1499
Denoising (Mean)	22.29	24.69	0.9031	Denoising (Mean)	26.91	23.74	1.1335
Denoising (75th)	24.47	24.81	0.9863	Denoising (75th)	26.96	23.72	1.1367
Denoising (90th)	24.35	24.35	0.9999	Denoising (90th)	26.88	23.72	1.1332
Shrinkage	21.48	25.97	0.8272	Shrinkage	26.38	24.61	1.0719
NL Shrinkage	21.47	25.79	0.8325	NL Shrinkage	26.99	24.31	1.1100
DCC	23.02	22.43	1.0260	DCC	25.06	22.52	1.1128
DCC Denoising	23.84	22.41	1.0640	DCC Denoising	25.12	23.50	1.0693
DCC Shrinkage	23.83	23.89	0.9976	DCC Shrinkage	25.37	25.54	0.9933
DCC-NL	26.15	23.10	1.1319	DCC-NL	28.14	22.87	1.2305
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	21.38	24.35	0.8778	Raw Covariance	23.21	24.33	0.9541
Denoising (Mean)	21.16	24.32	0.8700	Denoising (Mean)	23.38	24.44	0.9582
Denoising (75th)	21.13	24.42	0.8651	Denoising (75th)	23.26	24.44	0.9518
Denoising (90th)	21.63	25.05	0.8637	Denoising (90th)	23.08	24.62	0.9376
Shrinkage	21.39	25.98	0.8233	Shrinkage	22.03	25.37	0.8686
NL Shrinkage	21.25	25.14	0.8454	NL Shrinkage	22.71	24.97	0.9094
DCC	21.62	23.17	0.9327	DCC	23.76	23.10	1.0284
DCC Denoising	21.17	23.31	0.9081	DCC Denoising	23.78	23.24	1.0229
DCC Shrinkage	19.93	24.51	0.8129	DCC Shrinkage	22.12	24.37	0.9079
DCC-NL	23.04	23.52	0.9797	DCC-NL	25.20	23.50	1.0721
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.44	24.42	0.9188	Raw Covariance	22.46	24.58	0.9136
Denoising (Mean)	22.54	24.68	0.9133	Denoising (Mean)	23.54	24.67	0.9540
Denoising (75th)	22.46	24.70	0.9093	Denoising (75th)	23.56	24.68	0.9547
Denoising (90th)	22.77	24.78	0.9190	Denoising (90th)	23.88	24.89	0.9595
Shrinkage	22.59	25.53	0.8846	Shrinkage	22.96	25.51	0.9001
NL Shrinkage	21.98	25.15	0.8742	NL Shrinkage	22.74	25.08	0.9067
DCC	21.72	23.33	0.9312	DCC	22.45	23.12	0.9713
DCC Denoising	21.72	23.33	0.9312	DCC Denoising	22.80	23.29	0.9789
DCC Shrinkage	21.37	24.50	0.8726	DCC Shrinkage	22.21	24.48	0.9071
DCC-NL	23.55	23.64	0.9959	DCC-NL	24.10	23.63	1.0196

LASSO

Under LASSO forecasts, the results confirm the same structural pattern observed for Random Forest. The best specification combines hierarchical clustering, DCC-NL, and Spearman distance, achieving a Sharpe ratio of 1.1284 (Table 3.10). Under the ONC algorithm (Table 3.9), DCC-Denoising paired with Pearson distance delivers the strongest result at 1.0633.

Figure 3.3 visualises the LASSO results across all covariance estimators and distance metrics. Two things stand out immediately. First, every NCO portfolio — across all distance metrics and covariance estimators — sits above the benchmark line and clearly outperforms the no-clustering bars, confirming that clustering adds value regardless of specification choice. Second, the distance metric effect is most pronounced when paired with dynamic covariance estimators:

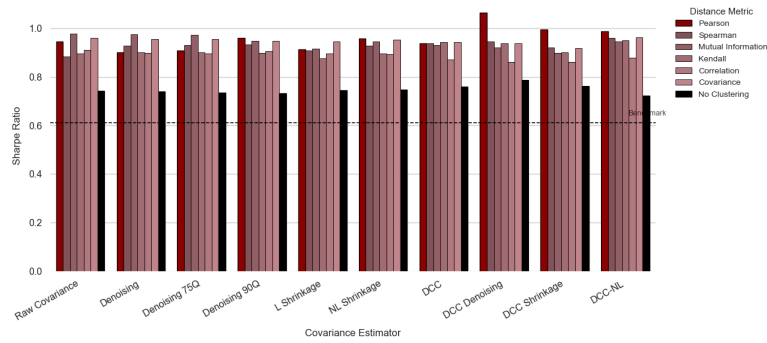
Table 3.9: **Out-of-sample portfolio performance under LASSO Regression forecasts (ONC Algorithm) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	26.93	28.53	0.9439	Raw Covariance	25.25	28.60	0.8830
Denoising (Mean)	26.64	29.57	0.9010	Denoising (Mean)	25.22	27.14	0.9290
Denoising (75th)	27.35	30.13	0.9077	Denoising (75th)	25.22	27.12	0.9301
Denoising (90th)	28.54	29.71	0.9605	Denoising (90th)	25.32	27.14	0.9327
Shrinkage	27.02	29.57	0.9137	Shrinkage	25.11	27.62	0.9089
NL Shrinkage	28.65	29.89	0.9584	NL Shrinkage	25.51	27.51	0.9273
DCC	24.65	26.28	0.9381	DCC	23.28	24.81	0.9386
DCC Denoising	28.31	26.63	1.0633	DCC Denoising	23.42	24.80	0.9444
DCC Shrinkage	28.10	28.22	0.9957	DCC Shrinkage	23.95	26.01	0.9208
DCC-NL	27.00	27.33	0.9879	DCC-NL	24.56	25.58	0.9599
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	29.15	29.85	0.9765	Raw Covariance	24.72	27.58	0.8961
Denoising (Mean)	29.03	29.80	0.9742	Denoising (Mean)	24.96	27.70	0.9010
Denoising (75th)	29.05	29.85	0.9732	Denoising (75th)	24.89	27.65	0.9000
Denoising (90th)	28.33	29.91	0.9472	Denoising (90th)	24.82	27.66	0.8972
Shrinkage	27.93	30.52	0.9152	Shrinkage	24.74	28.23	0.8763
NL Shrinkage	28.55	30.18	0.9460	NL Shrinkage	25.10	28.07	0.8944
DCC	25.92	27.86	0.9302	DCC	23.64	25.10	0.9418
DCC Denoising	25.74	27.94	0.9211	DCC Denoising	23.62	25.17	0.9384
DCC Shrinkage	26.11	29.08	0.8976	DCC Shrinkage	23.80	26.41	0.9012
DCC-NL	26.70	28.21	0.9463	DCC-NL	24.22	25.51	0.9495
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	26.34	28.94	0.9102	Raw Covariance	26.38	27.48	0.9599
Denoising (Mean)	26.08	29.04	0.8979	Denoising (Mean)	26.20	27.47	0.9538
Denoising (75th)	26.00	29.04	0.8955	Denoising (75th)	26.22	27.42	0.9562
Denoising (90th)	26.31	29.09	0.9045	Denoising (90th)	26.04	27.48	0.9478
Shrinkage	26.55	29.65	0.8953	Shrinkage	26.50	28.06	0.9441
NL Shrinkage	26.30	29.43	0.8936	NL Shrinkage	26.50	27.85	0.9513
DCC	23.61	27.15	0.8696	DCC	23.92	25.37	0.9429
DCC Denoising	23.51	27.33	0.8600	DCC Denoising	23.72	25.32	0.9366
DCC Shrinkage	24.40	28.32	0.8613	DCC Shrinkage	24.41	26.63	0.9170
DCC-NL	24.47	27.82	0.8794	DCC-NL	24.68	25.62	0.9633

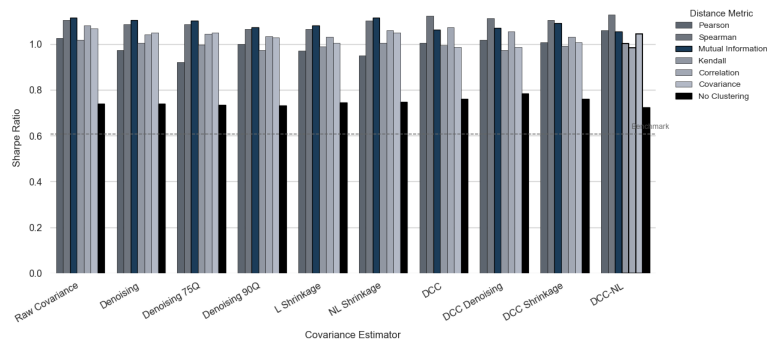
under the ONC algorithm, Pearson distance pulls ahead most clearly at DCC-Denoising, reaching the highest bar in Figure 3.3a. Under hierarchical clustering, Spearman distance outperforms other distance metrics when combined with DCC estimators, as visible in Figure 3.3b. Across static estimators, the distance metrics perform more closely together, suggesting that the choice of distance metric matters most when the covariance estimator is capturing richer dependence structure.

Gradient Boosting

Gradient Boosting follows the same broad pattern, with one notable exception. Under the ONC algorithm (Table 3.11), Denoising (90th) paired with Pearson distance achieves the highest



(a) Sharpe ratios under LASSO forecasts — ONC algorithm.



(b) Sharpe ratios under LASSO forecasts — Hierarchical Clustering (Ward Linkage)).

Figure 3.3: Comparison of Sharpe Ratios under LASSO forecasts – MQ Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.

Sharpe ratio of 1.0296, the only instance in the MQ analysis where a static estimator edges out the DCC variants. Gradient Boosting’s noisier forecasting signal appears to benefit more from eigenvalue stabilisation than from dynamic correlation modelling under the ONC algorithm. Under hierarchical clustering (Table 3.12), DCC-NL with Spearman distance recovers to an essentially equivalent SR of 1.0291. The gap between clustering regimes is negligible for Gradient Boosting (0.0005), compared to 0.037 for Random Forest and 0.065 for LASSO, suggesting that the choice of covariance estimator is the more consequential decision for this model.

Figure 3.4 confirms the dominance of Pearson distance under the ONC algorithm and Spearman distance under hierarchical clustering across all covariance estimators, consistent with the distance metric findings in Section 3.5.4 and the pattern observed for Random Forest and LASSO. As with Random Forest and LASSO, all NCO portfolios clear the benchmark line and exceed the no-clustering bars throughout.

Overall, DCC estimators dominate across all three models. In five of the six model-clustering

Table 3.10: **Out-of-sample portfolio performance under LASSO Regression forecasts (Hierarchical Clustering) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	29.72	28.93	1.0272	Raw Covariance	31.24	28.26	1.1054
Denoising (Mean)	29.44	30.22	0.9741	Denoising (Mean)	30.71	28.27	1.0862
Denoising (75th)	28.02	30.43	0.9208	Denoising (75th)	30.77	28.29	1.0877
Denoising (90th)	28.97	28.96	1.0002	Denoising (90th)	30.28	28.44	1.0649
Shrinkage	28.68	29.49	0.9724	Shrinkage	30.88	28.94	1.0672
NL Shrinkage	27.88	29.30	0.9516	NL Shrinkage	31.54	28.62	1.1021
DCC	27.23	27.08	1.0055	DCC	29.94	26.65	1.1233
DCC Denoising	27.62	27.14	1.0176	DCC Denoising	29.66	26.66	1.1123
DCC Shrinkage	28.44	28.22	1.0078	DCC Shrinkage	30.55	27.64	1.1054
DCC-NL	26.37	24.88	1.0597	DCC-NL	30.34	26.89	1.1284
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	32.58	29.20	1.1159	Raw Covariance	29.31	28.78	1.0185
Denoising (Mean)	32.59	29.45	1.1067	Denoising (Mean)	29.15	28.99	1.0054
Denoising (75th)	32.43	29.42	1.1024	Denoising (75th)	29.03	29.07	0.9987
Denoising (90th)	31.83	29.65	1.0734	Denoising (90th)	28.63	29.36	0.9752
Shrinkage	32.38	29.95	1.0814	Shrinkage	29.36	29.64	0.9904
NL Shrinkage	32.98	29.56	1.1155	NL Shrinkage	29.38	29.23	1.0054
DCC	28.91	27.21	1.0624	DCC	26.58	26.74	0.9942
DCC Denoising	29.47	27.51	1.0713	DCC Denoising	26.23	26.94	0.9735
DCC Shrinkage	31.03	28.39	1.0929	DCC Shrinkage	27.71	27.95	0.9916
DCC-NL	29.57	28.04	1.0546	DCC-NL	27.71	27.59	1.0042
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	31.51	29.15	1.0810	Raw Covariance	29.08	27.20	1.0687
Denoising (Mean)	30.41	29.17	1.0425	Denoising (Mean)	28.64	27.29	1.0495
Denoising (75th)	30.55	29.22	1.0455	Denoising (75th)	28.71	27.31	1.0511
Denoising (90th)	30.26	29.26	1.0340	Denoising (90th)	28.03	27.24	1.0289
Shrinkage	30.90	29.92	1.0327	Shrinkage	28.07	27.91	1.0059
NL Shrinkage	31.53	29.73	1.0605	NL Shrinkage	29.13	27.73	1.0506
DCC	26.75	24.92	1.0736	DCC	26.70	27.06	0.9867
DCC Denoising	26.54	25.14	1.0557	DCC Denoising	26.70	27.06	0.9867
DCC Shrinkage	26.93	26.13	1.0306	DCC Shrinkage	28.50	28.26	1.0084
DCC-NL	27.11	27.56	0.9834	DCC-NL	26.94	25.75	1.0462

combinations, the strongest specification pairs a DCC variant with the preferred distance metric for that regime. Within the DCC family, DCC-NL and DCC-Denoising consistently produce the highest Sharpe ratios. DCC-NL leads most often under hierarchical clustering, while DCC-Denoising is the stronger performer under the ONC algorithm. The distance metric preferences established in Section 3.5.4 hold throughout: Pearson distance identifies the top-performing specifications under the ONC algorithm and Spearman distance under hierarchical clustering, consistently across all three models.

Table 3.11: **Out-of-sample portfolio performance under Gradient Boosting forecasts (ONC Algorithm) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	25.32	27.28	0.9282	Raw Covariance	23.82	27.34	0.8711
Denoising (Mean)	27.51	26.95	1.0208	Denoising (Mean)	23.81	27.29	0.8725
Denoising (75th)	27.09	26.66	1.0161	Denoising (75th)	23.85	27.32	0.8730
Denoising (90th)	27.83	27.03	1.0296	Denoising (90th)	23.97	27.32	0.8775
Shrinkage	26.02	27.18	0.9571	Shrinkage	24.57	28.19	0.8713
NL Shrinkage	26.34	27.69	0.9512	NL Shrinkage	24.34	28.02	0.8686
DCC	23.04	25.74	0.8952	DCC	22.62	25.32	0.8931
DCC Denoising	23.93	26.52	0.9021	DCC Denoising	22.69	25.32	0.8961
DCC Shrinkage	24.41	26.20	0.9316	DCC Shrinkage	23.14	26.51	0.8726
DCC-NL	26.79	26.99	0.9926	DCC-NL	24.45	26.58	0.9198
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	24.34	27.72	0.8782	Raw Covariance	24.9	27.25	0.9136
Denoising (Mean)	24.45	27.63	0.8849	Denoising (Mean)	24.89	27.22	0.9144
Denoising (75th)	24.73	27.74	0.8914	Denoising (75th)	24.96	27.25	0.9162
Denoising (90th)	24.5	27.62	0.8870	Denoising (90th)	25.02	27.24	0.9183
Shrinkage	25.44	28.55	0.8911	Shrinkage	25.28	27.87	0.9072
NL Shrinkage	24.67	28.25	0.8734	NL Shrinkage	25.28	27.86	0.9072
DCC	22.83	26.63	0.8574	DCC	23.57	25.58	0.9215
DCC Denoising	23.04	26.59	0.8665	DCC Denoising	23.88	25.67	0.9304
DCC Shrinkage	24.14	27.70	0.8716	DCC Shrinkage	23.93	26.56	0.9011
DCC-NL	24.66	27.46	0.8981	DCC-NL	25.03	26.40	0.9481
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.45	26.01	0.8631	Raw Covariance	26.85	28.29	0.9491
Denoising (Mean)	22.27	26.04	0.8552	Denoising (Mean)	27.05	28.32	0.9550
Denoising (75th)	22.36	26.05	0.8584	Denoising (75th)	27.14	28.35	0.9572
Denoising (90th)	23.18	26.00	0.8916	Denoising (90th)	27.32	28.3	0.9655
Shrinkage	23.69	26.72	0.8866	Shrinkage	27.32	28.56	0.9567
NL Shrinkage	22.9	26.55	0.8626	NL Shrinkage	27.29	28.63	0.9532
DCC	22.28	25.20	0.8841	DCC	26.00	26.83	0.9690
DCC Denoising	22.14	25.22	0.8777	DCC Denoising	26.35	27.07	0.9735
DCC Shrinkage	22.78	25.93	0.8784	DCC Shrinkage	26.67	27.86	0.9574
DCC-NL	23.17	26.14	0.8862	DCC-NL	27.53	27.62	0.9967

Structural Decomposition of Risk–Return Improvements

The performance gains can be decomposed into separate structural components. Recall that the classical benchmark portfolio, constructed using historical mean returns and the sample covariance matrix without clustering, delivers an annualised return of 16.43% with a volatility of 26.76%, resulting in a Sharpe ratio of 0.6138. To understand the source of performance improvement, we will compare the benchmark risk and return to the top-performing portfolio—Random Forest and the DCC-NL estimator.

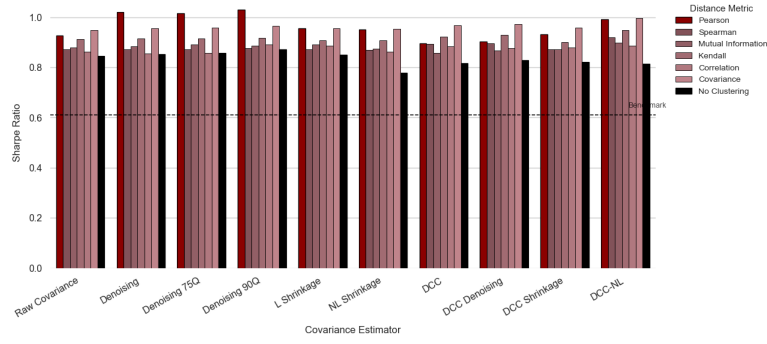
Introducing machine learning forecasts alone—specifically Random Forest combined with the DCC-NL estimator under no clustering—substantially increases the annualised return to 25.66%, but at the cost of higher volatility of around 28.91%.

Table 3.12: **Out-of-sample portfolio performance under Gradient Boosting forecasts (Hierarchical Clustering) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

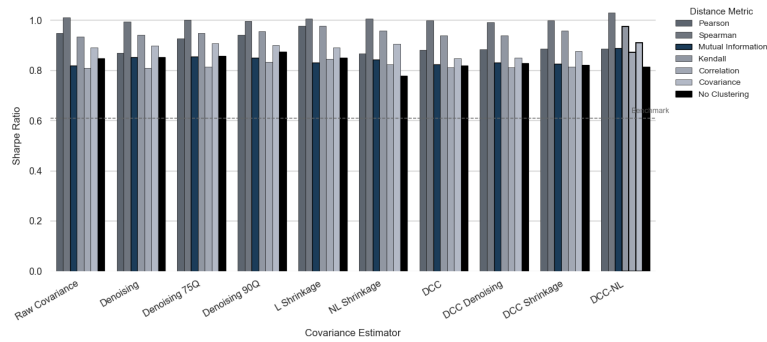
(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	24.88	26.25	0.9481	Raw Covariance	25.41	25.15	1.0103
Denoising (Mean)	22.41	25.77	0.8694	Denoising (Mean)	25.14	25.28	0.9944
Denoising (75th)	24.93	26.87	0.9278	Denoising (75th)	25.35	25.33	1.0007
Denoising (90th)	25.45	27.07	0.9401	Denoising (90th)	25.29	25.40	0.9954
Shrinkage	26.82	27.43	0.9778	Shrinkage	25.91	25.77	1.0054
NL Shrinkage	25.33	29.25	0.8658	NL Shrinkage	25.47	25.29	1.0069
DCC	22.58	25.63	0.8810	DCC	24.70	24.73	0.9987
DCC Denoising	22.78	25.77	0.8838	DCC Denoising	24.69	24.94	0.9903
DCC Shrinkage	23.43	26.46	0.8857	DCC Shrinkage	25.27	25.32	0.9981
DCC-NL	22.74	25.68	0.8858	DCC-NL	25.87	25.14	1.0291
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.52	27.52	0.8185	Raw Covariance	24.92	26.70	0.9335
Denoising (Mean)	22.27	26.16	0.8512	Denoising (Mean)	25.08	26.66	0.9408
Denoising (75th)	22.40	26.21	0.8546	Denoising (75th)	25.28	26.65	0.9485
Denoising (90th)	22.46	26.43	0.8497	Denoising (90th)	25.57	26.75	0.9561
Shrinkage	22.33	26.90	0.8300	Shrinkage	26.66	27.28	0.9770
NL Shrinkage	22.36	26.51	0.8433	NL Shrinkage	25.82	26.96	0.9577
DCC	20.61	25.01	0.8240	DCC	24.51	26.14	0.9377
DCC Denoising	20.84	25.10	0.8305	DCC Denoising	24.60	26.20	0.9388
DCC Shrinkage	21.27	25.76	0.8254	DCC Shrinkage	25.71	26.86	0.9571
DCC-NL	23.15	26.09	0.8875	DCC-NL	25.84	26.51	0.9746
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	20.88	25.81	0.8091	Raw Covariance	23.73	26.66	0.8899
Denoising (Mean)	20.86	25.76	0.8098	Denoising (Mean)	23.93	26.64	0.8979
Denoising (75th)	20.98	25.78	0.8139	Denoising (75th)	24.20	26.66	0.9080
Denoising (90th)	21.49	25.80	0.8326	Denoising (90th)	24.07	26.75	0.9000
Shrinkage	22.21	26.30	0.8443	Shrinkage	24.08	27.01	0.8914
NL Shrinkage	21.36	25.93	0.8237	NL Shrinkage	24.19	26.70	0.9060
DCC	20.53	25.30	0.8117	DCC	21.85	25.77	0.8479
DCC Denoising	20.46	25.19	0.8123	DCC Denoising	22.03	25.92	0.8499
DCC Shrinkage	21.13	25.97	0.8138	DCC Shrinkage	23.28	26.53	0.8774
DCC-NL	22.51	25.85	0.8707	DCC-NL	23.74	26.07	0.9105

This intermediate step highlights an important mechanism: forecasting enhances expected returns but does not automatically improve risk efficiency. In other words, Random Forest returns forecasts generate a meaningful predictive signal; however, the classical flat mean–variance optimisation framework remains exposed to high volatility.

The additional performance gains emerge when the NCO framework is introduced. The top-performing portfolio, Random Forest forecasts combined with hierarchical clustering and the DCC-NL estimator under the Spearman distance, achieves an annualised return of 28.14% while reducing volatility to 22.87%. The economic magnitude of these improvements is substantial: relative to the benchmark portfolio, returns increase by nearly 71%, while portfolio risk declines by approximately 14%, resulting in almost a doubling of the Sharpe ratio. This portfolio also clearly dominates the corresponding specification constructed using Random Forest forecasts and



(a) Sharpe ratios under Gradient Boosting forecasts — ONC algorithm.



(b) Sharpe ratios under Gradient Boosting forecasts — Hierarchical Clustering (Ward Linkage)).

Figure 3.4: Comparison of Sharpe Ratios under Gradient Boosting forecasts – MQ Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.

the same covariance estimator within the classical mean–variance framework, simultaneously delivering higher returns and lower risk.

This comparison shows the layered structure of the improvement. Machine learning forecasts generate stronger return signals, but these signals alone do not necessarily translate into efficient portfolios under the classical mean–variance framework. The introduction of clustering through the NCO framework improves portfolio allocation by distributing capital across clusters of related assets rather than individual securities. This structure reduces concentration and stabilises the allocation process. As a result, the predictive signal generated by the forecasting model can be translated into higher returns while simultaneously reducing portfolio volatility, thereby improving risk-adjusted performance.

3.5.6 Portfolio Performance under the Maximum Sharpe Ratio

Tables 3.13–3.18 summarise the out-of-sample MS performance of the NCO portfolios across Random Forest, LASSO, and Gradient Boosting forecasts under both clustering regimes. The MS objective maximises return relative to portfolio risk by optimising the ratio of expected return to volatility. In this formulation, risk appears in the denominator of the objective function rather than as a direct linear penalty, as in the MQ objective. This difference in structure can lead to changes in the portfolio allocations obtained under the two optimisation objectives.

The classical maximum Sharpe ratio benchmark, constructed using historical mean returns and the sample covariance matrix without clustering, delivers:

- annualised Return: 12.44%
- annualised Volatility: 13.15%
- Sharpe ratio: 0.9465

Even without forecasting or clustering, the benchmark already achieves a relatively well-balanced risk–return profile. Despite this solid starting point, the NCO framework further improves performance.

Random Forest

The strongest MS portfolio is obtained using Random Forest forecasts combined with hierarchical clustering, DCC-denoising, and the Spearman distance, delivering an annualised return of **16.35%**, volatility of **12.99%**, and a Sharpe ratio of **1.259** as shown in Table 3.14. The improvement here over the benchmark is mainly driven by the approximately 31% increase in returns, whereas volatility is already tightly controlled at the benchmark level.

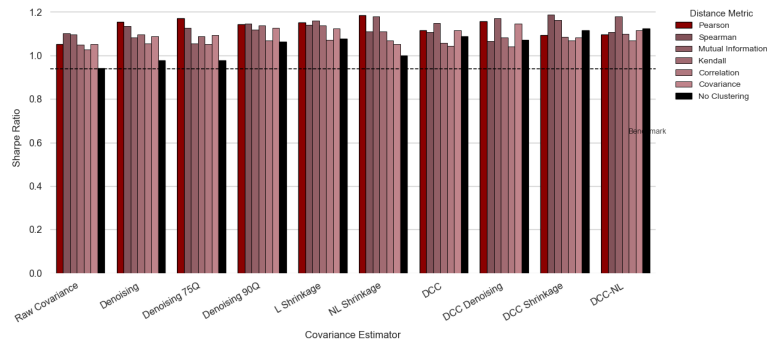
It is also informative to compare the NCO specification with the corresponding Random Forest and DCC-denoising portfolio constructed without clustering. The no-clustering version delivers a Sharpe ratio of 1.0719, with a return of 14.07% and volatility of 13.12%. Introducing the NCO structure increases return to 16.35% while leaving volatility nearly unchanged. This suggests that, under the MS objective, clustering does not primarily reduce risk further but instead

Table 3.13: **Out-of-sample portfolio performance under Random Forest (ONC Algorithm) – MS Objective** — Comparison of covariance estimators under different distance metrics

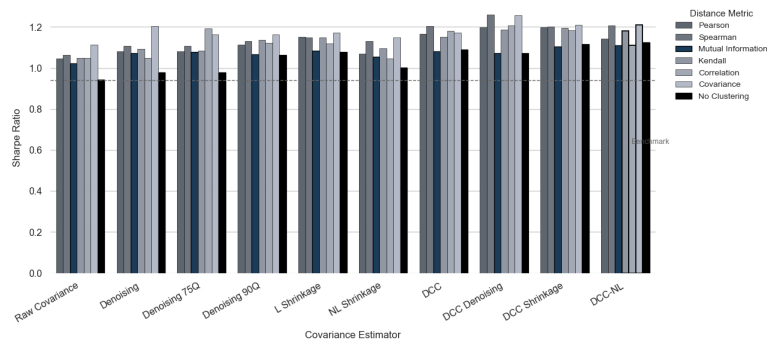
(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.43	14.67	1.0516	Raw Covariance	15.66	14.21	1.1017
Denoising (Mean)	17.01	14.75	1.1531	Denoising (Mean)	16.31	14.36	1.1355
Denoising (75th)	16.77	14.32	1.1711	Denoising (75th)	16.22	14.38	1.1273
Denoising (90th)	16.05	14.05	1.1425	Denoising (90th)	15.71	13.70	1.1466
Shrinkage	18.87	16.40	1.1508	Shrinkage	14.81	12.98	1.1412
NL Shrinkage	18.87	15.92	1.1856	NL Shrinkage	14.41	13.00	1.1088
DCC	14.54	13.02	1.1168	DCC	14.41	13.00	1.1087
DCC Denoising	14.96	12.92	1.1580	DCC Denoising	14.25	13.37	1.0663
DCC Shrinkage	17.98	16.45	1.0930	DCC Shrinkage	16.33	13.76	1.1868
DCC-NL	14.44	13.17	1.0967	DCC-NL	14.41	13.00	1.1087
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	17.02	15.52	1.0965	Raw Covariance	15.25	14.53	1.0498
Denoising (Mean)	17.15	15.83	1.0838	Denoising (Mean)	16.02	14.61	1.0966
Denoising (75th)	16.79	15.92	1.0551	Denoising (75th)	15.93	14.65	1.0867
Denoising (90th)	17.00	15.19	1.1186	Denoising (90th)	16.93	14.89	1.1365
Shrinkage	18.56	15.98	1.1612	Shrinkage	16.93	14.89	1.1365
NL Shrinkage	15.68	13.30	1.1786	NL Shrinkage	16.84	15.15	1.1112
DCC	15.32	13.34	1.1486	DCC	13.98	13.22	1.0573
DCC Denoising	15.77	13.45	1.1719	DCC Denoising	14.21	13.12	1.0831
DCC Shrinkage	15.49	13.33	1.1623	DCC Shrinkage	14.37	13.24	1.0856
DCC-NL	15.71	13.34	1.1782	DCC-NL	14.48	13.19	1.0985
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.18	14.77	1.0281	Raw Covariance	15.43	14.67	1.0516
Denoising (Mean)	15.66	14.84	1.0555	Denoising (Mean)	16.04	14.74	1.0881
Denoising (75th)	15.59	14.81	1.0527	Denoising (75th)	16.18	14.79	1.0939
Denoising (90th)	14.63	13.70	1.0675	Denoising (90th)	17.08	15.15	1.1276
Shrinkage	16.40	15.31	1.0714	Shrinkage	17.19	15.30	1.1237
NL Shrinkage	16.46	15.38	1.0699	NL Shrinkage	14.15	13.43	1.0531
DCC	14.36	13.77	1.0429	DCC	14.54	13.02	1.1168
DCC Denoising	14.36	13.80	1.0407	DCC Denoising	16.27	14.20	1.1457
DCC Shrinkage	14.63	13.70	1.0675	DCC Shrinkage	14.53	13.42	1.0826
DCC-NL	14.63	13.70	1.0675	DCC-NL	14.54	13.02	1.1168

improves the allocation of capital across assets. Notably, the portfolio constructed using DCC-Denoising under hierarchical clustering with the covariance distance achieves near-identical results, delivering a return of 16.37%, volatility of 13.02%, and a Sharpe ratio of 1.2577, further indicating the strong performance of the covariance distance under the MS objective as shown in section 3.5.4.

Under the ONC algorithm, no single distance metric dominates consistently. Figure 3.5 confirms this: in Figure 3.5a, the bars are tightly compressed across all covariance estimators, with every NCO portfolio clearing the benchmark line but no metric pulling ahead clearly. In Figure 3.5b, the spread widens at the DCC estimators, where Spearman and covariance distance produce the tallest bars, while across the static estimator panels, the covariance distance bars are nearly consistently among the highest, reflecting the stability advantage it carries under



(a) Sharpe ratios under Random Forest forecasts — ONC algorithm.



(b) Sharpe ratios under Random Forest forecasts — Hierarchical Clustering (Ward Linkage)).

Figure 3.5: Comparison of Sharpe Ratios under Random Forest forecasts – MS Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.

hierarchical clustering’s irreversible merge structure.

LASSO

Close behind, the strongest LASSO specification is obtained under the ONC algorithm, combining DCC-Denoising with Pearson distance and achieving a Sharpe ratio of 1.2499, while the Spearman distance under the same specification achieves a closely competing Sharpe ratio of 1.217, as shown in Table 3.15.

A recurring pattern across these results is the prominence of DCC-based estimators near the top, regardless of the forecasting model employed. This is noteworthy: it suggests that allowing correlations to evolve dynamically over time continues to add meaningful value regardless of the optimisation objective. Equally, the consistent outperformance of denoised covariance matrices points to a broader finding — that removing noise from the eigenvalue structure does not merely offer marginal gains in isolated cases, but contributes to more stable and reliable portfolio performance across the board.

Table 3.14: **Out-of-sample portfolio performance under Random Forest forecasts (Hierarchical Clustering) – MS Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.15	14.48	1.0463	Raw Covariance	15.40	14.48	1.0640
Denoising (Mean)	16.30	15.07	1.0817	Denoising (Mean)	16.33	14.75	1.1070
Denoising (75th)	16.21	15.00	1.0807	Denoising (75th)	16.34	14.78	1.1062
Denoising (90th)	15.65	14.07	1.1120	Denoising (90th)	15.88	14.06	1.1296
Shrinkage	17.20	14.94	1.1513	Shrinkage	17.21	14.98	1.1489
NL Shrinkage	15.97	14.93	1.0696	NL Shrinkage	16.94	14.98	1.1309
DCC	15.01	12.86	1.1671	DCC	15.88	13.17	1.2053
DCC Denoising	15.58	13.02	1.1968	DCC Denoising	16.35	12.99	1.2590
DCC Shrinkage	16.61	13.86	1.1986	DCC Shrinkage	16.82	14.00	1.2011
DCC-NL	15.01	13.13	1.1434	DCC-NL	15.77	13.08	1.2058
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.06	14.74	1.0216	Raw Covariance	15.43	14.73	1.0475
Denoising (Mean)	15.95	14.90	1.0708	Denoising (Mean)	16.40	14.99	1.0940
Denoising (75th)	16.03	14.88	1.0771	Denoising (75th)	16.29	15.01	1.0848
Denoising (90th)	15.09	14.15	1.0666	Denoising (90th)	16.26	14.29	1.1379
Shrinkage	16.54	15.28	1.0825	Shrinkage	17.53	15.27	1.1481
NL Shrinkage	16.06	15.25	1.0534	NL Shrinkage	16.79	15.32	1.0957
DCC	14.44	13.35	1.0817	DCC	15.50	13.48	1.1499
DCC Denoising	14.31	13.34	1.0723	DCC Denoising	15.94	13.44	1.1866
DCC Shrinkage	15.77	14.28	1.1041	DCC Shrinkage	17.05	14.26	1.1949
DCC-NL	14.90	13.41	1.1109	DCC-NL	15.85	13.44	1.1799
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.60	14.86	1.0497	Raw Covariance	15.81	14.19	1.1140
Denoising (Mean)	15.60	14.86	1.0497	Denoising (Mean)	17.28	14.36	1.2037
Denoising (75th)	15.66	13.12	1.1929	Denoising (75th)	16.09	13.85	1.1617
Denoising (90th)	16.02	14.27	1.1232	Denoising (90th)	16.09	13.85	1.1617
Shrinkage	16.94	15.13	1.1196	Shrinkage	17.17	14.66	1.1714
NL Shrinkage	15.82	15.14	1.0449	NL Shrinkage	16.79	14.60	1.1498
DCC	15.34	13.00	1.1792	DCC	15.52	13.24	1.1720
DCC Denoising	15.71	13.02	1.2066	DCC Denoising	16.37	13.02	1.2577
DCC Shrinkage	16.63	14.03	1.1848	DCC Shrinkage	16.86	13.92	1.2112
DCC-NL	15.21	13.71	1.1094	DCC-NL	15.65	12.95	1.2090

Figure 3.6 shows that under both clustering regimes, DCC-Denoising consistently produces the strongest results, with all NCO portfolios clearing the benchmark line throughout.

Gradient Boosting

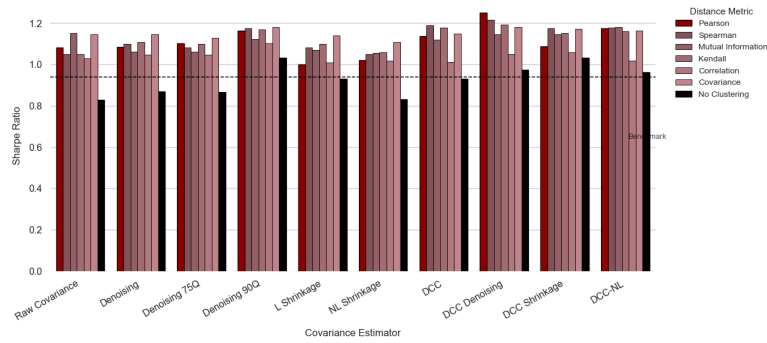
Gradient Boosting produces the weakest portfolio performance of the three models across both clustering regimes, though it clears the MS benchmark in most specifications. The best Sharpe ratio, 1.1076, is obtained under hierarchical clustering with Denoising (Mean) and mutual information distance Table 3.18. Under the ONC algorithm, DCC estimators collapse almost uniformly: as Table 3.17 and Figure 3.7a show, base DCC falls below the benchmark of 0.9465 in five of the six distance metrics, reaching 0.8716 at its lowest, with mutual information distance

Table 3.15: Out-of-sample portfolio performance under LASSO (ONC Algorithm) – MS Objective — Comparison of covariance estimators under different distance metrics

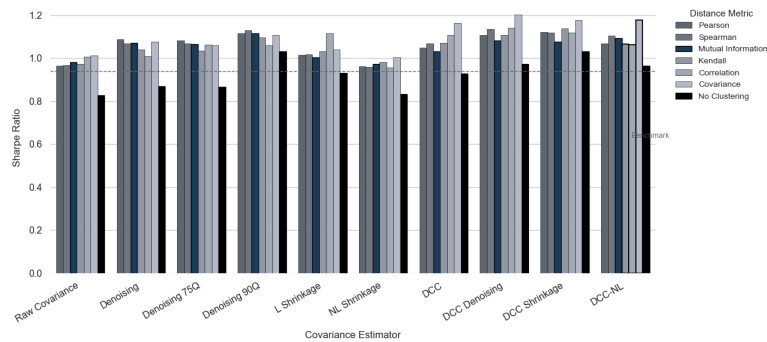
(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	19.82	18.30	1.0827	Raw Covariance	19.38	18.44	1.0510
Denoising (Mean)	20.23	18.65	1.0847	Denoising (Mean)	20.44	18.59	1.0994
Denoising (75th)	20.46	18.54	1.1036	Denoising (75th)	20.27	18.72	1.0826
Denoising (90th)	20.26	17.41	1.1639	Denoising (90th)	20.31	17.28	1.1753
Shrinkage	20.27	20.28	0.9992	Shrinkage	20.06	18.56	1.0809
NL Shrinkage	20.81	20.40	1.0199	NL Shrinkage	20.29	19.33	1.0500
DCC	18.06	15.89	1.1362	DCC	18.51	15.56	1.1899
DCC Denoising	19.77	15.82	1.2499	DCC Denoising	19.22	15.80	1.2170
DCC Shrinkage	20.84	19.19	1.0864	DCC Shrinkage	19.21	16.36	1.1745
DCC-NL	18.06	15.36	1.1758	DCC-NL	18.46	15.66	1.1790
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	18.37	15.95	1.1520	Raw Covariance	19.26	18.33	1.0508
Denoising (Mean)	20.71	19.52	1.0608	Denoising (Mean)	20.63	18.62	1.1077
Denoising (75th)	20.74	19.55	1.0607	Denoising (75th)	20.49	18.65	1.0986
Denoising (90th)	20.56	18.33	1.1214	Denoising (90th)	20.28	17.34	1.1690
Shrinkage	20.92	19.54	1.0709	Shrinkage	20.29	18.48	1.0979
NL Shrinkage	20.92	19.80	1.0562	NL Shrinkage	20.41	19.27	1.0588
DCC	17.63	15.73	1.1211	DCC	18.56	15.74	1.1787
DCC Denoising	19.62	17.14	1.1445	DCC Denoising	19.18	16.09	1.1917
DCC Shrinkage	19.62	17.14	1.1445	DCC Shrinkage	19.09	16.57	1.1520
DCC-NL	18.38	15.56	1.1811	DCC-NL	18.42	15.89	1.1597
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	19.26	18.72	1.0287	Raw Covariance	18.22	15.89	1.1465
Denoising (Mean)	19.90	19.00	1.0472	Denoising (Mean)	21.55	18.83	1.1446
Denoising (75th)	19.87	18.98	1.0468	Denoising (75th)	21.29	18.89	1.1274
Denoising (90th)	19.63	17.81	1.1018	Denoising (90th)	18.78	15.92	1.1794
Shrinkage	19.51	19.32	1.0098	Shrinkage	21.30	18.69	1.1394
NL Shrinkage	19.95	19.60	1.0182	NL Shrinkage	21.38	19.28	1.1094
DCC	17.16	16.98	1.0109	DCC	18.28	15.89	1.1500
DCC Denoising	17.70	16.88	1.0485	DCC Denoising	19.22	16.29	1.1797
DCC Shrinkage	18.83	17.79	1.0584	DCC Shrinkage	18.70	15.96	1.1720
DCC-NL	17.75	17.45	1.0168	DCC-NL	18.58	15.99	1.1624

the sole exception at 0.9667. DCC Denoising follows the same pattern throughout. Static estimators take over, with Shrinkage under Pearson distance reaching 1.0761. Under hierarchical clustering, DCC Shrinkage with Spearman distance recovers to 1.1063, and Figure 3.7b shows a more uniform picture above the benchmark, with mutual information distance producing the strongest static results.

Comparing results across the two optimisation objectives reveals an important distinction in how the NCO framework generates performance improvements. Under the Maximum Quadratic Utility objective, clustering simultaneously increases returns and reduces volatility. This reflects the direct variance penalty built into the objective function, which means the framework actively rewards both a stronger return signal and a more stable covariance structure. Under the Maximum Sharpe Ratio objective, the benchmark already operates at relatively low volatility, so there is



(a) Sharpe ratios under LASSO forecasts — ONC algorithm.



(b) Sharpe ratios under LASSO forecasts — Hierarchical Clustering (Ward Linkage)).

Figure 3.6: Comparison of Sharpe Ratios under LASSO forecasts – MS Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.

limited scope for further risk reduction. The performance gains here come primarily from higher returns, as the NCO framework improves capital allocation across asset clusters and enables the optimiser to better utilise the forecasting signal.

Across both optimisation objectives, Random Forest produces the strongest results: Sharpe ratios of 1.2305 under MQ and 1.259 under MS, both driven by hierarchical clustering with a DCC estimator and Spearman distance. As a bagging estimator, it averages predictions across many trees, each trained on a bootstrap sample of the data and restricted to a random subset of the 214 predictors at every split, compressing forecast variance without narrowing the signal base and producing a cross-sectional ranking that holds up across the full stock universe. The consistent dominance of hierarchical clustering with Spearman distance across the top specifications reflects two reinforcing properties. Ward’s linkage builds compact, well-separated asset groups that share genuine return co-movement, producing stable clusters that translate directly into more reliable within-cluster optimisation under the NCO framework. Spearman distance reinforces this by capturing rank-based co-movement rather than linear correlation,

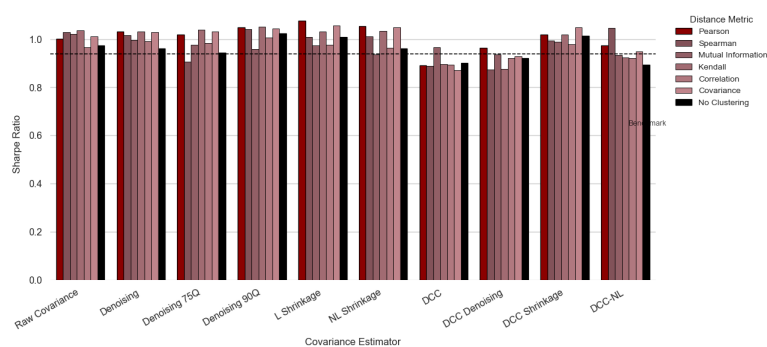
Table 3.16: **Out-of-sample portfolio performance under LASSO forecasts (Hierarchical Clustering) – MS Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	18.18	18.84	0.9647	Raw Covariance	18.01	18.64	0.9659
Denoising (Mean)	20.54	18.90	1.0865	Denoising (Mean)	19.84	18.57	1.0683
Denoising (75th)	20.43	18.89	1.0817	Denoising (75th)	19.86	18.61	1.0672
Denoising (90th)	19.70	17.67	1.1154	Denoising (90th)	19.42	17.20	1.1288
Shrinkage	19.58	19.30	1.0146	Shrinkage	19.42	19.09	1.0175
NL Shrinkage	18.79	19.55	0.9608	NL Shrinkage	18.77	19.59	0.9582
DCC	16.45	15.68	1.0492	DCC	17.17	16.09	1.0672
DCC Denoising	17.78	16.04	1.1085	DCC Denoising	18.30	16.11	1.1356
DCC Shrinkage	18.81	16.77	1.1218	DCC Shrinkage	18.85	16.84	1.1197
DCC-NL	16.67	15.60	1.0681	DCC-NL	17.54	15.88	1.1044
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	18.18	18.54	0.9805	Raw Covariance	19.40	19.97	0.9714
Denoising (Mean)	19.92	18.61	1.0703	Denoising (Mean)	20.55	19.75	1.0403
Denoising (75th)	19.82	18.63	1.0640	Denoising (75th)	20.49	19.81	1.0342
Denoising (90th)	19.33	17.35	1.1143	Denoising (90th)	19.87	18.11	1.0973
Shrinkage	19.22	19.16	1.0033	Shrinkage	20.45	19.86	1.0302
NL Shrinkage	19.05	19.57	0.9735	NL Shrinkage	19.93	20.34	0.9799
DCC	17.02	16.50	1.0312	DCC	17.90	16.70	1.0716
DCC Denoising	17.87	16.51	1.0824	DCC Denoising	18.35	16.57	1.1073
DCC Shrinkage	18.55	17.23	1.0765	DCC Shrinkage	19.82	17.42	1.1379
DCC-NL	17.80	16.28	1.0931	DCC-NL	17.51	16.45	1.0639
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	19.71	19.58	1.0070	Raw Covariance	19.27	19.02	1.0131
Denoising (Mean)	19.75	19.56	1.0096	Denoising (Mean)	20.45	18.98	1.0773
Denoising (75th)	20.43	19.25	1.0614	Denoising (75th)	20.12	19.00	1.0588
Denoising (90th)	20.41	19.25	1.0602	Denoising (90th)	20.29	18.33	1.1071
Shrinkage	19.88	17.82	1.1156	Shrinkage	20.19	19.42	1.0394
NL Shrinkage	18.98	19.87	0.9553	NL Shrinkage	19.56	19.47	1.0045
DCC	18.31	16.54	1.1073	DCC	19.05	16.39	1.1623
DCC Denoising	19.62	17.18	1.1420	DCC Denoising	19.78	16.46	1.2015
DCC Shrinkage	19.33	17.27	1.1195	DCC Shrinkage	20.16	17.11	1.1784
DCC-NL	17.16	16.15	1.0628	DCC-NL	19.04	16.17	1.1778

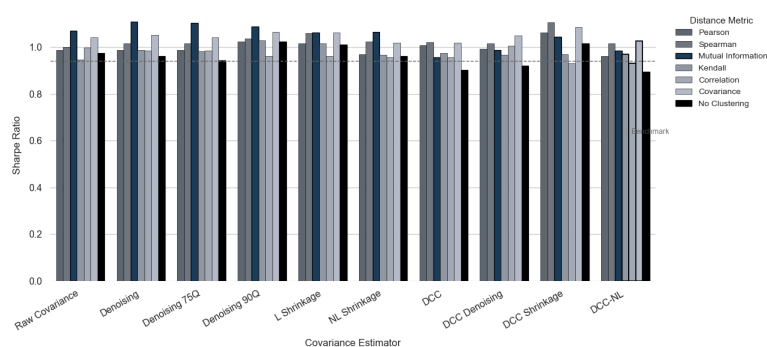
making the dependence structure robust to the fat tails and extreme observations that characterise equity return distributions. The combination filters estimation error at two levels before the optimiser receives the covariance matrix, which is why this pairing leads consistently across both objectives and both forecasting models. LASSO follows with competitive results, reaching 1.1284 under MQ and 1.2499 under MS, confirming that a well-regularised linear model can extract meaningful signal when paired with a robust optimisation framework. Gradient Boosting lags considerably across both objectives, an outcome consistent with the forecasting results in Section 3.5.1, where its predictive strength was shown to concentrate in the extreme deciles with irregular ordering in the middle. That structural limitation in the forecasting signal carries through directly into portfolio performance.

Table 3.17: **Out-of-sample portfolio performance under Gradient Boosting (ONC Algorithm) – MS Objective** — Comparison of covariance estimators under different distance metrics

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	16.68	16.63	1.0027	Raw Covariance	16.76	16.28	1.0294
Denoising (Mean)	17.16	16.64	1.0313	Denoising (Mean)	16.70	16.41	1.0176
Denoising (75th)	16.96	16.63	1.0198	Denoising (75th)	13.71	15.15	0.9053
Denoising (90th)	16.45	15.66	1.0501	Denoising (90th)	16.04	15.40	1.0415
Shrinkage	20.32	18.88	1.0761	Shrinkage	16.88	16.74	1.0085
NL Shrinkage	19.67	18.67	1.0535	NL Shrinkage	17.08	16.89	1.0111
DCC	13.62	15.28	0.8912	DCC	13.44	15.14	0.8876
DCC Denoising	14.21	14.75	0.9636	DCC Denoising	13.17	15.08	0.8733
DCC Shrinkage	18.74	18.40	1.0184	DCC Shrinkage	15.48	15.57	0.9941
DCC-NL	14.64	15.02	0.9751	DCC-NL	17.07	16.31	1.0468
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	17.29	16.94	1.0212	Raw Covariance	16.93	16.33	1.0371
Denoising (Mean)	16.91	16.97	0.9966	Denoising (Mean)	17.00	16.46	1.0328
Denoising (75th)	16.66	17.08	0.9756	Denoising (75th)	17.07	16.45	1.0380
Denoising (90th)	15.78	16.43	0.9602	Denoising (90th)	16.09	15.31	1.0510
Shrinkage	17.21	17.66	0.9746	Shrinkage	17.16	16.61	1.0329
NL Shrinkage	16.68	17.82	0.9360	NL Shrinkage	17.41	16.86	1.0330
DCC	14.60	15.10	0.9667	DCC	13.43	14.97	0.8972
DCC Denoising	14.20	15.18	0.9358	DCC Denoising	13.17	15.02	0.8772
DCC Shrinkage	16.13	16.32	0.9883	DCC Shrinkage	15.74	15.43	1.0199
DCC-NL	14.42	15.45	0.9329	DCC-NL	13.88	15.02	0.9244
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	16.18	16.73	0.9672	Raw Covariance	16.77	16.60	1.0106
Denoising (Mean)	16.36	16.50	0.9917	Denoising (Mean)	17.26	16.77	1.0295
Denoising (75th)	16.28	16.54	0.9846	Denoising (75th)	17.32	16.78	1.0324
Denoising (90th)	15.62	15.53	1.0058	Denoising (90th)	16.18	15.51	1.0432
Shrinkage	16.45	16.86	0.9757	Shrinkage	17.90	16.96	1.0558
NL Shrinkage	16.64	17.26	0.9643	NL Shrinkage	18.14	17.29	1.0492
DCC	14.02	15.68	0.8940	DCC	13.27	15.22	0.8716
DCC Denoising	14.28	15.50	0.9216	DCC Denoising	13.75	14.80	0.9294
DCC Shrinkage	15.64	15.96	0.9799	DCC Shrinkage	16.39	15.62	1.0488
DCC-NL	14.28	15.50	0.9216	DCC-NL	14.10	14.84	0.9500



(a) Sharpe ratios under Gradient Boosting forecasts — ONC algorithm.



(b) Sharpe ratios under Gradient Boosting forecasts — Hierarchical Clustering (Ward Linkage)).

Figure 3.7: Comparison of Sharpe Ratios under Gradient Boosting forecasts – MS Objective. Each subfigure reports Sharpe ratios across covariance estimators and dependence measures.

Table 3.18: **Out-of-sample portfolio performance under Gradient Boosting forecasts (Hierarchical Clustering) – MS Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.32	15.53	0.9868	Raw Covariance	15.97	15.97	1.0002
Denoising (Mean)	15.32	15.53	0.9868	Denoising (Mean)	16.41	16.16	1.0154
Denoising (75th)	15.61	15.82	0.9866	Denoising (75th)	16.40	16.16	1.0146
Denoising (90th)	15.49	15.15	1.0228	Denoising (90th)	16.00	15.45	1.0356
Shrinkage	16.47	16.24	1.0144	Shrinkage	17.50	16.52	1.0593
NL Shrinkage	15.45	15.96	0.9685	NL Shrinkage	16.89	16.49	1.0244
DCC	14.32	14.19	1.0089	DCC	14.61	14.32	1.0203
DCC Denoising	14.12	14.22	0.9926	DCC Denoising	14.43	14.20	1.0160
DCC Shrinkage	16.00	15.06	1.0623	DCC Shrinkage	16.71	15.11	1.1063
DCC-NL	14.16	14.72	0.9621	DCC-NL	14.63	14.40	1.0159
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	17.21	16.09	1.0697	Raw Covariance	15.50	16.39	0.9459
Denoising (Mean)	18.01	16.26	1.1076	Denoising (Mean)	16.50	16.70	0.9879
Denoising (75th)	17.98	16.31	1.1026	Denoising (75th)	16.42	16.72	0.9824
Denoising (90th)	16.60	15.26	1.0879	Denoising (90th)	16.08	15.65	1.0272
Shrinkage	17.70	16.67	1.0621	Shrinkage	17.06	16.80	1.0155
NL Shrinkage	17.72	16.63	1.0651	NL Shrinkage	16.27	16.82	0.9670
DCC	14.18	14.83	0.9559	DCC	14.13	14.52	0.9729
DCC Denoising	14.43	14.61	0.9876	DCC Denoising	14.10	14.57	0.9672
DCC Shrinkage	15.92	15.27	1.0426	DCC Shrinkage	14.33	14.79	0.9689
DCC-NL	14.55	14.79	0.9836	DCC-NL	14.33	14.79	0.9689
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	16.45	16.51	0.9962	Raw Covariance	16.61	15.94	1.0420
Denoising (Mean)	16.33	16.58	0.9852	Denoising (Mean)	16.92	16.09	1.0516
Denoising (75th)	16.33	16.58	0.9852	Denoising (75th)	16.76	16.11	1.0405
Denoising (90th)	16.10	16.75	0.9612	Denoising (90th)	17.48	16.43	1.0637
Shrinkage	16.10	16.75	0.9611	Shrinkage	17.31	16.31	1.0612
NL Shrinkage	13.77	14.42	0.9552	NL Shrinkage	14.80	14.53	1.0185
DCC	13.77	14.42	0.9552	DCC	14.80	14.53	1.0185
DCC Denoising	15.31	15.24	1.0044	DCC Denoising	15.05	14.34	1.0497
DCC Shrinkage	13.76	14.78	0.9307	DCC Shrinkage	16.65	15.34	1.0852
DCC-NL	13.75	14.78	0.9299	DCC-NL	15.18	14.79	1.0260

3.5.7 Portfolio Weight and Diversification Analysis

Figures 3.8 and 3.9 report allocation diagnostics for the best-performing portfolio specification identified in the preceding analysis, namely, Random Forest return forecasts paired with hierarchical clustering and a dynamic covariance estimator under the Spearman distance metric. For consistency, the comparison portfolios employ the same forecasting model and covariance estimator but use mean–variance optimisation to isolate the effect of the NCO framework on portfolio allocation and diversification. Two aspects of portfolio structure are examined: the number of active positions, taken as assets assigned weights greater than 0.001, and the maximum portfolio weight, reflecting the single largest allocation in the portfolio. Given that weights are bounded between zero and 0.5, the maximum weight additionally reveals how far the optimiser is drawn toward this upper constraint.

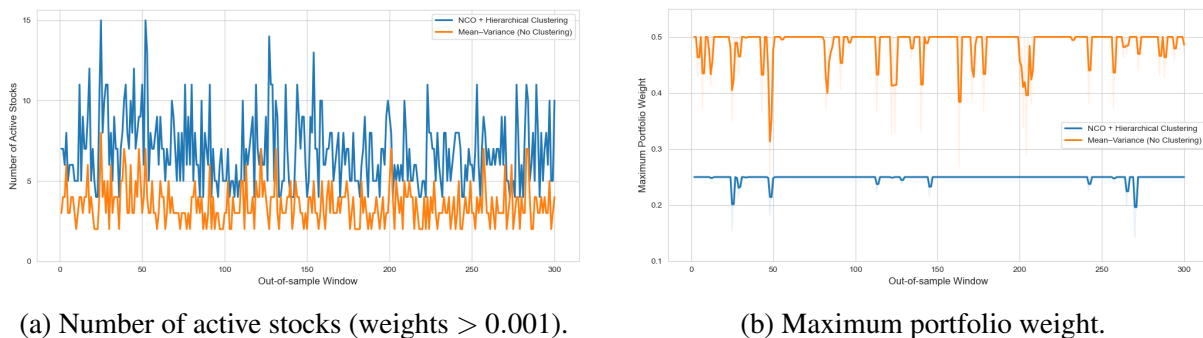
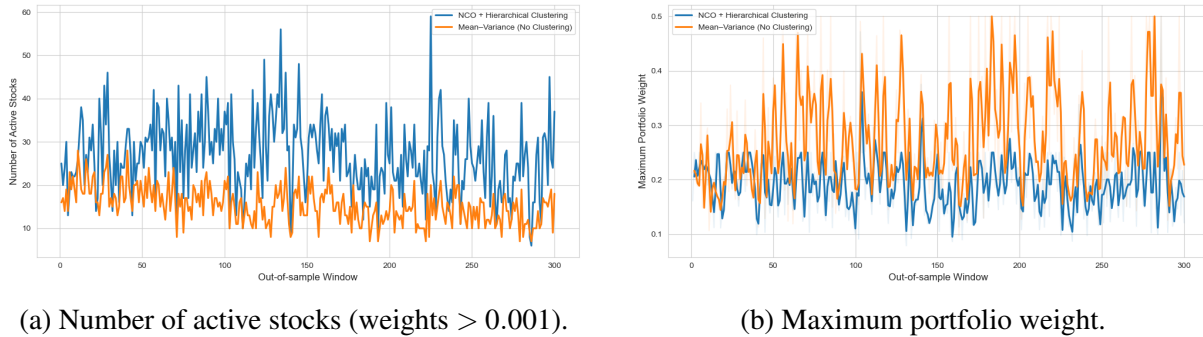


Figure 3.8: Portfolio allocation characteristics under the Maximum Quadratic Utility objective.

In Figure 3.8, the MQ mean–variance portfolio (shown in orange) is heavily concentrated across all windows, with an average number of active positions of just 3.5 and an average maximum weight of 0.487, which is very close to the imposed upper bound of 0.5. On the other hand, NCO-constructed portfolios show improvement, with active positions roughly doubled, averaging approximately 7, and the maximum weight dropping to around 0.24, reflecting a more measured distribution of capital.

Shifting to the Maximum Sharpe Ratio objective, the contrast between the two approaches persists, although at a higher overall level of diversification. In Figure 3.9, the mean–variance portfolio average around 15.4 active positions across all out-of-sample windows, with the largest single allocation reaching 0.272. The NCO portfolios distribute capital considerably more

broadly, engaging an average of 26.8 assets while holding the maximum weight of around 0.193. These results indicate that the NCO framework not only increases the number of assets with active weights but also reduces concentration in individual positions.



(a) Number of active stocks (weights > 0.001).

(b) Maximum portfolio weight.

Figure 3.9: Portfolio allocation characteristics under the Maximum Sharpe Ratio objective.

Since both portfolios rely on the same inputs, the diversification benefit can be attributed to the behaviour of the NCO framework, which enforces balance within clusters before aggregating allocations across them. The framework effectively introduces a structural discipline that the mean–variance optimiser alone does not impose, as the latter treats the asset universe as a single optimisation problem. By contrast, the hierarchical process through which capital is allocated under the NCO framework results in portfolios that are more evenly distributed across the opportunity set.

On average, the NCO portfolios contain 7 active positions under the maximum quadratic utility objective and 26.8 under the maximum Sharpe ratio objective, selected from roughly 300 eligible stocks per window. Both figures are consistent with the diversification literature, which finds that the required number of stocks depends on the target level of risk reduction. Evans and Archer [61] put the threshold for meaningful diversification at 8–10 stocks, Tang [62] showed analytically that 20 stocks eliminates 95% of diversifiable risk on average, and Alexeev and Dungey [63] found that 7 stocks is sufficient for an 85% reduction, rising to 10 for a 90% target.

3.6 Conclusion

This chapter builds on the material in Chapter 2 by moving beyond traditional mean–variance optimisation and developing a more robust investment framework. The aim was to combine three elements: improved covariance matrix estimation, machine-learning-based return forecasts, and an optimisation approach that reduces dimensionality by clustering assets and allocating weights at both the cluster and asset level. These components were brought together within the NCO framework, using López de Prado’s original ONC algorithm and, in parallel, hierarchical clustering. The NCO framework is itself motivated by robustness: by decomposing the portfolio problem into cluster-level and asset-level steps, it limits the propagation of estimation error that causes classical mean–variance optimisation to produce unstable and concentrated allocations. Different distance metrics were also examined to understand how they influence the clustering and, ultimately, portfolio performance.

Across the forecasting models, Random Forest emerged as the strongest overall performer, closely followed by LASSO in both cross-sectional forecasting ability and portfolio optimisation results under both optimisation objectives. What both models share, despite their structural differences, is the ability to handle a large and noisy predictor set without overfitting: Random Forest through ensemble averaging that reduces forecast variance across the full asset universe, and LASSO through aggressive regularisation that discards uninformative predictors and retains a stable linear structure. Incorporating Random Forest forecasts within the NCO framework led to substantial improvement over portfolios constructed using historical mean returns and over forecast-based portfolios optimised without clustering, establishing that the predictive signal and the allocation structure each contribute independently. Among the clustering configurations, hierarchical clustering paired with rank-based distance metrics, particularly Spearman distance, produced the strongest portfolios under both objectives. This framework achieves a Sharpe ratio of 1.2305 under Random Forest and MQ objective, a 100.5% increase over the classical benchmark’s 0.6138, while under the MS objective it yields a Sharpe ratio of 1.2590, marking a 33.1% improvement over the corresponding benchmark. Unlike the ONC algorithm, which can revisit and correct poorly formed clusters, hierarchical clustering locks in every merge decision

as it is made, so any distortion in the distance metric propagates throughout the entire tree. Rank-based metrics are more tolerant of outliers and extreme return values than correlation-based ones, making them better suited to a procedure that offers no opportunity for correction.

The dynamic covariance estimators were consistent in their dominance throughout. The DCC-NL and DCC with denoising produced the strongest portfolios across forecasting models and clustering configurations, reflecting the importance of allowing correlations to evolve over time and of further refining the dynamic structure through nonlinear shrinkage and denoising. Static estimators, by assuming a fixed correlation structure, fail to capture the time-varying nature of asset relationships that dynamic models accommodate naturally.

The importance of the integrated NCO framework is further established through its diversification benefit. Relative to classical mean–variance optimisation applied to identical inputs, NCO portfolios held more active positions and reduced concentration in individual assets through its two-stage allocation mechanism, a structural advantage that cannot be attributed solely to the forecasting inputs.

The results establish the importance of each component and confirm that integrating refined forecasts and dynamic covariance estimation within the NCO framework produces risk-adjusted performance that the classical mean–variance approach does not match. The consistency of this result across models, estimators, and clustering configurations suggests that the gains are not incidental to any particular specification but are the consequence of building every component of the framework around the same principle: reducing sensitivity to estimation error at each stage of the portfolio construction process.

Machine Learning Forecasting and Nested Clustered Optimisation Across Market Regimes: Evidence from Hedge Fund Strategy Indices

Chapter 3 developed and validated a portfolio construction framework combining machine-learning return forecasts, various covariance-matrix estimators, and Nested Clustered Optimisation (NCO) across a large equity universe. This chapter extends the framework to hedge funds, a more complex asset class. Hedge funds present particular complexities: non-normal return distributions driven by leverage and derivative exposure, illiquidity-induced autocorrelation that distorts conventional risk measures, and correlation structures that can change markedly during periods of market stress. The main objectives of this chapter are twofold. The first is to assess whether machine learning models produce meaningful out-of-sample return forecasts for hedge fund strategy indices and whether predictability varies systematically across strategy types. The second is to evaluate a range of portfolio construction strategies applied to Bloomberg hedge fund strategy indices, including classical optimisation, multiple NCO variants, and a regime-switching strategy that dynamically allocates between portfolio objectives depending on prevailing market conditions. The aim is to determine which strategies deliver the most robust risk-adjusted performance across three market regimes (pre-COVID, COVID, and post-COVID). NCO's advantage over classical mean—variance optimisation proves to be regime-specific rather

than constant: it is largest during the COVID period, when hedge fund correlations converge and the covariance matrix becomes poorly conditioned, and narrows or reverses in the calmer pre- and post-COVID periods.

The remainder of the chapter is organised as follows. Section 4.1 presents an overview of hedge funds. Section 4.2 discusses the characteristics of hedge fund returns. Section 4.3 describes the data. Section 4.4 sets out the methodology. Sections 4.5 and 4.6 report the empirical results. Section 4.7 concludes.

4.1 Introduction

Over recent decades, hedge funds have grown into a substantial element of modern financial markets, attracting growing allocations from high-net-worth individuals and institutional investors. At their core, hedge funds bring together capital from sophisticated investors and place it under the discretion of skilled managers, who are free to deploy it across a diverse range of asset classes and instruments. The term "hedge" reflects the original intent of the strategy: to protect against market downturns. Though in practice, modern hedge funds pursue far broader mandates. Benefiting from comparatively loose regulatory oversight, managers are not obliged to report portfolio holdings and investment strategies. Moreover, they are not limited to conventional investment approaches; instead, they can actively employ leverage and short selling, trade non-traditional assets, and apply complex strategies, all with the objective of generating above-average returns in any market environment [64]. This active and flexible approach stands in stark contrast to traditional mutual funds, which typically employ long-only, buy-and-hold strategies across standard asset classes, operate under narrower mandates, and face more binding regulatory constraints. Hedge fund investors, however, trade liquidity for this access. Capital is typically locked up for an initial period before redemptions are permitted, reflecting the illiquid nature of many of the strategies employed.

Central to the hedge fund model is how managers are compensated. Traditional mutual funds typically charge a flat annual management fee, whereas hedge funds rely on a more layered arrangement that ties manager rewards directly to the returns they generate. The structure historically associated with hedge funds, widely known as "two and twenty", combines a fixed

management fee of around 2% of assets under management with a performance fee of roughly 20% of any profits made [64]. The management fee covers the fund's operational costs. However, the performance fee is where the real incentive lies. Managers can earn significantly more, but only when they deliver returns that genuinely exceed what an investor could have achieved through a passive vehicle. Most funds reinforce this further through a high-water mark provision, which prevents managers from collecting performance fees until prior losses have been recovered. Some funds go further still, imposing a hurdle rate: a minimum return threshold that must be met before any performance fee applies. This structure is designed to align the interests of managers and investors by rewarding genuine outperformance. At the same time, performance-based compensation is asymmetric in nature: managers share in the upside but face limited downside, and this asymmetry can encourage risk-taking behaviour, which in turn contributes to the non-normal return distributions commonly observed in hedge fund performance [65].

Hedge funds tend to be classified along two dimensions: first, by investment style and the approach adopted by the manager, and second, by the market or sector in which they operate. The composition of the industry has shifted considerably over time, with the dominance of individual strategies rising and falling in response to market conditions and performance cycles [66]. These strategies can be broadly grouped into five categories: relative value, event-driven, directional and trading, long/short equity, and a range of other approaches.

Relative value strategies aim to generate returns by exploiting pricing discrepancies between related securities, focusing on relative price movements rather than directional exposure to the broader market. These typically take the form of pairs trading or arbitrage, where managers construct long-short positions to profit from temporary mispricings between them. However, since these strategies rely on capturing small pricing differentials, transaction costs represent a meaningful constraint. The anticipated gains from these opportunities must be sufficient to clear execution costs and the cost of maintaining positions until the mispricing corrects, introducing both timing risk and capital efficiency considerations [67].

Other managers, by contrast, focus less on pricing relationships between securities and more on specific corporate events as the primary source of their returns. Event-driven strategies seek to capitalise on opportunities created by extraordinary corporate catalysts. These are events that

have the potential to drive a company's value towards a fundamentally different value. These might include mergers, acquisitions, restructurings, or distressed situations. Each represents a moment of corporate uncertainty that managers seek to turn into an investment opportunity. Success in this strategy depends heavily on accurately assessing the probability and timing of such events, as any delay in the anticipated catalyst can meaningfully reduce returns [66].

Not all managers, however, concern themselves with individual securities or corporate events. Some take a far broader view, positioning entire portfolios around anticipated movements in macroeconomic conditions, interest rates, currencies, and commodity prices. These are broadly referred to as directional and trading strategies, a category that includes a range of approaches, most notably macro and managed futures.

Global macro funds represent the broadest expression of the directional side. Managers invest in virtually any market using any financial instrument, with decisions driven by the manager's own macroeconomic views, top-down analysis, and market forecasts generated through econometric models [66]. While managed futures strategies share the same directional nature, human discretion plays a far more limited role. Systematic, computerized models generate the majority of trading decisions across futures markets worldwide, with the manager's role limited to periodically recalibrating the model's parameters.

Finally, and perhaps the most classic of all hedge fund strategies, long/short equity stands as the most widely adopted approach in the industry today. The underlying logic suggests that if research can identify both undervalued and overvalued stocks, there is no reason to bet on only one direction. The strategy traces its roots to 1949, when Alfred Winslow Jones, widely regarded as the father of the modern hedge fund, demonstrated that pairing long positions in undervalued stocks with short positions in overvalued ones could produce returns that reflect the manager's stock picking ability rather than the general direction of the market [64]. These categories, among others, capture the dominant approaches within the industry, though hedge fund strategies in practice extend well beyond any predefined set.

Since different strategies produce markedly different return profiles and respond differently to market conditions, institutional investors and high-net-worth individuals have increasingly turned to the fund of hedge funds structure as a means of gaining diversified exposure across

multiple strategies and managers through a single vehicle.

4.2 Hedge Fund Returns

4.2.1 Return Characteristics

Hedge fund returns exhibit several structural features that set them apart from traditional asset classes, with direct consequences for how performance is evaluated and risk is measured.

Hedge fund returns can be decomposed into two components: beta, which captures exposure to various asset classes and style risk factors, and alpha, which captures the portion attributable to manager skill [68]. In the literature, alpha and beta are estimated using factor models that range from single-factor specifications, where the market return alone serves as the benchmark, to multi-factor frameworks that account for exposures across equity, fixed income, credit, and other risk dimensions. Hedge funds were originally designed to be market neutral, combining long and short positions to cancel out broad market exposure and isolate returns from manager skill rather than market direction. A fund achieving this would carry near-zero market beta, earning its returns from alpha alone rather than from riding general market movements. In practice, this has proved difficult to maintain. Liang [69] uses an eight-asset-class factor model to test this and finds that hedge funds, while less correlated with stocks and bonds than mutual funds, are not genuinely neutral. Patton [70] tests neutrality across four dimensions separately, mean, variance, VaR, and tail neutrality, and finds that most funds classified as market neutral are not. The tail result is the most consequential: a fund can appear uncorrelated with the market across years of normal data and then co-move with it sharply in a downturn, which is precisely the scenario that matters for an investor holding it as a hedge.

This failure of market neutrality is largely a product of how hedge funds actually operate. The dynamic use of leverage, short selling, and derivatives produces a risk-return profile that standard linear factor models cannot adequately capture. Agarwal and Naik [65] make this concrete: across a wide range of equity-oriented strategies, they found that the return profile resembles writing a put option on a market index. Collecting steady small gains most of the time and then sustaining heavy losses when markets fall sharply, these strategies generate

return distributions with negative skewness and fat left tails, since losses during market stress can be large relative to the steady gains accumulated in normal periods. Liang and Park [71] compare a range of alternative risk measures and find that standard deviation alone fails to capture the true risk profile of hedge funds, with skewness and kurtosis playing a material role in explaining return behaviour across strategies. Mean–variance optimisation is therefore poorly suited to hedge fund portfolios, since it assumes normally distributed returns and systematically underestimates downside risk. Agarwal and Naik [65] demonstrate this empirically, finding that mean–variance optimal portfolios substantially understate tail losses relative to those constructed under a mean-CVaR framework.

Serial autocorrelation is another feature that distinguishes hedge fund returns from most asset classes. Getmansky et al. [72] attribute this to illiquidity exposure: when funds hold assets that trade infrequently and lack observable market prices, the returns they report tend to lag true economic value, producing a smoother series than actually occurred. This artificial smoothness suppresses measured volatility and inflates risk-adjusted performance metrics such as the Sharpe ratio. The degree of autocorrelation varies systematically across hedge fund style categories, with strategies concentrated in illiquid holdings exhibiting substantially higher serial correlation than those invested in more actively traded securities. Getmansky et al. [72] find that illiquidity is the most likely explanation for this pattern, though they acknowledge that other sources of serial correlation cannot be fully excluded.

4.2.2 Return Predictability

The question of whether hedge fund returns are forecastable has generated a substantial body of evidence, with conclusions that vary depending on the method employed, the sample period, and the strategy under examination. Early work by Amenc et al. [73] was among the first to address this question systematically, using multi-factor models built from economic variables to assess out-of-sample predictability across hedge fund styles, providing early evidence that hedge fund returns contain a forecastable component. Hamza et al. [74] extended this analysis using a broader set of risk factors and a longer time series, similarly finding evidence in favour of predictability.

The risk factor literature established both the drivers of hedge fund returns and the natural inputs for predictive models. Fung and Hsieh [75, 76] showed that a compact set of trend-following and equity-oriented factors captures substantial variation in returns across strategies. Bali et al. [77] found that the default premium beta is positively associated with future returns, whereas the inflation beta is negatively associated with them. Beyond individual factor betas, broader risk structure also predicts returns: Titman and Tiu [78] found that funds with lower R-squares on risk factors tend to achieve higher Sharpe ratios, suggesting that less factor-dependent managers generate more persistent alpha, while Bali et al. [79] found that systematic risk dominates the cross-sectional dispersion of future returns. Fund-level characteristics add a further predictive layer: Sun et al. [80] showed that strategy distinctiveness predicts subsequent performance, pointing to information beyond what factor exposures alone can explain. Macroeconomic variables extend this further: Avramov et al. [81] demonstrated that the default spread and VIX improve on pure factor-based benchmarks, while Wegener et al. [82] showed that explicitly accounting for non-normality, heteroskedasticity, and time-varying exposures yields further gains in out-of-sample accuracy. At the fund level, Avramov et al. [83] found that predictability is partly explained by leverage and capacity constraints, with forecast combination consistently outperforming any individual predictor model.

Panopoulou and Vrontos [84] brought these lines together in the most directly relevant prior study. Using predictors from the Fung-Hsieh factor library, Fama-French factors, and macroeconomic variables, they evaluated out-of-sample predictability across 15 HFR hedge fund strategy indices, comparing forecast combination methods with information combination approaches including kitchen sink, bagging, LASSO, ridge, and elastic net specifications. Simple forecast combinations outperformed the more sophisticated information combination methods statistically, and dynamically constructing portfolios based on combination forecasts delivered considerably improved portfolio performance relative to the historical average, with forecast combination proving the superior approach overall.

The structural features of hedge fund return prediction, including high predictor dimensionality, regime sensitivity, and non-normal return distributions, suggest that linear forecasting approaches face inherent limitations and that methods capable of capturing nonlinear relation-

ships and predictor interactions may offer further predictive gains. Nonlinear machine learning models are well-suited to this setting, and their application to return prediction across broad asset universes has produced substantial improvements over traditional factor models [52, 85]. The present study applies this class of methods to nineteen Bloomberg hedge fund strategy indices and, rather than evaluating forecast performance in isolation, integrates the resulting return forecasts directly into the NCO portfolio construction process alongside covariance estimation, allowing the contribution of forecasting to portfolio performance to be assessed explicitly.

4.3 Data

4.3.1 Hedge Fund Indices

We use monthly return data on nineteen Bloomberg hedge fund indices from the Bloomberg Terminal, covering January 2014 to March 2026. The indices are constructed by Bloomberg L.P. from hedge funds reporting their performance and assets under management to Bloomberg's database. Within each strategy category, funds are weighted by assets under management, meaning larger funds have proportionally more influence on the index return. The constituent universe is reviewed quarterly, and the indices are rebalanced monthly. Returns are reported net of all fees and expenses, which is consistent with standard practice in the fund-of-funds literature.

The nineteen indices span four broad strategy groups consistent with standard industry classification practice. The Bloomberg All Hedge Fund Index (BHEDGE) is a broad composite covering all strategies and serves as the benchmark throughout the analysis. The remaining eighteen indices form the investable universe from which portfolios are constructed. The equity strategies comprise Equity Hedge Fund Index (BHEQTY), Equity Long/Short (BHEQLS), Equity Multi-Strategy (BHEQMS), Fundamental Equity Market Neutral (BHFEMN), and Activist (BHACT). Equity hedge funds take both long and short positions in equities across a range of investment processes and market capitalisations. The event-driven strategies include Event Driven Hedge Fund Index (BHEVT), Event Driven Multi-Strategy (BHEVMS), Merger Arbitrage (BHMARB), and Distressed Debt (BHDIST), which exploit corporate events such

as mergers, restructurings, and distressed situations. The macro strategies comprise the Macro Hedge Fund Index (BHMACR), Macro Multi-Strategy (BHMMS), and Commodity (BHCOM), covering discretionary and systematic strategies driven by top-down macroeconomic views across currencies, rates, and commodities. The relative value and credit strategies include the Relative Value Hedge Fund Index (BHRV), Fixed Income Arbitrage (BHFIA), Convertible Bond Arbitrage (BHCBA), Credit Hedge Fund Index (BHCRTD), Credit Multi-Strategy (BHCRTS), and Emerging Market Debt (BHEMDB), which pursue arbitrage opportunities in fixed income, convertible securities, and credit markets.

All indices are converted to monthly simple returns for the empirical analysis. BHEDGE serves as the benchmark throughout, representing broad hedge fund market exposure against which the constructed fund-of-funds portfolios are evaluated. The in-sample estimation period runs from January 2014 to December 2018, yielding 60 monthly observations, and the out-of-sample evaluation period runs from January 2019 to March 2026, yielding 87 monthly observations.

4.3.2 Predictive Factors

We model hedge fund index returns using a set of 24 predictive factors drawn from three groups of variables along the lines of Agarwal and Naik [65], Fung and Hsieh [75, 76], and Wegener et al. [82]. All factors are lagged by one month to ensure that only information available at the time of forecast formation is used in the predictive models.

The first set of factors comprises the Fung and Hsieh [76] risk factors, which have been shown to achieve considerable explanatory power in describing hedge fund returns. Five of these are trend-following risk factors, constructed as returns on portfolios of lookback straddle options on bonds (PTFSBD), currencies (PTFSFX), commodities (PTFSCOM), short-term interest rates (PTFSIR), and stock indices (PTFSSTK), designed to replicate the maximum achievable return from trend-following strategies in their respective underlying asset classes. These factors are obtained from David Hsieh's data library at Duke University.¹ We also include two equity-oriented factors: the S&P 500 total return index (SP500) and the size spread factor, defined as the

¹Data available at: <https://people.duke.edu/~dah7/DataLibrary/TF-Fac.xls>

Russell 2000 monthly total return minus the S&P 500 monthly total return (SIZE_SPREAD), both sourced from LSEG Datastream. The bond-oriented factors consist of the bond market factor, measured as the monthly change in the 10-year Treasury constant maturity yield (BOND_MKT), and the credit spread factor, measured as the monthly change in the difference between Moody's BAA yield and the 10-year Treasury constant maturity yield (CREDIT_SPREAD), both obtained from the Federal Reserve Bank of St. Louis. Finally, we include the emerging market factor, represented by the MSCI Emerging Markets index monthly total return (MSCI_EM), sourced from LSEG Datastream. This factor was added to the original seven-factor model by Fung and Hsieh as an eighth factor to capture exposure to emerging market risk.

The second set of factors relates to style investing and captures size and value mispricings in equity markets. Specifically, we employ the Fama and French factors, namely the market excess return (Mkt-RF), the size factor (SMB), the value factor (HML), and the risk-free rate (RF), along with the momentum factor (Mom), the short-term reversal factor (ST_Rev), and the long-term reversal factor (LT_Rev). These factors are obtained from Kenneth French's data library.² The third set of factors consists of macroeconomic and business cycle variables following Wegener et al. [82]. We employ the US industrial production index growth rate (INDPRO), the monthly growth rate in non-farm payrolls (PAYEMS), the change in the consumer price index as a measure of inflation (CPIAUCSL), the monthly change in the trade-weighted US dollar index against advanced foreign economies (TWEXAFEGSMTHx), the term spread defined as the difference between the 10-year Treasury yield and the 3-month Treasury bill rate (TERM_SPREAD), the default spread defined as the difference between Moody's BAA and AAA corporate bond yields (DEFAULT_SPREAD), and the change in the CBOE Volatility Index (VIXCLSx) as a measure of implied equity market volatility. All macroeconomic variables are sourced from the FRED-MD database.

4.4 Methodology

This section describes the two-step framework used throughout Chapter 4. The first step is forecasting: generating one-month-ahead return predictions for each of the 19 Bloomberg hedge

²<http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>

fund indices using machine learning models. The second step is portfolio construction. Both steps are evaluated over the same out-of-sample period, January 2019 to March 2026, and performance is assessed using a set of metrics chosen to reflect the non-normal distributional features of hedge fund returns documented in Section 4.1.1.

4.4.1 Return Forecasting Framework

Forecasts are generated using both an expanding and a fixed 60-month rolling window, applied in parallel across all models. Under the expanding window, the estimation sample grows by one observation at each step, running from January 2014 through to the month immediately preceding the forecast. This accumulates an increasingly long history but assigns equal weight to observations from structurally different periods. The rolling window fixes the estimation sample at 60 months, dropping the oldest observation as each new one is added. It concentrates on recent return dynamics and adapts more quickly when factor-return relationships shift, at the cost of discarding older data that may still carry useful information.

At each step t , the model is re-estimated on the relevant estimation window up to $t - 1$ and used to produce a one-month-ahead forecast for each of the 19 indices. The underlying predictive regression takes the form

$$\hat{r}_{i,t+1} = \mathbf{x}'_t \boldsymbol{\beta}_i + \varepsilon_{i,t+1}, \quad (4.1)$$

where $\hat{r}_{i,t+1}$ is the one-month-ahead return forecast for index i , $\mathbf{x}'_t \in R^{1 \times 24}$ is a vector of lagged macroeconomic and financial predictors described in Section 4.3.2, $\boldsymbol{\beta}_i \in R^{24 \times 1}$ is the parameter vector for index i , and $\varepsilon_{i,t+1}$ is the forecast error. All predictors are lagged by one month to ensure no future information enters the estimation. Three machine learning models are used to estimate this relationship. Hyperparameters are selected once on the initial 60-month in-sample window using time-series cross-validation with three folds, minimising mean squared error, and are then used for the entire out-of-sample period. Fixing hyperparameters in this way prevents any form of look-ahead bias entering through the tuning process.

Random Forest

Random Forest [39] builds an ensemble of B regression trees, each trained on a different

bootstrap sample of the training data with a random subset of predictors considered at each split. Unlike linear models, it can capture nonlinear relationships between predictors and returns. The final forecast is the average across all trees

$$\hat{r}_{i,t+1}^{RF} = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}_t), \quad (4.2)$$

where $T_b(\mathbf{x}_t)$ is the prediction of the b -th tree. Averaging across many weakly correlated trees tends to reduce forecast variance, which is particularly valuable given the limited monthly observations available for hedge fund index returns.

Gradient Boosting

Gradient Boosting [40] constructs the ensemble sequentially. At each step m , a regression tree is fitted to the pseudo-residuals, defined as the negative gradient of the loss function evaluated at the current model predictions, and the model is updated

$$F_m(\mathbf{x}_t) = F_{m-1}(\mathbf{x}_t) + \eta \cdot h_m(\mathbf{x}_t), \quad (4.3)$$

where h_m is the tree fitted at step m and $\eta \in (0, 1]$ is the learning rate, which controls the contribution of each new tree to the ensemble. The sequential fitting process can increase sensitivity to noise in the training data, particularly in low signal-to-noise environments.

Support Vector Regression

Support Vector Regression [41] seeks a function that deviates from observed returns by no more than ε , while remaining as flat as possible. The primal optimisation problem is

$$\min_{\mathbf{w}, \xi, \xi^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{t=1}^T (\xi_t + \xi_t^*), \quad (4.4)$$

subject to $r_{i,t+1} - \mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}_t) \leq \varepsilon + \xi_t$ and $\mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}_t) - r_{i,t+1} \leq \varepsilon + \xi_t^*$, where C controls the penalty on deviations beyond the ε -tube and $\xi_t, \xi_t^* \geq 0$ are slack variables. A radial basis function kernel is employed throughout this work to capture nonlinear relationships between predictors and returns.

Penalised Linear Models

LASSO and ElasticNet were also evaluated, but are not reported in the main results. Cross-

validation consistently selected regularisation parameters that shrank most coefficients to zero, effectively reducing both models to near-constant forecasts. Relaxing the regularisation produced severe out-of-sample overfitting, with Theil's U values exceeding 10 for several indices. With limited monthly observations and 24 predictors, these models appear to sit in an unfavourable regime where neither tight shrinkage nor flexible estimation is reliable. The three nonlinear models do not face this constraint and are therefore the focus of the forecasting analysis.

4.4.2 Portfolio Construction Framework

Nine portfolio strategies are evaluated over the full out-of-sample period, January 2019 to March 2026, under both expanding and rolling window schemes. Given that the COVID-19 crisis falls within the evaluation period, the full-period results are supplemented by a subperiod analysis across three regimes: Pre-COVID (January 2019 – December 2019), COVID (January 2020 – December 2021), and Post-COVID (January 2022 – March 2026). For the subperiod analysis, the expanding window is the primary focus, as it produces a more stable long-run estimate of the covariance structure; rolling window results are reported in the full-period table for completeness.

The nine strategies differ in their objective functions. Minimum Volatility and Minimum CVaR minimise portfolio risk. Maximum Sharpe Ratio and Maximum Sortino Ratio maximise risk-adjusted returns using the forecasted returns as the expected return input. Hierarchical Risk Parity (HRP), discussed in detail in Chapter 2, allocates capital using inverse-variance weighting within a hierarchically clustered universe without requiring matrix inversion. The regime-switching strategy combines HRP and Minimum CVaR based on a market volatility signal. The NCO algorithm, introduced in Chapter 3, applies the same objective functions within the two-stage clustering framework of López de Prado [12]; three NCO variants are evaluated: NCO Minimum Volatility, NCO Minimum CVaR, and NCO Max Sharpe, where the latter uses Random Forest forecasts as inputs. All strategies impose weight bounds $(0, 0.5)$ and transaction costs $k = 0.001$ applied at each monthly rebalancing. The upper bound of 0.5 prevents any single index from dominating the portfolio, which is a particular concern given the small 18-index investable universe. Without this constraint, risk-minimising or forecast-based objectives could concentrate heavily in one index with unusually low volatility or a strong forecasted return in

a given window, undermining diversification. Capping weights at 0.5 ensures every strategy retains exposure to at least two indices while still allowing meaningful concentration where risk or return differences are genuine.

Minimum Volatility

The Minimum Volatility portfolio minimises portfolio variance subject to the weight constraints

$$\min_{\mathbf{w}} \mathbf{w}^\top \hat{\Sigma} \mathbf{w} \quad (4.5)$$

$$\text{subject to } \mathbf{w}^\top \mathbf{1} = 1, \quad 0 \leq w_i \leq 0.5,$$

where $\mathbf{w} \in R^{N \times 1}$ is the vector of portfolio weights, $\hat{\Sigma} \in R^{N \times N}$ is the sample covariance matrix of hedge fund index returns estimated over the in-sample window, and $\mathbf{1}$ is an $N \times 1$ vector of ones.

Minimum CVaR

Standard deviation is a symmetric risk measure that treats upside and downside deviations equally. For hedge fund returns, which exhibit negative skewness and fat left tails as documented in Section 4.1.1, this is a poor approximation of the risk an investor actually faces. The Minimum CVaR objective directly minimises tail loss. Following Rockafellar and Uryasev [86], CVaR at confidence level $\alpha = 0.95$ is computed as

$$\min_{\mathbf{w}, \gamma} \gamma + \frac{1}{T(1-\alpha)} \sum_{t=1}^T \max(-\mathbf{w}^\top \mathbf{r}_t - \gamma, 0) \quad (4.6)$$

$$\text{s.t. } \mathbf{w}^\top \mathbf{1} = 1, \quad 0 \leq w_i \leq 0.5,$$

where γ is the Value-at-Risk at level α and $\mathbf{r}_t \in R^{N \times 1}$ is the vector of realised returns at time t . At $\alpha = 0.95$, CVaR captures the expected loss in the worst 5% of monthly outcomes.

Maximum Sharpe Ratio

For strategies incorporating the return forecasts, the Maximum Sharpe Ratio objective is defined as

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top \hat{\boldsymbol{\mu}} - r_f}{\sqrt{\mathbf{w}^\top \hat{\Sigma} \mathbf{w}}} \quad (4.7)$$

$$\text{s.t. } \mathbf{w}^\top \mathbf{1} = 1, \quad 0 \leq w_i \leq 0.5,$$

where $\hat{\boldsymbol{\mu}} \in R^{N \times 1}$ is the vector of one-month-ahead return forecasts and r_f is the risk-free rate.

Maximum Sortino Ratio

The Sharpe ratio penalises all return deviations equally, whether the portfolio is doing better or worse than expected. For hedge fund returns that are negatively skewed, this symmetry is misleading. Investors usually care far more about large losses than large gains. The Sortino ratio [87] addresses this by replacing total volatility in the denominator with a measure of downside risk only. The portfolio is obtained by solving

$$\begin{aligned} \max_{\mathbf{w}} \quad & \frac{\mathbf{w}^\top \hat{\boldsymbol{\mu}} - r_f}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma}^- \mathbf{w}}} \\ \text{s.t.} \quad & \mathbf{w}^\top \mathbf{1} = 1, \quad 0 \leq w_i \leq 0.5, \end{aligned} \quad (4.8)$$

where the semi-variance matrix $\boldsymbol{\Sigma}^- \in R^{N \times N}$ captures only the months where returns fell below the risk-free rate:

$$\boldsymbol{\Sigma}^- = \frac{1}{T} \sum_{t=1}^T \min(0, \mathbf{r}_t - r_f \mathbf{1}) \min(0, \mathbf{r}_t - r_f \mathbf{1})^\top, \quad (4.9)$$

where $\mathbf{r}_t \in R^{N \times 1}$ is the vector of realised returns at time t and T is the number of months in the sample.

Regime-Switching Strategy

The regime-switching strategy combines HRP and Minimum CVaR within a simple volatility-based rule. The allocation at time t depends on the prevailing level of broad hedge fund market volatility:

$$\text{Strategy}_t = \begin{cases} \text{Min CVaR}, & \text{if } \hat{\sigma}_t^{\text{BHEDGE}} < 0.02 \\ \text{HRP}, & \text{if } \hat{\sigma}_t^{\text{BHEDGE}} \geq 0.02 \end{cases}, \quad (4.10)$$

where $\hat{\sigma}_t^{\text{BHEDGE}}$ is the 3-month rolling standard deviation of BHEDGE monthly returns computed at the end of month t . When market volatility is below the 2% monthly threshold, the mechanism allocates to Minimum CVaR to capture better risk-adjusted returns in calm conditions. When volatility exceeds the threshold, it switches to HRP, which provides broad diversification without relying on covariance matrix inversion.

The threshold and strategy pair were selected empirically by evaluating all combinations of candidate thresholds (1%, 1.25%, 1.5%, 1.75%, and 2%) and strategy pairs over the in-sample estimation period. HRP combined with Minimum CVaR at a 2% threshold produced the highest Sharpe ratio and was therefore retained. HRP was preferred as the high-volatility allocation because it avoids covariance matrix inversion and remains stable precisely when the covariance matrix is most ill-conditioned.

Transaction Costs

All strategies are evaluated net of transaction costs, applied at each monthly rebalancing as

$$r_{p,t+1}^{\text{net}} = r_{p,t+1} - k \sum_{i=1}^N |w_{i,t+1} - w_{i,t}^+|, \quad (4.11)$$

where $w_{i,t}^+$ is the pre-rebalancing weight of asset i after market moves and $k = 0.001$.

4.5 Forecasting Results

This section presents out-of-sample forecasting results for Random Forest, Gradient Boosting, and Support Vector Regression under both expanding and rolling window schemes. All models are evaluated against an AR(1) benchmark, which serves as the natural reference point in this setting. It coincides with a linear regression model using only the lagged hedge fund return as a predictor, capturing univariate return persistence without any factor information. A model that cannot outperform this benchmark offers no practical value as a forecasting tool [84].

Forecast accuracy is assessed using Theil's U statistic, defined as the ratio of the Mean Squared Forecast Error of the competing model to that of the AR(1) benchmark:

$$\text{Theil's } U = \frac{MSFE_i}{MSFE_{AR(1)}},$$

where $MSFE_i$ denotes the average squared forecast error over the out-of-sample period for model i and $MSFE_{AR(1)}$ is the corresponding value for the AR(1) benchmark. Values below unity indicate superior forecasting accuracy relative to the benchmark; values above unity indicate the opposite. Table 4.1 reports Theil's U statistics for all three models across the full out-of-sample

period under both window schemes, and Tables 4.2 and 4.3 report results for the expanding and rolling windows respectively across three subperiods: Pre-COVID (January 2019 – December 2019), COVID (January 2020 – December 2021), and Post-COVID (January 2022 – March 2026). Given that the Pre-COVID subperiod comprises only 12 monthly observations, results for this period should be interpreted with caution.

Overall Performance

Table 4.1 reveals a differentiated pattern of forecasting performance across model classes. Random Forest is the dominant specification, beating the AR(1) benchmark on 14 out of 19 indices under the expanding window and 18 out of 19 under the rolling window, with mean Theil's U of 0.9367 and 0.9071, respectively. SVR produces competitive results, with 15 out of 19 indices beating the benchmark under both window schemes and mean Theil's U of 0.9217 and 0.9170, respectively. Gradient Boosting lags considerably behind both models across all specifications. One possible explanation for the performance gap is that, in the noisy monthly hedge-fund-return setting, methods that place greater emphasis on variance control may have a structural advantage.

Subperiod analysis

The full-sample results mask substantial variation across subperiods, as shown in Tables 4.2 and 4.3. During the Pre-COVID period, Random Forest already demonstrated forecasting ability, beating the AR(1) benchmark on 11 out of 19 indices under the expanding window and 16 out of 19 under the rolling window, with a mean Theil's U of 0.965 and 0.956, respectively. SVR produces comparable results with 14 out of 19 indices beating the benchmark under both schemes. Gradient Boosting, however, already struggles in this period, with only 8 out of 19 indices beating the benchmark under the expanding window and just 5 out of 19 under rolling, with mean values of 1.217 and 1.260.

During the COVID period, predictability improves sharply across all models. Under the rolling window, Random Forest beats the AR(1) benchmark on all 19 indices with a mean Theil's U of 0.834, the strongest result across any model and period. SVR beats the benchmark on 18

Table 4.1: Theil's U statistics for Random Forest, Gradient Boosting, and Support Vector Regression under expanding and rolling window schemes. Values below unity indicate superior predictive accuracy relative to the AR(1) benchmark.

Index	Random Forest		Gradient Boosting		SVR	
	Expanding	Rolling	Expanding	Rolling	Expanding	Rolling
BHCRDT	0.6736	0.6370	0.7253	0.7184	0.6386	0.6218
BHCRMS	0.7361	0.6868	0.7950	0.7755	0.7988	0.7592
BHMARB	0.8438	0.8154	0.8550	0.8286	0.8144	0.7945
BHFIA	0.8454	0.8238	0.8362	0.8095	0.7986	0.7882
BHDIST	0.9075	0.8831	1.0078	1.0580	0.9816	1.0143
BHCBA	0.9080	0.8760	0.8952	0.8764	0.8578	0.8696
BHEVT	0.9162	0.8681	1.2791	1.1733	0.8607	0.8285
BHEDGE	0.9574	0.9325	1.2968	1.2766	0.9625	0.9671
BHEVMS	0.9647	0.9247	1.3006	1.2489	0.9167	0.8871
BHEQMS	0.9650	0.9464	1.3162	1.1898	0.9474	0.9518
BHRV	0.9726	0.9425	1.3175	1.2553	0.9101	0.9128
BHACT	0.9752	0.9502	0.9822	0.9864	1.0305	1.0208
BHEMDB	0.9779	0.9837	0.9599	1.0034	0.9613	0.9783
BHCOM	0.9949	0.9747	1.3412	1.3909	1.0268	1.0261
BHEQTY	1.0063	0.9825	1.3123	1.2877	0.9489	0.9584
BHEQLS	1.0181	0.9887	1.3438	1.3810	0.9765	0.9768
BHMACR	1.0196	0.9849	1.3459	1.3685	1.1044	1.1177
BHMMS	1.0526	1.0356	1.3528	1.4504	1.0069	0.9891
BHFEMN	1.0629	0.9986	1.3698	1.2151	0.9702	0.9618
Mean Theil's U	0.9367	0.9071	1.1386	1.1207	0.9217	0.9170
Indices < 1	14/19	18/19	7/19	6/19	15/19	15/19

out of 19 indices, and even Gradient Boosting, which struggles across the full sample, improves substantially to 16 out of 19. This convergence of all model classes toward strong predictability during the COVID period suggests that the 24 systematic factors captured unusually clear and persistent signals during the crisis. Extreme movements in credit spreads, volatility, and equity risk premia created strong and exploitable factor-return relationships that all models could detect.

The Post-COVID period tells the opposite story. Under both window schemes, Random Forest beats the benchmark on at most 1 of 19 indices, while GB and SVR beat the benchmark on 1 to 2 indices. The mean Theil's U values rise sharply and exceed 1 in all cases, indicating that all models are actively worse than the AR(1) benchmark in the post-crisis environment. This complete breakdown in the post-COVID period points to a structural shift in hedge fund return dynamics after 2022: the interest rate hiking cycle, elevated geopolitical uncertainty, and changing cross-asset correlations appear to have disrupted the factor-return relationships that

Table 4.2: Subperiod Statistical Evaluation — Theil's U Statistics by Model and Period (Expanding Window)

Index	Pre-COVID			COVID			Post-COVID		
	RF	GB	SVR	RF	GB	SVR	RF	GB	SVR
BHCRDT	0.9051	0.9931	0.6224	0.5935	0.6022	0.5595	1.1447	1.4609	1.1341
BHCRMS	0.9750	0.9796	0.7183	0.6372	0.6453	0.6641	1.1451	1.4342	1.4011
BHMARB	0.9658	1.0074	1.1273	0.8032	0.8202	0.7489	1.0469	1.0239	1.1290
BHFIA	0.8944	0.9850	0.7221	0.7584	0.7387	0.7165	1.0701	1.0742	1.0293
BHDIST	0.6987	1.0180	0.6126	0.8572	0.9322	0.8517	1.0990	1.1831	1.4173
BHCBA	0.9285	0.9284	0.7303	0.8685	0.8402	0.7889	1.0459	1.0868	1.1282
BHEVT	0.8745	1.3889	0.7455	0.8535	1.0127	0.7613	1.0852	1.9348	1.1393
BHEDGE	0.9218	1.0332	0.8874	0.8846	1.0886	0.8266	1.0704	1.5987	1.1748
BHEVMS	0.9455	1.5110	0.7994	0.8982	1.0425	0.8255	1.1101	1.8078	1.1341
BHEQMS	1.0779	0.8574	1.0562	0.8458	1.1322	0.7746	1.0628	1.5088	1.0927
BHRV	1.0046	1.1732	0.9228	0.9130	1.0839	0.7967	1.0640	1.7356	1.0950
BHACT	0.9257	0.8884	1.1407	0.9408	1.0218	0.9373	1.0407	0.9781	1.0927
BHEMDB	0.9276	0.9458	0.8106	0.9362	0.9338	0.9238	1.0397	0.9961	1.0316
BHCOM	1.0823	2.0074	0.9958	0.9262	0.8971	0.9457	1.0189	1.5564	1.0804
BHEQTY	1.0018	1.0101	0.9073	0.9377	1.3276	0.8395	1.0956	1.5314	1.0976
BHEQLS	1.0124	1.3619	0.9364	0.9423	1.1239	0.8443	1.1151	1.6182	1.1519
BHMACR	1.0555	1.6830	1.1240	0.9899	1.2512	1.0487	1.0310	1.3467	1.1334
BHMMS	1.1372	2.3603	1.1743	0.9962	0.9916	0.9786	1.0760	1.4719	0.9929
BHFEMN	1.0039	0.9893	0.9107	1.0012	1.2779	0.9076	1.1300	1.5209	1.0383
Mean Theil's U	0.9652	1.2169	0.8918	0.8728	0.9875	0.8284	1.0785	1.4141	1.1312
Indices < 1	11/19	8/19	14/19	18/19	9/19	18/19	0/19	2/19	1/19

underpinned earlier predictability, leaving systematic predictors with little predictive power.

Rolling versus Expanding Window

Across all three models, the rolling window produces lower mean Theil's U values than the expanding window for the majority of indices. The improvement is most pronounced for Random Forest, where the rolling window adds four additional indices that beat the benchmark in the full sample, reducing the mean Theil's U from 0.937 to 0.907. The subperiod results reinforce this pattern: under the rolling window, Random Forest beats the benchmark on all 19 indices during the COVID period, compared to 18 under the expanding window. An expanding window assigns equal weight to all historical observations regardless of how far back they lie, so as the out-of-sample period extends, earlier observations from structurally different regimes increasingly dilute the signal from more recent return dynamics. A fixed 60-month rolling window concentrates estimation on the most recent return environment, adapting more rapidly when factor-return relationships shift. The consistent advantage of the rolling window across

Table 4.3: Subperiod Statistical Evaluation — Theil's U Statistics by Model and Period (Rolling Window)

Index	Pre-COVID			COVID			Post-COVID		
	RF	GB	SVR	RF	GB	SVR	RF	GB	SVR
BHCRDT	0.8925	1.1303	0.6306	0.5628	0.5679	0.5374	1.0836	1.6371	1.1639
BHCRMS	0.9511	1.0695	0.7272	0.5943	0.5926	0.6274	1.0834	1.5846	1.3688
BHMARB	0.9840	1.0673	1.1593	0.7722	0.7744	0.7237	1.0375	1.1044	1.1486
BHFIA	0.8658	0.9022	0.7277	0.7254	0.7518	0.6946	1.0925	0.9566	1.0588
BHDIST	0.7068	1.0210	0.6205	0.8104	0.9985	0.8234	1.1200	1.2146	1.6118
BHCBA	0.9366	0.8991	0.7215	0.8374	0.8454	0.7702	1.0036	0.9836	1.2519
BHEVT	0.8991	1.4390	0.7441	0.7841	0.8302	0.7113	1.0855	2.0311	1.1595
BHEDGE	0.9324	1.2138	0.8715	0.8428	0.9460	0.8033	1.0681	1.6950	1.2311
BHEVMS	0.9646	1.5834	0.7908	0.8394	0.8829	0.7726	1.1085	2.0051	1.1644
BHEQMS	1.0621	0.9865	1.0329	0.8053	0.9923	0.7565	1.0652	1.3803	1.1215
BHRV	0.9995	1.2906	0.9300	0.8620	0.9785	0.7711	1.0704	1.7318	1.1570
BHACT	0.9366	0.8772	1.1403	0.9105	1.0079	0.9399	1.0040	1.0111	1.0624
BHEMDB	0.9270	0.9485	0.8354	0.9370	0.9625	0.9385	1.0534	1.0652	1.0507
BHCOM	1.0432	1.9668	0.9980	0.9008	0.8619	0.9575	1.0062	1.7210	1.0735
BHEQTY	0.9933	1.0737	0.8962	0.8895	1.1528	0.8176	1.1030	1.5854	1.1556
BHEQLS	0.9995	1.5529	0.9260	0.8961	0.9624	0.8236	1.1059	1.8853	1.1850
BHMACR	0.9976	1.6027	1.1249	0.9615	1.1267	1.0420	0.9969	1.4763	1.1624
BHMMS	1.0930	1.9991	1.1516	0.9582	0.9493	0.9589	1.0809	1.9395	0.9773
BHFEMN	0.9847	1.3157	0.9208	0.9505	0.9668	0.8897	1.0480	1.4387	1.0395
Mean Theil's U	0.9563	1.2600	0.8921	0.8337	0.9027	0.8084	1.0640	1.4972	1.1655
Indices < 1	16/19	5/19	14/19	19/19	16/19	18/19	1/19	2/19	1/19

models and subperiods suggests that hedge fund return dynamics are sufficiently non-stationary that the cost of discarding older observations is outweighed by the benefit of focusing on recent factor behaviour.

Predictability by Strategy Type

Under the rolling window scheme, forecasting performance varies systematically across hedge fund strategy types, revealing a clear gradient from the most to the least predictable. Credit, relative value, and merger arbitrage strategies (BHCRDT, BHCRMS, BHFIA, BHCBA, BHRV, and BHMARB) sit at the top of this gradient, with RF and SVR achieving substantial reductions in forecast error relative to the AR(1) benchmark — as large as 38% for BHCRDT. These strategies share a common characteristic: their returns are driven by persistent, identifiable exposures to credit spreads, fixed-income risk premia, and trend-following factors that the predictor set directly captures. Event driven and equity strategies (BHEVT, BHEVMS, BHDIST, BHEQMS, BHEQTY, BHEQLS, BHFEMN, BHACT, and BHEMDB) occupy the middle of

the gradient, where RF and SVR generate meaningful improvements while Gradient Boosting struggles, suggesting that variance control matters more than bias reduction in this low signal-to-noise environment. Macro strategies (BHMCCR, BHMMS, and BHCOM) sit at the bottom, with Theil's U values remaining close to or above unity across all models, reflecting the discretionary and top-down nature of their return drivers that systematic predictors cannot reliably capture.

Across the machine learning specifications evaluated, Random Forest and SVR emerge as the most reliable forecasting models, demonstrating greater robustness than Gradient Boosting across both window schemes. Predictability is strongest for strategies with persistent, rule-based return drivers, namely credit, relative value, and merger arbitrage, where the predictor set used is structurally well aligned with the underlying factor exposures. Macro strategies are largely resistant to systematic prediction, reflecting the broader, less observable forces that drive their returns.

Predictability is also not uniform through time. The COVID period represents a distinct regime of unusually strong forecastability across all models, while the post-COVID period sees a near-complete breakdown, pointing to a structural shift in the factor-return relationships underpinning hedge fund return dynamics after 2022.

Taken together, the results provide evidence that machine learning methods can extract economically meaningful information from a broad predictor set and improve upon traditional time-series benchmarks, with the degree of improvement dependent on both strategy type and market regime.

Penalised Linear Models

LASSO and ElasticNet were evaluated but are not reported because neither model produced satisfactory forecasting performance. Cross-validation consistently selected regularisation parameters that shrank most coefficients to zero across indices, effectively reducing the models to near-constant forecasts. Lowering the degree of regularisation increased model flexibility but led to substantial out-of-sample overfitting, with Theil's U values exceeding 10 for several indices.

These results suggest that penalised linear models struggle to extract stable predictive relationships from this dataset. The sample consists of only 147 monthly observations and

exhibits a high noise-to-signal ratio, making coefficient estimates highly sensitive to the choice of regularisation. In contrast, Random Forest generated meaningful forecasting improvements using the same set of 24 predictors. This finding indicates that the information contained in the predictors may be captured more effectively through nonlinear relationships and interaction effects than through a purely linear specification.

4.6 Portfolio Performance

This section evaluates the out-of-sample performance of nine portfolio strategies against two benchmarks: BHEDGE, the Bloomberg All Hedge Fund Index, which represents passive broad market exposure to hedge funds, and an equally weighted portfolio allocating uniformly across the 18 investable indices. The Maximum Sharpe Ratio and Maximum Sortino Ratio strategies are each evaluated under three return forecast inputs, including Random Forest, Gradient Boosting, and Support Vector Regression, to isolate the contribution of forecast quality to portfolio outcomes; all remaining strategies do not incorporate return forecasts. All strategies are assessed under both expanding and rolling window schemes across five metrics: annualised return, volatility, Sharpe ratio, Sortino ratio, and maximum drawdown. Maximum drawdown captures the largest peak-to-trough decline in cumulative portfolio value over the evaluation period and is defined as

$$\text{Max DD} = \min_t \left(\frac{V_t}{\max_{s \leq t} V_s} - 1 \right), \quad (4.12)$$

where V_t is the cumulative portfolio value at time t and $\max_{s \leq t} V_s$ is the running peak value up to time t ; a smaller absolute magnitude indicates greater resilience to sustained losses. The Sharpe ratio is the primary metric for comparison; the Sortino ratio and maximum drawdown serve as complementary measures of downside risk.

Table 4.4 reports full-period results. NCO Minimum Volatility delivers the strongest risk-adjusted performance under the expanding window, with a Sharpe ratio of 1.8064, a Sortino ratio of 3.18, and a maximum drawdown of 4.78% — less than half that of BHEDGE at 10.95% and the equally weighted benchmark at 10.70%.

NCO Minimum CVaR produces nearly identical risk-adjusted results, with a marginally

Table 4.4: Full-Period Portfolio Performance (January 2019 – March 2026).

Strategy	Expanding Window					Rolling Window				
	Return	Std	Sharpe	Sortino	Max DD	Return	Std	Sharpe	Sortino	Max DD
<i>Benchmarks</i>										
Equally Weighted	7.12%	6.17%	1.1549	1.7057	−10.70%	7.12%	6.17%	1.1549	1.7057	−10.70%
BHEDGE	7.13%	6.72%	1.0612	1.6333	−10.95%	7.13%	6.72%	1.0612	1.6333	−10.95%
<i>Standalone Strategies</i>										
Min Volatility	6.98%	4.11%	1.6968	2.3714	−7.77%	7.11%	4.08%	1.7415	2.4602	−7.83%
RF Max Sharpe	7.14%	5.18%	1.3789	2.0119	−9.55%	7.02%	5.38%	1.3054	1.9493	−9.50%
GB Max Sharpe	5.19%	5.10%	1.0192	1.5094	−8.29%	5.28%	7.08%	0.7455	0.9077	−15.14%
SVR Max Sharpe	5.54%	5.16%	1.0737	1.3928	−9.55%	5.92%	5.08%	1.1655	1.5517	−10.53%
Min CVaR	7.66%	4.79%	1.5996	2.1989	−9.36%	7.62%	4.76%	1.6001	2.1944	−9.33%
RF Max Sortino	6.92%	5.13%	1.3496	1.8925	−9.80%	7.83%	5.49%	1.4255	2.1658	−9.64%
GB Max Sortino	5.27%	5.24%	1.0076	1.3451	−10.30%	6.43%	5.37%	1.1979	1.653	−10.12%
SVR Max Sortino	5.92%	5.06%	1.1702	1.5166	−10.42%	6.86%	5.15%	1.3324	1.7655	−10.44%
HRP	6.53%	4.61%	1.4186	2.0436	−8.33%	6.49%	4.78%	1.3597	1.9015	−9.02%
Regime (HRP/Min CVaR)	8.03%	4.57%	1.7565	2.5798	−8.18%	7.83%	4.75%	1.6491	2.3370	−8.85%
<i>NCO Strategies</i>										
NCO Min Volatility	6.70%	3.71%	1.8064	3.1840	−4.78%	6.42%	4.20%	1.5288	2.2747	−7.41%
NCO Max Sharpe (RF)	6.82%	4.77%	1.4305	2.5855	−6.41%	7.41%	5.55%	1.3345	2.2551	−8.64%
NCO Min CVaR	6.47%	3.62%	1.7884	3.1264	−4.75%	6.74%	4.40%	1.5336	2.2513	−7.78%

lower drawdown of 4.75% and the lowest portfolio volatility in the full sample at 3.62%. For an investor whose primary concern is limiting downside losses, the Minimum CVaR specification is the more suitable choice: by directly minimising expected tail loss rather than total variance, it is structurally better aligned with the negative skewness of hedge fund returns. The regime portfolio achieves the highest annualised return at 8.03% under the expanding window, with a Sharpe ratio of 1.7565, closely competitive with NCO Minimum Volatility, and a Sortino ratio of 2.58. Among forecast-dependent strategies, RF-based portfolios deliver the strongest risk-adjusted performance, with RF Max Sharpe recording a Sharpe ratio of 1.3789 and RF Max Sortino recording a Sharpe ratio of 1.3496 under the expanding window. GB-based portfolios are the weakest across both objective functions and both window schemes. GB Max Sharpe records a Sharpe ratio of just 0.7455 under the rolling window, falling below both benchmarks, consistent with GB's poor forecasting accuracy in Section 4.4. SVR performs moderately, with Sharpe ratios ranging from 1.0737 to 1.3324 across specifications and window schemes.

The forecasting and portfolio optimisation steps respond differently to the choice of estimation window. In the forecasting step, the rolling window proved superior because it kept estimation focused on recent factor-return dynamics. For portfolio optimisation, the expanding window demonstrates a clear advantage for NCO and regime strategies — NCO Minimum Volatility records a Sharpe of 1.8064 under expanding against 1.5288 under rolling, and the

regime portfolio falls from 1.7565 to 1.6491, reflecting the benefit of accumulating a more stable long-run covariance structure. For subperiod analysis, the expanding window is the natural choice, as it continuously incorporates all available history and avoids severely under-identified covariance matrices in shorter subperiods. All subsequent subperiod analyses are therefore conducted under the expanding window.

Figure 4.1 plots the growth of \$1 invested in January 2019 under the outperforming strategies through March 2026. Before COVID, all strategies tracked closely. March 2020 marks the split in the chart: BHEDGE and the equally weighted portfolio fall sharply, while the NCO portfolios experience substantially shallower drawdowns. The benchmarks recover more strongly through 2021 and briefly retake the lead. From 2022 onwards, the regime portfolio pulls clear, ending the sample at approximately \$1.80.

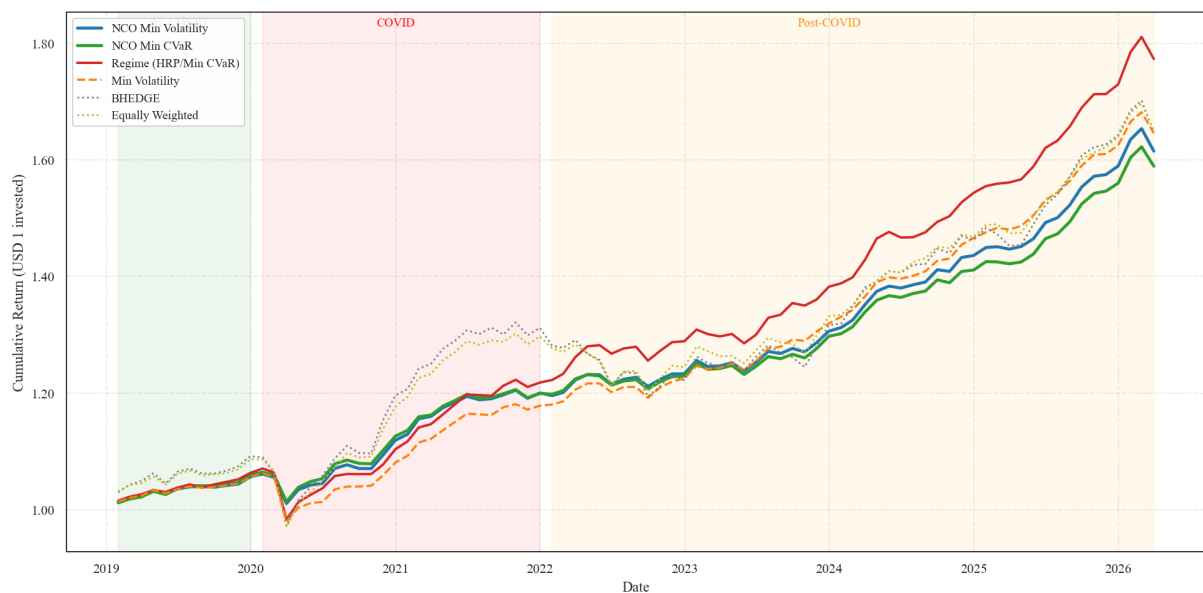


Figure 4.1: Cumulative portfolio performance under the expanding window scheme (January 2019 – March 2026). Shaded regions denote the Pre-COVID (January 2019 – December 2019), COVID (January 2020 – December 2021), and Post-COVID (January 2022 – March 2026) subperiods.

Subperiod Analysis

Table 4.5 reports out-of-sample performance across three subperiods: 12 months pre-COVID, 24 months during COVID, and 51 months post-COVID. Subperiod analysis focuses on RF-based forecast strategies as the dominant model. Given the limited pre-COVID sample, the COVID

and post-COVID subperiods provide a more informative basis for comparison.

Before COVID, most portfolios performed well, with Sharpe ratios mostly between 2 and 3.5 and maximum drawdowns well below both benchmarks. With only 12 monthly observations, little can be inferred beyond confirming that all strategies behave sensibly in benign conditions.

During COVID period, correlations between hedge fund indices increased sharply as assets that ordinarily provide diversification moved together under a dominant market risk factor. Figure 4.2 captures this directly: the condition number of the covariance matrix rose by approximately 55% relative to its pre-COVID mean, indicating a substantial deterioration in matrix conditioning during the crisis. A higher condition number indicates that the matrix is more sensitive to estimation error; the classical mean–variance optimiser, which inverts the covariance matrix directly, produces noisy, unstable weights under these conditions. NCO avoids this by restricting each optimisation to a cluster of internally correlated assets before aggregating across clusters, keeping instability local rather than letting it propagate through the full allocation. The NCO Minimum Volatility portfolio records the strongest risk-adjusted performance in the COVID period, with a Sharpe ratio of 1.3447 and a maximum drawdown of 4.78%, against 0.9098 and 7.77% for the mean–variance Minimum Volatility portfolio. NCO Minimum CVaR records a Sharpe of 1.3255 and a maximum drawdown of 4.75%, while mean–variance Minimum CVaR records 0.7133 and 9.36%. Both benchmarks, BHEDGE and the equally weighted portfolio, have drawdowns exceeding 10%.

It is worth noting that every strategy that structurally manages diversification records a Sharpe ratio above 1 during the COVID period: HRP (1.0255), the regime portfolio (1.0479), and all three NCO variants (1.3447, 1.1931, 1.3255). Every classical mean–variance portfolio falls below it. The performance gap reflects the importance of robust risk management: as systematic risk dominated returns and genuine diversification across hedge fund strategies compressed, portfolios that preserve diversification and manage risk by construction proved more resilient than those that rely on a favourable correlation structure as an assumption.

The condition number declines steadily in the post-COVID period as hedge fund strategy correlations normalise. The Federal Reserve tightening cycle introduced greater return dispersion across strategies, unwinding much of the near-multicollinearity that had built up during the

Table 4.5: Subperiod portfolio performance across market regimes (expanding window). All strategies impose weight bounds $(0, 0.5)$ and transaction costs $k = 0.001$. Returns and standard deviations are annualised. Max DD denotes maximum drawdown. Bold entries denote the best value in each column within each subperiod.

Strategy	Return	Std	Sharpe	Sortino	Max DD
<i>Panel A: Pre-COVID (January 2019 – December 2019)</i>					
Min Volatility	5.45%	1.77%	3.0795	9.3725	−0.53%
RF Max Sharpe	6.27%	2.12%	2.9554	11.5742	−0.44%
Min CVaR	6.35%	1.80%	3.5301	10.3652	−0.54%
RF Max Sortino	5.13%	2.05%	2.5046	5.6174	−0.73%
HRP	6.69%	2.32%	2.8873	8.2912	−0.74%
Regime (HRP/Min CVaR)	6.13%	1.78%	3.4399	10.0133	−0.54%
NCO Min Volatility	5.50%	1.88%	2.9314	8.6495	−0.62%
NCO Max Sharpe	8.13%	4.09%	1.9850	12.7098	−0.64%
NCO Min CVaR	5.78%	1.78%	3.2527	10.8976	−0.53%
BHEDGE	8.88%	4.05%	2.1905	4.8209	−1.66%
Equally Weighted	8.37%	3.90%	2.1489	5.4895	−1.31%
<i>Panel B: COVID (January 2020 – December 2021)</i>					
Min Volatility	5.71%	6.27%	0.9098	1.1103	−7.77%
RF Max Sharpe	7.03%	8.03%	0.8753	1.1225	−9.55%
Min CVaR	5.23%	7.34%	0.7133	0.8403	−9.36%
RF Max Sortino	5.04%	8.21%	0.6139	0.7399	−10.31%
HRP	7.12%	6.94%	1.0255	1.3237	−8.33%
Regime (HRP/Min CVaR)	7.06%	6.74%	1.0479	1.3082	−8.18%
NCO Min Volatility	6.52%	4.85%	1.3447	2.0444	−4.78%
NCO Max Sharpe	7.74%	6.48%	1.1931	1.8595	−6.41%
NCO Min CVaR	6.34%	4.78%	1.3255	1.9870	−4.75%
BHEDGE	9.66%	9.27%	1.0428	1.4876	−10.78%
Equally Weighted	9.30%	8.85%	1.0515	1.4098	−10.68%
<i>Panel C: Post-COVID (January 2022 – March 2026)</i>					
Min Volatility	7.94%	3.07%	2.5866	5.2806	−2.14%
RF Max Sharpe	7.40%	3.79%	1.9551	4.2968	−2.16%
Min CVaR	9.12%	3.55%	2.5655	5.9019	−2.09%
RF Max Sortino	5.42%	3.75%	1.4466	2.6469	−2.87%
HRP	6.22%	3.50%	1.7782	3.2412	−3.65%
Regime (HRP/Min CVaR)	8.94%	3.66%	2.4424	5.3979	−2.09%
NCO Min Volatility	7.07%	3.40%	2.0801	4.3209	−2.34%
NCO Max Sharpe	6.08%	3.86%	1.5738	3.1938	−2.74%
NCO Min CVaR	6.70%	3.29%	2.0357	4.2741	−2.05%
BHEDGE	5.53%	5.68%	0.9736	1.6050	−7.52%
Equally Weighted	5.80%	4.91%	1.1814	1.9662	−6.32%

crisis. Risk-minimising portfolios, Minimum Volatility and Minimum CVaR, perform strongly under both the mean–variance and NCO frameworks, with Sharpe ratios above 2 and maximum drawdowns below 2.5% for all four specifications. Notably, standalone Minimum Volatility and Minimum CVaR record Sharpe ratios of 2.5866 and 2.5655 respectively, outperforming their NCO counterparts at 2.0801 and 2.0357, suggesting that NCO’s diversification benefit is most pronounced during periods of market stress.



Figure 4.2: Condition Number of the Rolling 60-Month Sample Covariance Matrix (January 2019 – March 2026)

BHEDGE records a Sharpe of 0.97 and a drawdown of 7.52% over the same period. The regime portfolio records a Sharpe ratio of 2.44 and one of the highest annualised return of 8.94%; with BHEDGE volatility remaining below the 2% switching threshold throughout, the mechanism allocated primarily to Minimum CVaR. In the post-COVID period, portfolios incorporating RF forecasts underperform risk-minimising approaches, consistent with RF’s deteriorating forecast accuracy during this period, as documented in Section 4.4

NCO portfolios under risk-minimising objectives, namely NCO Minimum CVaR and NCO Minimum Volatility, are the most consistent performers across all three subperiods, outperforming both benchmarks in every period on risk-adjusted returns and maximum drawdown. NCO Minimum CVaR records Sharpe ratios of 3.2527, 1.3255, and 2.0357 and its maximum drawdown never exceeds 4.75%; for an investor whose primary concern is limiting losses across market cycles, this is the stronger case. NCO Minimum Volatility records Sharpe ratios of 2.9314, 1.3447, and 2.0801, with performance broadly comparable to NCO Minimum CVaR across all three subperiods. The regime-switching strategy records Sharpe ratios of 3.4399, 1.0479, and

2.4424 and pulls ahead on total return post-COVID, though the margin over NCO Minimum CVaR is modest and comes at the cost of weaker drawdown protection during the crisis. Overall, the NCO risk-minimising strategies deliver the most consistent risk-adjusted performance across varying market conditions.

4.7 Conclusion

This chapter contributes evidence on two interconnected questions in the hedge fund literature. The first concerns the conditions under which machine learning models generate economically meaningful return forecasts. The second concerns which portfolio construction approaches, spanning risk-minimising objectives, forecast-dependent strategies, and NCO-based allocation, deliver robust performance across distinct market regimes.

The central finding on forecasting is that predictability is regime-dependent rather than a stable property of hedge fund returns. Random Forest is the strongest model across periods, reflecting its robustness to the non-linear and noisy nature of hedge fund factor exposures. The COVID period represents a regime of unusually high predictability for all models. In the post-2022 period, predictability collapsed entirely. The post-2022 collapse in predictability, with no model beating the AR(1) benchmark on more than one index under either window scheme, points to a genuine shift in return dynamics following the 2022 tightening cycle. Sharp movements in the risk-free rate and term spread appear to have broken down the factor-return relationships that drove earlier predictability. This suggests that machine learning forecasting performance in this setting cannot be separated from the broader macro environment in which it operates.

On portfolio construction, NCO-based risk-minimising strategies deliver the strongest performance over the full sample. The source of that advantage is not uniform across regimes. During the COVID period, when correlations across hedge fund strategies converged sharply and classical optimisation degraded under compressed diversification, NCO's clustering mechanism proved most consequential: the performance gap between NCO and classical strategies was at its widest. Post-COVID, as cross-strategy correlations receded from their crisis levels, both NCO and classical risk-minimising strategies performed strongly relative to benchmarks and the differential narrowed considerably. The full-sample superiority of NCO is therefore real,

but it is concentrated in periods of market stress, where robust covariance estimation matters most. Forecast-dependent strategies are ultimately constrained by forecast quality; when RF predictability collapses after 2022, portfolio performance follows. The regime-switching strategy contributes a complementary dimension through dynamic adaptation to changing market conditions, leading on total return in the post-COVID period, while NCO strategies lead on drawdown protection during crisis.

The primary limitation of this analysis is the length of the available Bloomberg BHEDGE history, which constrains inference to a single broad market cycle and leaves the pre-COVID subperiod statistically underpowered. Extending the framework to individual fund-level returns, a longer sample spanning multiple rate cycles, or alternative hedge fund databases would allow stronger conclusions about the generalisability of the regime-dependent patterns documented here.

Conclusion

The central objective of this thesis has been to build an integrated framework for portfolio optimisation, one that addresses the principal sources of instability in mean–variance optimisation rather than treating them as separate problems. Three sources of instability stand out: noise in the covariance matrix, unreliable estimates of expected returns, and treating a heterogeneous asset universe as a single undifferentiated block. Covariance refinement, return forecasting, and allocation structure are usually studied separately. This thesis instead builds them as three connected stages of the same construction process, where each stage feeds into and conditions the one that follows.

The thesis is organised so that each empirical chapter builds directly on the one before it. Chapter 2 isolates the first source of instability by testing a wide range of covariance matrix estimators, linear and nonlinear shrinkage, Random Matrix Theory denoising and its quantile-based extension, DCC-GARCH, and hybrid combinations of these, within purely risk-based allocation strategies that depend solely on the covariance matrix. This isolates the effect of covariance refinement from any confounding influence of return estimation. Chapter 3 then builds on these findings by constructing an integrated framework that combines the covariance estimators identified in Chapter 2 with machine-learning forecasts of expected returns, allocated through Nested Clustered Optimisation (NCO). Because expected returns now enter the problem, this chapter moves from purely risk-based objectives to return-based ones, Maximum Quadratic Utility and Maximum Sharpe Ratio, that draw on both covariance and return information

simultaneously. The chapter takes the NCO framework a step further by testing it under two distinct clustering structures, López de Prado's Optimal Number of Clusters (ONC) algorithm and hierarchical clustering. Each clustering algorithm is run under different distance metrics: correlation-based, covariance-based, rank-based, and mutual-information-based. From there, the chapter brings together every covariance estimator, forecasting model, clustering algorithm, and distance metric within the same testing framework, and compares the results across every combination. Chapter 4 extends this integrated framework to a different asset class, hedge fund strategy indices, and tests whether the same conclusions hold outside equities and across distinct market regimes rather than in a single pooled sample.

Chapter 2 established that the value of covariance refinement is conditional rather than universal. Under monthly sampling, where a large asset universe is estimated from a comparatively short observation window, the sample covariance matrix is heavily contaminated by noise, and refinement techniques deliver a meaningful improvement over the unadjusted sample estimate. Under monthly S&P 500-scale data specifically, this improvement holds at the level of each estimator family, denoising, shrinkage, dynamic estimation. Under daily sampling, where the number of observations is large relative to the number of assets, the sample covariance matrix is already reasonably well estimated. Further adjustment tends in most cases to harm performance rather than help it, removing genuine signal along with noise. Only a handful of techniques, mainly nonlinear shrinkage, still deliver a modest improvement. This establishes a precise account of when refinement should be applied and when it should not, rather than treating refinement as beneficial by default.

Within the monthly setting, the chapter further shows that no single estimator dominates across objectives. DCC-NL is best suited to minimum volatility and maximum decorrelation, because it combines a dynamic, time-varying correlation structure with nonlinear shrinkage applied to the eigenvalues, capturing genuine shifts in correlation while still correcting for the eigenvalue distortion that a purely dynamic model would otherwise leave uncorrected. The quantile-based extension of the denoising framework, an original contribution of this thesis, delivers the strongest results under maximum diversification, where estimation error enters through the full covariance structure rather than through a single dynamic. Within that extension,

replacing noisy eigenvalues with a higher quantile rather than their mean tends to perform better, since the condition number of the matrix, $\kappa(C) = \lambda_{\max}/\lambda_{\min}$, is governed by the smallest eigenvalue: a higher substituted value raises $\lambda_{\min}$ further than the mean would, lowering the condition number by more and leaving the matrix better conditioned under inversion. Hierarchical Risk Parity (HRP) stands apart from this pattern entirely: because its allocation mechanism proceeds through hierarchical clustering rather than matrix inversion, it is largely indifferent to which covariance estimator feeds it. That indifference is exactly what recommends it as a fallback when covariance estimation cannot be trusted. Together, these results give the thesis a set of estimators to carry forward.

Chapter 3 builds its findings in layers, each resting on the one before it. Random Forest emerges as the strongest forecasting model. It captures nonlinear return dynamics that linear models miss, and averaging across many trees keeps its predictions stable. Incorporating these forecasts into mean–variance optimisation improves on a portfolio built from historical mean returns under the Maximum Quadratic Utility (MQ) objective, even before any clustering is applied. Under the Maximum Sharpe Ratio (MS) objective the historical-mean benchmark is a harder target, since it directly optimises the ratio being measured. Forecast-based portfolios paired with the raw covariance matrix alone do not consistently clear it under MS. Only once the covariance estimator is refined does the forecast-based portfolio pull ahead, even without clustering. The improvement is enhanced further once NCO is introduced, adding structural discipline to the allocation that the forecast alone does not provide. The strongest specification throughout combines Random Forest forecasts with hierarchical clustering and Spearman distance. Hierarchical clustering, built here using Ward’s linkage, forms compact, well-separated asset groups with genuine return co-movement, producing more stable clusters than the ONC algorithm achieves in the top specifications. Spearman distance reinforces this by capturing rank-based co-movement rather than linear correlation, making the clustering robust to the fat tails and extreme returns common in equity markets. This strongest specification is completed by a dynamic covariance estimator, DCC-NL under MQ and DCC combined with denoising under MS. Dynamic estimation matters here because asset correlations shift with market conditions rather than staying fixed, and a covariance matrix that adapts to this captures structure that a

static estimate cannot. Under both objectives, this specification delivers the highest Sharpe ratios anywhere in the thesis for monthly S&P 500 data. These two portfolios are also considerably more diversified than their mean–variance equivalents built from the same inputs, holding more active positions and a substantially lower maximum single-asset weight in both cases.

Chapter 4 tests whether the same layered logic holds on hedge fund indices, a different and harder asset class. Random Forest again emerges as the strongest forecasting model. It beats the AR(1) benchmark on 18 of 19 hedge fund indices under the rolling window. This predictive edge is not stable, it strengthens sharply during COVID, when Random Forest beats the benchmark on all 19 indices. Extreme, correlated moves in credit spreads, volatility, and equity risk premia during the crisis created clear, exploitable factor–return relationships that the model could detect. That signal disappears almost entirely after 2022, when Random Forest beats the benchmark on at most one index under either window scheme. The 2022 rate-hiking cycle disrupted the same factor–return relationships the model had relied on, leaving little for it to learn from. Forecast quality is therefore regime-dependent rather than a fixed property of the model. On portfolio construction, NCO again improves on the classical mean–variance equivalent. The improvement is concentrated specifically in the COVID period, when hedge fund correlations converged and the covariance matrix became poorly conditioned. Classical mean–variance optimisation inverts this matrix directly, so instability in the matrix passes straight into the portfolio weights. NCO instead restricts each optimisation to a cluster of internally correlated assets before aggregating across clusters, which keeps that instability local rather than letting it spread through the whole allocation. This is precisely why NCO’s advantage is largest exactly when it is needed most, during periods of market stress. Every strategy that structurally manages diversification outperforms every classical mean–variance portfolio during this period. Across all three subperiods, NCO’s risk-minimising variants are the most consistent performers. The regime-switching strategy pulls ahead on total return once markets calm.

The three empirical chapters establish a single connected argument. Refined covariance estimation stabilises risk, machine-learning forecasting recovers a genuine return signal, and Nested Clustered Optimisation converts both into diversified, resilient allocations. Dynamic, time-varying covariance estimators proved decisive once folded into the full framework: DCC-

NL and DCC combined with denoising deliver the thesis's highest Sharpe ratios in Chapter 3, evidence that correlations shift with market conditions and that estimators built to track this consistently outperform those that assume a fixed structure. Chapter 4 then showed where the framework earns its return: NCO's advantage over classical optimisation is largest exactly when correlations converge under market stress, the moment portfolio construction matters most. This thesis therefore delivers more than three separate improvements. It identifies, with evidence, when refinement pays off, which specification wins under which objective, and where the gains concentrate.

Finally, this research has limitations that point directly to future work. Chapter 4's hedge fund results are constrained by the limited availability of long-history hedge fund index data, which restricts how far the regime-dependent pattern documented there can be confirmed against a genuinely different market cycle. The interaction analysis carried out in Chapter 3, spanning ten covariance estimators, six distance metrics, and two clustering algorithms, is not repeated at the same depth in Chapter 4, which evaluates a fixed set of nine portfolio strategies rather than the full combinatorial space. Extending that interaction analysis to hedge fund indices, and testing the regime-dependent findings against a longer or independent hedge fund database, would strengthen confidence in how far the conditional patterns established in this thesis extend beyond the settings tested here.

Bibliography

1. Markowitz H. Portfolio Selection. *The Journal of Finance* 1952; 7:77–91
2. Ledoit O and Wolf M. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 2004; 88:365–411
3. Chopra VK and Ziemba WT. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management* 1993; 19:6–11
4. Black F and Litterman R. Asset Allocation: Combining Investor Views with Market Equilibrium. *Goldman Sachs Fixed Income Research* 1990; 115:7–18
5. Maillard S, Roncalli T, and Teïletche J. The Properties of Equally Weighted Risk Contributions Portfolios. *Journal of Portfolio Management* 2010; 36:60–70
6. Choueifaty Y. Towards Maximum Diversification. SSRN Working Paper 4063676. 2008
7. Laloux L, Cizeau P, Bouchaud JP, and Potters M. Noise Dressing of Financial Correlation Matrices. *Physical Review Letters* 1999; 83:1467–70
8. Ledoit O and Wolf M. Honey, I Shrunk the Sample Covariance Matrix. Tech. rep. 691. Universitat Pompeu Fabra, 2003
9. Ledoit O and Wolf M. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 2003; 10:603–21
10. Ledoit O and Wolf M. Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices. *Annals of Statistics* 2012; 40:1024–60
11. López de Prado M. Building Diversified Portfolios that Outperform Out-of-Sample. *Journal of Portfolio Management* 2016; 42:59–69

12. López de Prado MM. Machine Learning for Asset Managers. Cambridge University Press, 2020
13. Engle RF. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalised Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 2002; 20:339–50
14. Sharpe WF. Mutual Fund Performance. *The Journal of Business* 1966; 39:119–38
15. Stein C. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1956; 1:197–206
16. Vu Anh Tuan E. How Does the Ledoit and Wolf Shrinkage Estimator Improve a Real Estate Portfolio? *Journal of Real Estate Portfolio Management* 2013; 19:89–101
17. Chen Y, Wiesel A, Eldar YC, and Hero AO. Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing* 2010; 58:5016–29
18. Marchenko VA and Pastur LA. Distribution of Eigenvalues for Some Sets of Random Matrices. *Matematicheskii Sbornik* 1967; 114:507–36
19. Ledoit O and Wolf M. Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimensions. *Journal of Multivariate Analysis* 2015; 139:360–84
20. Ledoit O and Wolf M. Quadratic Shrinkage for Large Covariance Matrices. *Bernoulli* 2022; 28:1519–47
21. Lee MC, Chiou JS, and Lin CM. A Study of Value-at-Risk on Portfolio in Stock Return Using DCC Multivariate GARCH. *Applied Financial Economics Letters* 2006; 2:183–8
22. Engle RF and Sheppard K. Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. Working Paper 8554. National Bureau of Economic Research, 2001
23. Pakel C, Shephard N, Sheppard K, and Engle RF. Fitting Vast Dimensional Time-Varying Covariance Models. Tech. rep. SSRN Working Paper, 2014

24. Engle RF, Ledoit O, and Wolf M. Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics* 2019; 37:363–75
25. Timmermann A. Elusive return predictability. *International Journal of Forecasting* 2008; 24:1–18
26. Box GEP and Jenkins GM. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970
27. Engle RF. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 1982; 50:987–1007
28. Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 1986; 31:307–27
29. Campbell JY. Stock Returns and the Term Structure. *Journal of Financial Economics* 1987; 18:373–99
30. Fama EF and French KR. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 1988; 22:3–25
31. Fama EF and French KR. Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics* 1989; 25:23–49
32. Campbell JY and Shiller RJ. The Dividend–Price Ratio, Expectations of Future Dividends and Discount Factors. *Review of Financial Studies* 1988; 1:195–227
33. Goyal A and Welch I. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 2008; 21:1455–508
34. Campbell JY and Thompson SB. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 2008; 21:1509–31
35. Rapach DE, Strauss JK, and Zhou G. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies* 2010; 23:821–62
36. Hoerl AE and Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* 1970; 12:55–67

37. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B* 1996; 58:267–88
38. Breiman L, Friedman JH, Olshen RA, and Stone CJ. *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984
39. Breiman L. Random forests. *Machine Learning* 2001; 45:5–32
40. Friedman J. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 2001; 29:1189–232
41. Drucker H, Burges C, Kaufman L, Smola A, and Vapnik V. Support vector regression machines. *Advances in Neural Information Processing Systems*. Vol. 9. 1997 :155–61
42. Mantegna RN. Hierarchical Structure in Financial Markets. *The European Physical Journal B* 1999; 11:193–7
43. López de Prado M and Lewis M. Detection of false investment strategies using unsupervised learning methods. *Quantitative Finance* 2019; 19:1555–65
44. Rousseeuw P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 1987; 20:53–65
45. Tumminello M, Lillo F, and Mantegna R. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior and Organization* 2010; 75:40–58
46. Ward JH. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 1963; 58:236–44
47. Yue S, Wang J, and Wei L. Application of Two-Order Difference to Gap Statistic. 2009
48. McCracken M and Ng S. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 2016; 34:574–89
49. Neely CJ, Rapach DE, Tu J, and Zhou G. Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science* 2014; 60:1772–91
50. Ledoit O and Wolf M. Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management* 2004; 30:110–9

51. Ledoit O and Wolf M. Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices. *Annals of Statistics* 2022; 50:1247–78
52. Gu S, Kelly B, and Xiu D. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 2020; 33:2223–73
53. Bailey DH and López de Prado M. The Sharpe Ratio Efficient Frontier. *Journal of Risk* 2012; 15:3–44
54. Pearson K. Notes on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* 1895; 58:240–2
55. Spearman C. The Proof and Measurement of Association Between Two Things. *The American Journal of Psychology* 1904; 15:72–101
56. Kendall MG. A New Measure of Rank Correlation. *Biometrika* 1938; 30:81–93
57. Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal* 1948; 27:379–423
58. Meilă M. Comparing Clusterings—an Information Based Distance. *Journal of Multivariate Analysis* 2007; 98:873–95
59. Gerber S, Markowitz H, Ernst P, Miao Y, Javid B, and Sargen P. The Gerber Statistic: A Robust Co-Movement Measure for Portfolio Optimization. *The Journal of Portfolio Management* 2022; 48:87–102
60. Székely GJ, Rizzo ML, and Bakirov NK. Measuring and Testing Dependence by Correlation of Distances. *The Annals of Statistics* 2007; 35:2769–94
61. Evans JL and Archer SH. Diversification and the Reduction of Dispersion: An Empirical Analysis. *Journal of Finance* 1968; 23:761–7
62. Tang GY. How Efficient is Naive Portfolio Diversification? An Educational Note. *Omega* 2004; 32:155–60
63. Alexeev V and Dungey M. Equity Portfolio Diversification with High Frequency Data. *Quantitative Finance* 2015; 15:1205–15
64. Lhabitant F. *Handbook of Hedge Funds*. Chichester: John Wiley & Sons, 2006

65. Agarwal V and Naik NY. Risks and Portfolio Decisions Involving Hedge Funds. *Review of Financial Studies* 2004; 17:63–98
66. Stefanini F. *Investment Strategies of Hedge Funds*. Chichester: John Wiley & Sons, 2006
67. Mitchell M and Pulvino T. Characteristics of risk and return in risk arbitrage. *Journal of Finance* 2001; 56:2135–75
68. El Kalak I, Azevedo A, and Hudson R. Reviewing the Hedge Funds Literature II: Hedge Funds' Returns and Risk Management Characteristics. *International Review of Financial Analysis* 2016; 48:55–66
69. Liang B. On the Performance of Hedge Funds. *Financial Analysts Journal* 1999; 55:72–85
70. Patton AJ. Are “Market Neutral” Hedge Funds Really Market Neutral? *Review of Financial Studies* 2009; 22:2495–530
71. Liang B and Park H. Risk Measures for Hedge Funds: A Cross-Sectional Approach. *European Financial Management* 2007; 13:333–70
72. Getmansky M, Lo AW, and Makarov I. An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns. *Journal of Financial Economics* 2004; 74:529–609
73. Amenc N, El Bied S, and Martellini L. Predictability in Hedge Fund Returns. *Financial Analysts Journal* 2003; 59:32–46. DOI: 10.2469/faj.v59.n5.2562
74. Hamza O, Kooli M, and Roberge M. Further Evidence on Hedge Fund Return Predictability. *The Journal of Wealth Management* 2006; 9:68–79
75. Fung W and Hsieh DA. The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers. *Review of Financial Studies* 2001; 14:313–41
76. Fung W and Hsieh DA. Hedge Fund Benchmarks: A Risk-Based Approach. *Financial Analysts Journal* 2004; 60:65–80
77. Bali TG, Brown SJ, and Caglayan MO. Do Hedge Funds' Exposures to Risk Factors Predict Their Future Returns? *Journal of Financial Economics* 2011; 101:36–68
78. Titman S and Tiu C. Do the Best Hedge Funds Hedge? *Review of Financial Studies* 2011; 24:123–68

79. Bali TG, Brown SJ, and Caglayan MO. Systematic Risk and the Cross-Section of Hedge Fund Returns. *Journal of Financial Economics* 2012; 106:114–31
80. Sun Z, Wang A, and Zheng L. The Road Less Traveled: Strategy Distinctiveness and Hedge Fund Performance. *Review of Financial Studies* 2012; 25:96–143
81. Avramov D, Kosowski R, Naik NY, and Teo M. Hedge Funds, Managerial Skill, and Macroeconomic Variables. *Journal of Financial Economics* 2011; 99:672–92
82. Wegener C, Nitzsch R von, and Cengiz C. An Advanced Perspective on the Predictability in Hedge Fund Returns. *Journal of Banking and Finance* 2010; 34:2694–708
83. Avramov D, Barras L, and Kosowski R. Hedge Fund Return Predictability Under the Magnifying Glass. *Journal of Financial and Quantitative Analysis* 2013; 48:1057–83
84. Panopoulou E and Vrontos S. Hedge Fund Return Predictability; To Combine Forecasts or Combine Information? *Journal of Banking and Finance* 2015; 56:103–22
85. Wu W, Chen J, Yang Z, and Tindall ML. A Cross-Sectional Machine Learning Approach for Hedge Fund Return Prediction and Selection. *Management Science* 2021; 67:4577–601
86. Rockafellar RT and Uryasev S. Optimization of Conditional Value-at-Risk. *Journal of Risk* 2000; 2:21–42
87. Sortino F and Meer R van der. Downside risk. *Journal of Portfolio Management* 1991; 17:27–31

Appendix

Table 5.1: Technical indicators employed as predictor variables in the forecasting models, following Neely et al. (2014).

Variable	Category	Trading signal (S_t)
MA(1,9)	Moving Average	$S_t = 1$ if $MA_{1,t} \geq MA_{9,t}$, otherwise $S_t = 0$.
MA(1,12)	Moving Average	$S_t = 1$ if $MA_{1,t} \geq MA_{12,t}$, otherwise $S_t = 0$.
MA(2,9)	Moving Average	$S_t = 1$ if $MA_{2,t} \geq MA_{9,t}$, otherwise $S_t = 0$.
MA(2,12)	Moving Average	$S_t = 1$ if $MA_{2,t} \geq MA_{12,t}$, otherwise $S_t = 0$.
MA(3,9)	Moving Average	$S_t = 1$ if $MA_{3,t} \geq MA_{9,t}$, otherwise $S_t = 0$.
MA(3,12)	Moving Average	$S_t = 1$ if $MA_{3,t} \geq MA_{12,t}$, otherwise $S_t = 0$.
MOM(9)	Momentum	$S_t = 1$ if $P_t \geq P_{t-9}$, otherwise $S_t = 0$.
MOM(12)	Momentum	$S_t = 1$ if $P_t \geq P_{t-12}$, otherwise $S_t = 0$.
VOL(1,9)	Volume (OBV)	$S_t = 1$ if $MA_{1,t}^{OBV} \geq MA_{9,t}^{OBV}$, otherwise $S_t = 0$.
VOL(1,12)	Volume (OBV)	$S_t = 1$ if $MA_{1,t}^{OBV} \geq MA_{12,t}^{OBV}$, otherwise $S_t = 0$.
VOL(2,9)	Volume (OBV)	$S_t = 1$ if $MA_{2,t}^{OBV} \geq MA_{9,t}^{OBV}$, otherwise $S_t = 0$.
VOL(2,12)	Volume (OBV)	$S_t = 1$ if $MA_{2,t}^{OBV} \geq MA_{12,t}^{OBV}$, otherwise $S_t = 0$.
VOL(3,9)	Volume (OBV)	$S_t = 1$ if $MA_{3,t}^{OBV} \geq MA_{9,t}^{OBV}$, otherwise $S_t = 0$.
VOL(3,12)	Volume (OBV)	$S_t = 1$ if $MA_{3,t}^{OBV} \geq MA_{12,t}^{OBV}$, otherwise $S_t = 0$.

Notes. P_t denotes the stock closing price at time t . The j -period moving average of price is $MA_{j,t} = \frac{1}{j} \sum_{i=0}^{j-1} P_{t-i}$. The On-Balance Volume is $OBV_t = \sum_{k=1}^t VOL_k D_k$, where $D_k = 1$ if $P_k - P_{k-1} \geq 0$ and $D_k = -1$ otherwise. Its j -period moving average is $MA_{j,t}^{OBV} = \frac{1}{j} \sum_{i=0}^{j-1} OBV_{t-i}$. All signals are binary (0 or 1).

Table 5.2: Firm-level fundamental variables sourced from the WRDS Financial Ratio used in Chapter 3.

Variable	Definition
<i>Valuation</i>	
capei	Price divided by 10-year inflation-adjusted average earnings.
bm	Book value of common equity divided by market capitalisation.
evm	Enterprise value divided by EBITDA.
pe_op_basic	Closing price divided by basic EPS from operations.
pe_op_dil	Closing price divided by diluted EPS from operations.
pe_exi	Closing price divided by EPS excluding extraordinary items.
pe_inc	Closing price divided by EPS including extraordinary items.
ps	Market capitalisation divided by net sales.
pcf	Closing price divided by cash flow per share.
ptb	Closing price divided by book value per share.
peg_trailing	P/E ratio divided by 3-year trailing earnings growth rate.
divyield	Annual dividends per share divided by closing price.
<i>Profitability</i>	
dpr	Dividends per share divided by earnings per share.
npm	Net income divided by sales.
opmbd	Operating income before depreciation divided by sales.
opmad	Operating income after depreciation divided by sales.
gpm	Gross profit divided by sales.
ptpm	Pre-tax income divided by sales.
cfm	Net income plus depreciation, divided by sales.
roa	Net income divided by total assets.
roe	Net income divided by book equity.
roce	Earnings before interest and tax divided by capital employed.

Continued on next page.

Table 5.2 continued.

Variable	Definition
efftax	Income tax expense divided by pre-tax income.
aftret_eq	After-tax return on average common equity.
aftret_invcapx	After-tax return on invested capital, including extraordinary items.
aftret_equity	After-tax return on total stockholders' equity.
pretret_noa	Pre-tax return on net operating assets.
pretret_earnat	Pre-tax return on total earning assets.
gprof	Gross profit divided by total assets.
<i>Capitalization</i>	
equity_invcap	Common equity divided by invested capital.
debt_invcap	Long-term debt divided by invested capital.
totdebt_invcap	Total debt divided by invested capital.
capital_ratio	Long-term debt divided by long-term debt plus common equity.
<i>Solvency</i>	
int_debt	Interest expense divided by average long-term debt.
int_totdebt	Interest expense divided by average total debt.
cash_lt	Cash and short-term investments divided by long-term debt.
debt_at	Total debt divided by total assets.
debt_ebitda	Total debt divided by EBITDA.
short_debt	Short-term debt divided by total debt.
curr_debt	Current liabilities divided by total liabilities.
lt_debt	Long-term debt divided by total liabilities.
lt_ppent	Total liabilities divided by net property, plant, and equipment.
dltt_be	Long-term debt divided by book equity.
debt_assets	Total debt divided by total assets.
debt_capital	Total debt divided by total debt plus equity.

Continued on next page.

Table 5.2 continued.

Variable	Definition
de_ratio	Total liabilities divided by book equity.
intcov	EBIT divided by interest expense.
intcov_ratio	EBIT plus depreciation, divided by interest expense.
<i>Liquidity</i>	
cash_ratio	Cash and short-term investments divided by current liabilities.
quick_ratio	Cash, short-term investments and receivables divided by current liabilities.
curr_ratio	Current assets divided by current liabilities.
cash_conversion	Days receivables plus days inventory, minus days payables.
profit_lct	Operating income divided by current liabilities.
ocf_lct	Operating cash flow divided by current liabilities.
cash_debt	Operating cash flow divided by total debt.
fcf_ocf	Free cash flow divided by operating cash flow.
inv_tact	Inventory divided by current assets.
rec_tact	Receivables divided by current assets.
<i>Efficiency</i>	
inv_turn	Cost of goods sold divided by average inventory.
at_turn	Sales divided by average total assets.
rec_turn	Sales divided by average receivables.
pay_turn	Cost of goods sold plus change in inventory, divided by average payables.
sale_invcap	Sales divided by invested capital.
sale_equity	Sales divided by book equity.
sale_nwc	Sales divided by net working capital.
rd_sale	R&D expense divided by sales.
adv_sale	Advertising expense divided by sales.

Continued on next page.

Table 5.2 continued.

Variable	Definition
staff_sale	Staff expense divided by sales.
<i>Other</i>	
accrual	Net income minus operating cash flow, divided by total assets.

Table 5.3: Macroeconomic variables from FRED-MD used as predictors in Chapter 3.

Category	Number of Variables
Output and Income	19
Labour Market	28
Housing	10
Orders and Inventories	6
Money and Credit	10
Interest and Exchange Rates	22
Prices	23
Stock Market	8
Total	126

Notes. All variables are sourced from the FRED-MD database (McCracken and Ng, 2016). Each variable is transformed according to the transformation code recommended by FRED-MD to ensure stationarity.

Table 5.4: Out-of-sample portfolio performance under Random Forest forecasts — Mean–Variance optimisation without clustering, MQ Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	24.95	29.95	0.8329
Denoising (Mean)	25.30	30.04	0.8422
Denoising (75th)	25.33	30.06	0.8426
Denoising (90th)	25.35	30.10	0.8422
Shrinkage	24.66	31.47	0.7835
NL Shrinkage	24.34	30.59	0.7956
DCC	22.85	27.95	0.8175
DCC Denoising	23.00	28.14	0.8174
DCC Shrinkage	22.88	29.90	0.7651
DCC-NL	25.66	28.91	0.8874

Table 5.5: Out-of-sample portfolio performance under LASSO Regression forecasts — Mean–Variance optimisation without clustering, MQ Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	26.93	36.34	0.7412
Denoising (Mean)	26.91	36.40	0.7395
Denoising (75th)	26.73	36.39	0.7345
Denoising (90th)	26.57	36.28	0.7324
Shrinkage	27.81	37.34	0.7449
NL Shrinkage	27.79	37.15	0.7480
DCC	25.22	33.16	0.7604
DCC Denoising	25.24	33.16	0.7610
DCC Shrinkage	27.28	34.72	0.7859
DCC-NL	24.26	33.55	0.7231

Table 5.6: Out-of-sample portfolio performance under Gradient Boosting forecasts — Mean–Variance optimisation without clustering, MQ Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	28.07	33.16	0.8465
Denoising (Mean)	28.25	33.13	0.8527
Denoising (75th)	28.47	33.18	0.8581
Denoising (90th)	28.91	33.12	0.8728
Shrinkage	29.12	34.27	0.8495
NL Shrinkage	26.27	33.79	0.7774
DCC	25.87	31.64	0.8177
DCC Denoising	26.01	31.39	0.8284
DCC Shrinkage	27.11	32.99	0.8219
DCC-NL	26.61	32.70	0.8139

Table 5.7: Out-of-sample portfolio performance under Support Vector Regression forecasts — Mean–Variance optimisation without clustering, MQ Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	27.26	31.98	0.8524
Denoising (Mean)	27.15	32.14	0.8446
Denoising (75th)	27.03	32.14	0.8411
Denoising (90th)	28.20	32.36	0.8713
Shrinkage	29.57	34.60	0.8545
NL Shrinkage	28.41	33.79	0.8407
DCC	24.22	30.28	0.7999
DCC Denoising	24.98	30.45	0.8204
DCC Shrinkage	26.76	32.98	0.8113
DCC-NL	26.98	31.55	0.8553

Table 5.8: Out-of-sample portfolio performance under Elastic Net Regression forecasts — Mean–Variance optimisation without clustering, MQ Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.24	31.26	0.7113
Denoising (Mean)	21.89	31.01	0.7059
Denoising (75th)	22.03	31.11	0.7082
Denoising (90th)	23.87	32.41	0.7364
Shrinkage	24.35	34.61	0.7036
NL Shrinkage	23.49	33.14	0.7087
DCC	18.27	28.90	0.6320
DCC Denoising	18.15	28.62	0.6342
DCC Shrinkage	19.37	31.64	0.6121
DCC-NL	19.86	28.73	0.6913

Table 5.9: Out-of-sample portfolio performance under Random Forest forecasts — Mean–Variance optimisation without clustering, MS Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	13.75	14.58	0.9426
Denoising (Mean)	14.73	15.07	0.9776
Denoising (75th)	14.72	15.04	0.9788
Denoising (90th)	14.80	13.93	1.0626
Shrinkage	16.08	14.91	1.0780
NL Shrinkage	15.54	15.53	1.0008
DCC	14.32	13.15	1.0884
DCC Denoising	14.07	13.12	1.0719
DCC Shrinkage	15.58	13.97	1.1157
DCC-NL	14.85	13.21	1.1244

Table 5.10: Out-of-sample portfolio performance under LASSO Regression forecasts — Mean–Variance optimisation without clustering, MS Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	16.14	19.50	0.8277
Denoising (Mean)	17.85	20.53	0.8695
Denoising (75th)	17.77	20.49	0.8669
Denoising (90th)	19.02	18.44	1.0314
Shrinkage	18.79	20.19	0.9307
NL Shrinkage	17.44	20.96	0.8318
DCC	15.92	17.13	0.9294
DCC Denoising	17.16	17.63	0.9733
DCC Shrinkage	18.61	18.05	1.0312
DCC-NL	16.70	17.34	0.9636

Table 5.11: Out-of-sample portfolio performance under Gradient Boosting forecasts — Mean–Variance optimisation without clustering, MS Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.61	16.03	0.9737
Denoising (Mean)	16.27	16.91	0.9623
Denoising (75th)	15.44	16.36	0.9435
Denoising (90th)	15.53	15.16	1.0243
Shrinkage	16.68	16.52	1.0091
NL Shrinkage	16.27	16.91	0.9624
DCC	13.35	14.80	0.9015
DCC Denoising	13.68	14.87	0.9202
DCC Shrinkage	15.54	15.32	1.0147
DCC-NL	13.40	15.00	0.8936

Table 5.12: Out-of-sample portfolio performance under Support Vector Regression forecasts — Mean–Variance optimisation without clustering, MS Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	11.98	13.74	0.8718
Denoising (Mean)	12.41	13.67	0.9075
Denoising (75th)	12.33	13.64	0.9046
Denoising (90th)	12.72	12.91	0.9854
Shrinkage	14.25	13.81	1.0324
NL Shrinkage	13.31	14.00	0.9505
DCC	11.05	13.27	0.8326
DCC Denoising	11.00	13.20	0.8335
DCC Shrinkage	13.64	13.47	1.0124
DCC-NL	10.88	13.48	0.8069

Table 5.13: Out-of-sample portfolio performance under Elastic Net Regression forecasts — Mean–Variance optimisation without clustering, MS Objective.

Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.21	19.95	0.7627
Denoising (Mean)	13.70	16.24	0.8432
Denoising (75th)	13.74	16.22	0.8474
Denoising (90th)	13.79	15.64	0.8818
Shrinkage	14.22	17.50	0.8127
NL Shrinkage	14.04	17.60	0.7980
DCC	13.34	13.96	0.9558
DCC Denoising	12.80	13.87	0.9229
DCC Shrinkage	13.72	13.99	0.9807
DCC-NL	12.24	13.98	0.8759

Table 5.14: **Out-of-sample portfolio performance under Support Vector Regression forecasts (ONC Algorithm) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.82	24.78	0.9208	Raw Covariance	22.76	24.99	0.9105
Denoising (Mean)	23.44	24.68	0.9498	Denoising (Mean)	23.19	24.92	0.9306
Denoising (75th)	23.67	24.95	0.9488	Denoising (75th)	23.28	25.76	0.9036
Denoising (90th)	26.83	25.75	1.0423	Denoising (90th)	24.07	25.08	0.9595
Shrinkage	25.89	26.97	0.9599	Shrinkage	24.21	26.34	0.9192
NL Shrinkage	26.24	26.39	0.9943	NL Shrinkage	23.46	25.88	0.9065
DCC	22.59	23.80	0.9491	DCC	21.80	23.53	0.9263
DCC Denoising	24.43	23.58	1.0360	DCC Denoising	22.39	23.62	0.9479
DCC Shrinkage	24.41	26.20	0.9316	DCC Shrinkage	23.97	25.02	0.9582
DCC-NL	24.28	24.69	0.9833	DCC-NL	22.75	24.15	0.9419
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.15	25.26	0.8770	Raw Covariance	21.34	25.24	0.8455
Denoising (Mean)	21.91	25.18	0.8703	Denoising (Mean)	21.82	25.20	0.8660
Denoising (75th)	21.74	25.22	0.8622	Denoising (75th)	21.75	25.17	0.8639
Denoising (90th)	22.61	25.81	0.8759	Denoising (90th)	22.25	25.38	0.8768
Shrinkage	23.52	26.84	0.8765	Shrinkage	22.26	26.53	0.8388
NL Shrinkage	22.87	26.25	0.8713	NL Shrinkage	21.78	26.15	0.8329
DCC	21.13	23.60	0.8952	DCC	19.97	23.65	0.8447
DCC Denoising	21.76	23.82	0.9133	DCC Denoising	20.38	23.72	0.8593
DCC Shrinkage	22.83	25.52	0.8945	DCC Shrinkage	21.97	25.19	0.8724
DCC-NL	22.63	24.26	0.9326	DCC-NL	21.07	24.19	0.8708
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.93	25.69	0.8927	Raw Covariance	21.79	25.01	0.8714
Denoising (Mean)	23.35	25.79	0.9054	Denoising (Mean)	21.64	25.00	0.8656
Denoising (75th)	23.28	25.76	0.9036	Denoising (75th)	21.64	24.98	0.8664
Denoising (90th)	24.20	26.24	0.9223	Denoising (90th)	22.57	25.34	0.8908
Shrinkage	24.21	27.37	0.8846	Shrinkage	23.14	26.50	0.8729
NL Shrinkage	23.34	26.60	0.8775	NL Shrinkage	22.45	25.80	0.8700
DCC	21.17	23.93	0.8849	DCC	21.40	23.68	0.9038
DCC Denoising	21.75	24.16	0.9004	DCC Denoising	21.71	23.73	0.9146
DCC Shrinkage	23.17	25.96	0.8922	DCC Shrinkage	23.49	25.46	0.9226
DCC-NL	23.14	24.77	0.9343	DCC-NL	22.14	23.93	0.9249

Table 5.15: Out-of-sample portfolio performance under Support Vector Regression forecasts (Hierarchical Clustering) - MQ Objective — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	23.72	25.58	0.9270	Raw Covariance	24.65	25.45	0.9687
Denoising (Mean)	22.36	26.01	0.8594	Denoising (Mean)	24.35	25.67	0.9484
Denoising (75th)	20.68	26.74	0.7733	Denoising (75th)	24.30	25.66	0.9468
Denoising (90th)	22.75	26.58	0.8558	Denoising (90th)	24.80	25.99	0.9541
Shrinkage	27.12	27.70	0.9789	Shrinkage	26.16	26.88	0.9733
NL Shrinkage	22.18	26.43	0.8393	NL Shrinkage	24.46	25.69	0.9521
DCC	22.92	24.45	0.9375	DCC	19.92	23.82	0.8362
DCC Denoising	23.24	24.65	0.9430	DCC Denoising	19.87	24.10	0.8245
DCC Shrinkage	23.71	26.49	0.8951	DCC Shrinkage	23.25	25.53	0.9109
DCC-NL	26.37	24.88	1.0597	DCC-NL	22.25	24.59	0.9046
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	23.06	26.13	0.8825	Raw Covariance	24.08	25.21	0.9551
Denoising (Mean)	23.67	25.42	0.9309	Denoising (Mean)	24.47	25.56	0.9574
Denoising (75th)	23.56	25.44	0.9262	Denoising (75th)	24.21	25.56	0.9472
Denoising (90th)	24.04	26.23	0.9166	Denoising (90th)	25.01	26.16	0.9563
Shrinkage	25.00	27.16	0.9203	Shrinkage	25.36	27.15	0.9339
NL Shrinkage	24.22	26.27	0.9220	NL Shrinkage	24.42	26.17	0.9333
DCC	21.44	24.79	0.8648	DCC	21.63	23.50	0.9206
DCC Denoising	21.46	24.98	0.8590	DCC Denoising	21.88	28.96	0.9237
DCC Shrinkage	23.57	26.54	0.8882	DCC Shrinkage	22.63	25.56	0.8852
DCC-NL	22.99	25.67	0.8955	DCC-NL	23.29	24.46	0.9519
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	24.53	25.99	0.9438	Raw Covariance	24.53	25.94	0.9457
Denoising (Mean)	25.60	26.37	0.9707	Denoising (Mean)	24.82	26.33	0.9425
Denoising (75th)	25.67	26.40	0.9725	Denoising (75th)	24.70	26.30	0.9390
Denoising (90th)	26.28	26.88	0.9780	Denoising (90th)	25.81	26.76	0.9646
Shrinkage	27.29	27.79	0.9819	Shrinkage	27.16	27.70	0.9806
NL Shrinkage	26.15	27.01	0.9682	NL Shrinkage	25.97	26.94	0.9641
DCC	21.47	24.19	0.8875	DCC	20.58	24.94	0.8251
DCC Denoising	21.20	24.42	0.8683	DCC Denoising	20.58	24.94	0.8251
DCC Shrinkage	24.07	26.11	0.9217	DCC Shrinkage	23.58	26.32	0.8959
DCC-NL	21.96	25.25	0.8696	DCC-NL	22.22	25.85	0.8595

Table 5.16: **Out-of-sample portfolio performance under Elastic Net forecasts (ONC Algorithm) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	18.56	24.06	0.7717	Raw Covariance	19.95	24.59	0.8113
Denoising (Mean)	18.82	24.14	0.7793	Denoising (Mean)	20.05	24.57	0.8160
Denoising (75th)	18.42	23.92	0.7702	Denoising (75th)	20.16	24.54	0.8216
Denoising (90th)	20.82	24.46	0.8509	Denoising (90th)	21.64	25.29	0.8554
Shrinkage	22.03	26.85	0.8208	Shrinkage	22.46	27.14	0.8274
NL Shrinkage	21.56	26.08	0.8268	NL Shrinkage	21.21	26.03	0.8147
DCC	17.53	22.90	0.7654	DCC	17.62	22.67	0.7772
DCC Denoising	17.65	22.81	0.7738	DCC Denoising	17.16	22.78	0.7531
DCC Shrinkage	18.86	25.10	0.7514	DCC Shrinkage	18.51	24.98	0.7412
DCC-NL	17.05	23.02	0.7405	DCC-NL	19.27	22.83	0.8442
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.66	24.14	0.9390	Raw Covariance	21.20	27.44	0.7726
Denoising (Mean)	22.22	24.20	0.9183	Denoising (Mean)	21.33	27.36	0.7797
Denoising (75th)	22.50	24.32	0.9253	Denoising (75th)	21.51	27.37	0.7860
Denoising (90th)	24.58	25.76	0.9541	Denoising (90th)	22.88	28.12	0.8135
Shrinkage	25.48	27.85	0.9150	Shrinkage	23.62	29.76	0.7937
NL Shrinkage	23.94	26.00	0.9208	NL Shrinkage	22.53	28.79	0.7824
DCC	20.78	22.46	0.9250	DCC	17.10	22.87	0.7474
DCC Denoising	20.05	22.56	0.8887	DCC Denoising	16.63	22.91	0.7259
DCC Shrinkage	29.43	33.31	0.8836	DCC Shrinkage	18.45	25.58	0.7211
DCC-NL	22.49	23.12	0.9726	DCC-NL	18.78	23.15	0.8114
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	18.94	23.87	0.7936	Raw Covariance	20.44	25.04	0.8160
Denoising (Mean)	19.27	24.02	0.8019	Denoising (Mean)	20.61	25.13	0.8200
Denoising (75th)	19.34	24.01	0.8053	Denoising (75th)	20.74	25.21	0.8224
Denoising (90th)	20.88	25.03	0.8339	Denoising (90th)	22.05	26.16	0.8426
Shrinkage	22.13	26.26	0.8428	Shrinkage	22.89	27.90	0.8204
NL Shrinkage	20.23	24.97	0.8103	NL Shrinkage	22.13	26.77	0.8268
DCC	17.75	22.97	0.7727	DCC	19.75	23.55	0.8386
DCC Denoising	17.26	23.01	0.7500	DCC Denoising	19.78	23.57	0.8394
DCC Shrinkage	18.89	24.67	0.7658	DCC Shrinkage	20.55	25.90	0.7934
DCC-NL	19.28	23.12	0.8341	DCC-NL	21.31	23.24	0.9171

Table 5.17: **Out-of-sample portfolio performance under Elastic Net forecasts (Hierarchical Clustering) - MQ Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	22.16	24.00	0.9231	Raw Covariance	19.32	23.94	0.8069
Denoising (Mean)	22.35	25.34	0.8818	Denoising (Mean)	19.55	24.35	0.8029
Denoising (75th)	20.16	24.62	0.8190	Denoising (75th)	19.65	24.40	0.8055
Denoising (90th)	22.66	26.34	0.8605	Denoising (90th)	19.96	25.36	0.7869
Shrinkage	26.06	31.88	0.8175	Shrinkage	21.03	26.64	0.7895
NL Shrinkage	24.11	27.93	0.8630	NL Shrinkage	20.44	25.06	0.8155
DCC	19.48	21.52	0.9050	DCC	15.88	21.78	0.7290
DCC Denoising	19.34	21.69	0.8918	DCC Denoising	15.64	21.95	0.7126
DCC Shrinkage	15.88	21.78	0.7290	DCC Shrinkage	17.26	23.89	0.7228
DCC-NL	19.12	23.27	0.8216	DCC-NL	16.94	22.03	0.7688
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	21.51	21.51	0.8505	Raw Covariance	22.32	24.12	0.9255
Denoising (Mean)	21.77	25.53	0.8529	Denoising (Mean)	21.72	24.56	0.8843
Denoising (75th)	21.77	25.67	0.8479	Denoising (75th)	21.84	24.73	0.8831
Denoising (90th)	22.98	26.87	0.8553	Denoising (90th)	21.30	25.88	0.8229
Shrinkage	23.80	28.39	0.8383	Shrinkage	22.47	27.01	0.8321
NL Shrinkage	22.28	26.68	0.8351	NL Shrinkage	22.69	25.38	0.8941
DCC	17.23	22.36	0.7708	DCC	17.35	22.02	0.7879
DCC Denoising	16.63	22.42	0.7416	DCC Denoising	16.86	22.24	0.7580
DCC Shrinkage	18.63	25.11	0.7420	DCC Shrinkage	18.59	24.66	0.7539
DCC-NL	17.84	22.46	0.7944	DCC-NL	18.58	22.20	0.8368
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	20.75	25.25	0.8217	Raw Covariance	19.76	25.24	0.7830
Denoising (Mean)	20.58	25.30	0.8132	Denoising (Mean)	18.81	25.71	0.7317
Denoising (75th)	20.65	25.35	0.8145	Denoising (75th)	18.98	25.79	0.7360
Denoising (90th)	20.64	26.32	0.7842	Denoising (90th)	19.60	26.48	0.7403
Shrinkage	21.56	27.88	0.7773	Shrinkage	21.37	27.77	0.7695
NL Shrinkage	21.73	26.57	0.8179	NL Shrinkage	20.32	26.46	0.7681
DCC	15.87	22.65	0.7008	DCC	16.64	23.15	0.7186
DCC Denoising	15.28	22.90	0.6675	DCC Denoising	16.60	23.51	0.7064
DCC Shrinkage	18.11	25.46	0.7113	DCC Shrinkage	17.20	25.62	0.6713
DCC-NL	17.41	22.89	0.7603	DCC-NL	17.25	23.39	0.7377

Table 5.18: **Out-of-sample portfolio performance under Support Vector Regression forecasts (ONC Algorithm) - MS Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	14.44	14.09	1.0246	Raw Covariance	14.48	14.03	1.0322
Denoising (Mean)	14.05	14.25	0.9861	Denoising (Mean)	14.83	14.01	1.0581
Denoising (75th)	14.45	14.37	1.0060	Denoising (75th)	14.73	14.07	1.0466
Denoising (90th)	14.16	13.64	1.0385	Denoising (90th)	14.16	13.34	1.0610
Shrinkage	16.62	16.45	1.0102	Shrinkage	15.31	14.39	1.0644
NL Shrinkage	16.74	16.25	1.0304	NL Shrinkage	15.18	14.64	1.0371
DCC	12.24	14.13	0.8664	DCC	12.55	13.67	0.9181
DCC Denoising	12.43	13.45	0.9240	DCC Denoising	12.63	13.62	0.9273
DCC Shrinkage	15.20	16.42	0.9258	DCC Shrinkage	14.06	13.92	1.0098
DCC-NL	11.47	13.59	0.8442	DCC-NL	12.08	13.81	0.8750
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	14.01	14.87	0.9422	Raw Covariance	14.54	14.08	1.0325
Denoising (Mean)	13.90	14.98	0.9281	Denoising (Mean)	14.98	13.99	1.0709
Denoising (75th)	13.78	14.99	0.9197	Denoising (75th)	14.81	14.01	1.0574
Denoising (90th)	14.19	14.28	0.9937	Denoising (90th)	14.29	13.29	1.0750
Shrinkage	14.63	15.20	0.9625	Shrinkage	15.25	14.29	1.0672
NL Shrinkage	14.84	15.43	0.9621	NL Shrinkage	15.34	14.61	1.0494
DCC	12.26	14.13	0.8673	DCC	12.65	13.51	0.9361
DCC Denoising	12.20	14.28	0.8546	DCC Denoising	12.85	13.42	0.9574
DCC Shrinkage	12.10	14.11	0.8571	DCC Shrinkage	14.11	13.79	1.0231
DCC-NL	12.03	14.15	0.8506	DCC-NL	12.58	13.56	0.9275
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	13.90	14.89	0.9333	Raw Covariance	14.44	14.09	1.0246
Denoising (Mean)	13.93	14.68	0.9494	Denoising (Mean)	14.42	14.03	1.0279
Denoising (75th)	13.79	14.69	0.9386	Denoising (75th)	14.39	14.07	1.0226
Denoising (90th)	13.48	13.84	0.9739	Denoising (90th)	14.14	13.22	1.0694
Shrinkage	14.71	14.94	0.9850	Shrinkage	15.54	14.25	1.0908
NL Shrinkage	14.52	15.24	0.9525	NL Shrinkage	15.38	14.45	1.0645
DCC	11.96	13.98	0.8558	DCC	12.77	13.68	0.9338
DCC Denoising	11.95	13.82	0.8648	DCC Denoising	12.29	13.65	0.9003
DCC Shrinkage	13.38	14.31	0.9347	DCC Shrinkage	14.32	13.86	1.0328
DCC-NL	11.53	14.00	0.8237	DCC-NL	12.23	13.74	0.8903

Table 5.19: **Out-of-sample portfolio performance under Support Vector Regression forecasts (Hierarchical Clustering) - MS Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	13.16	13.79	0.9542	Raw Covariance	13.12	13.54	0.9694
Denoising (Mean)	13.07	13.95	0.9369	Denoising (Mean)	13.03	13.78	0.9458
Denoising (75th)	13.00	13.96	0.9318	Denoising (75th)	12.97	13.80	0.9399
Denoising (90th)	13.47	13.33	1.0110	Denoising (90th)	13.12	13.54	0.9694
Shrinkage	14.86	14.21	1.0459	Shrinkage	14.17	13.87	1.0218
NL Shrinkage	13.97	14.15	0.9872	NL Shrinkage	13.55	13.85	0.9786
DCC	11.22	13.38	0.8381	DCC	11.88	13.11	0.9060
DCC Denoising	11.59	13.37	0.8668	DCC Denoising	11.55	13.24	0.8723
DCC Shrinkage	13.59	13.67	0.9940	DCC Shrinkage	13.68	13.50	1.0132
DCC-NL	11.52	13.29	0.8671	DCC-NL	11.57	13.15	0.8803
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	12.52	14.10	0.8880	Raw Covariance	13.79	13.60	1.0134
Denoising (Mean)	12.67	14.09	0.8993	Denoising (Mean)	13.83	13.66	1.0126
Denoising (75th)	12.56	14.12	0.8893	Denoising (75th)	13.04	13.03	1.0001
Denoising (90th)	12.66	13.42	0.9428	Denoising (90th)	14.56	13.82	1.0531
Shrinkage	14.26	14.43	0.9888	Shrinkage	14.11	13.80	1.0231
NL Shrinkage	13.16	14.38	0.9151	NL Shrinkage	12.22	12.96	0.9434
DCC	11.85	13.57	0.8736	DCC	13.97	13.43	1.0404
DCC Denoising	11.28	13.66	0.8257	DCC Denoising	12.22	12.96	0.9434
DCC Shrinkage	13.13	14.07	0.9332	DCC Shrinkage	13.97	13.43	1.0404
DCC-NL	11.31	13.62	0.8298	DCC-NL	12.10	13.03	0.9286
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	13.05	13.65	0.9562	Raw Covariance	13.54	13.73	0.9866
Denoising (Mean)	12.95	13.62	0.9511	Denoising (Mean)	13.11	13.87	0.9448
Denoising (75th)	13.02	13.63	0.9556	Denoising (75th)	13.02	13.86	0.9391
Denoising (90th)	12.66	13.16	0.9621	Denoising (90th)	13.16	13.24	0.9937
Shrinkage	14.09	14.03	1.0045	Shrinkage	14.63	14.07	1.0394
NL Shrinkage	13.59	14.00	0.9710	NL Shrinkage	14.14	14.06	1.0056
DCC	11.22	12.97	0.8649	DCC	12.18	13.29	0.9169
DCC Denoising	11.22	13.06	0.8589	DCC Denoising	11.87	13.32	0.8907
DCC Shrinkage	13.58	13.73	0.9891	DCC Shrinkage	14.05	13.62	1.0316
DCC-NL	11.07	13.25	0.8353	DCC-NL	11.89	13.30	0.8943

Table 5.20: Out-of-sample portfolio performance under Elastic Net forecasts (ONC Algorithm) - MS Objective — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.24	15.12	1.0082	Raw Covariance	13.71	13.78	0.9950
Denoising (Mean)	15.56	16.71	0.9308	Denoising (Mean)	15.34	16.71	0.9179
Denoising (75th)	15.71	16.53	0.9500	Denoising (75th)	15.41	16.71	0.9220
Denoising (90th)	13.16	14.07	0.9348	Denoising (90th)	14.78	15.31	0.9651
Shrinkage	16.01	17.95	0.8917	Shrinkage	15.82	16.32	0.9692
NL Shrinkage	16.60	17.90	0.9277	NL Shrinkage	16.09	16.74	0.9614
DCC	13.16	14.07	0.9348	DCC	14.54	13.82	1.0524
DCC Denoising	13.67	13.74	0.9945	DCC Denoising	14.13	14.03	1.0075
DCC Shrinkage	14.45	16.50	0.8753	DCC Shrinkage	14.90	14.09	1.0572
DCC-NL	13.22	13.97	0.9464	DCC-NL	13.71	13.78	0.9950
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	16.56	15.18	1.0908	Raw Covariance	16.63	16.11	1.0322
Denoising (Mean)	15.98	15.63	1.0225	Denoising (Mean)	16.01	16.66	0.9609
Denoising (75th)	16.00	15.62	1.0241	Denoising (75th)	16.18	16.73	0.9673
Denoising (90th)	15.77	15.64	1.0085	Denoising (90th)	15.52	15.95	0.9727
Shrinkage	16.38	16.30	1.0046	Shrinkage	16.58	16.63	0.9969
NL Shrinkage	16.74	16.00	1.0466	NL Shrinkage	17.00	17.04	0.9973
DCC	14.95	13.69	1.0923	DCC	15.12	13.80	1.0963
DCC Denoising	14.78	13.90	1.0627	DCC Denoising	14.42	14.00	1.0300
DCC Shrinkage	15.30	14.23	1.0751	DCC Shrinkage	15.16	14.23	1.0657
DCC-NL	14.39	13.61	1.0575	DCC-NL	14.14	13.94	1.0147
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	15.08	17.50	0.8618	Raw Covariance	15.24	15.12	1.0082
Denoising (Mean)	14.68	16.54	0.8874	Denoising (Mean)	14.56	15.81	0.9207
Denoising (75th)	14.74	16.53	0.8913	Denoising (75th)	14.91	15.70	0.9496
Denoising (90th)	14.24	15.59	0.9134	Denoising (90th)	14.77	15.03	0.9826
Shrinkage	15.18	16.46	0.9221	Shrinkage	15.85	15.51	1.0222
NL Shrinkage	15.28	16.79	0.9099	NL Shrinkage	15.97	15.88	1.0056
DCC	14.22	14.36	0.9907	DCC	15.37	14.06	1.0933
DCC Denoising	13.67	14.60	0.9368	DCC Denoising	15.05	14.23	1.0577
DCC Shrinkage	14.43	15.03	0.9595	DCC Shrinkage	15.52	14.28	1.0870
DCC-NL	14.06	14.56	0.9657	DCC-NL	14.63	13.81	1.0593

Table 5.21: **Out-of-sample portfolio performance under Elastic Net forecasts (Hierarchical Clustering) - MS Objective** — Comparison of covariance estimators under different distance metrics.

(a) Pearson Distance				(b) Spearman Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	13.96	14.75	0.9465	Raw Covariance	13.28	15.24	0.8712
Denoising (Mean)	13.69	15.50	0.8829	Denoising (Mean)	13.49	15.64	0.8625
Denoising (75th)	13.68	15.55	0.8797	Denoising (75th)	13.57	15.67	0.8661
Denoising (90th)	13.52	14.92	0.9062	Denoising (90th)	12.86	15.06	0.8541
Shrinkage	14.48	15.56	0.9308	Shrinkage	13.83	15.79	0.8759
NL Shrinkage	14.21	15.47	0.9180	NL Shrinkage	13.44	15.42	0.8713
DCC	13.60	12.76	1.0659	DCC	12.90	13.78	0.9368
DCC Denoising	12.30	13.06	0.9417	DCC Denoising	12.73	13.70	0.9290
DCC Shrinkage	13.93	13.64	1.0214	DCC Shrinkage	13.56	13.91	0.9749
DCC-NL	12.30	13.06	0.9417	DCC-NL	11.61	13.43	0.8642
(c) Mutual Information Distance				(d) Kendall Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	14.17	15.53	0.9128	Raw Covariance	13.28	15.24	0.8712
Denoising (Mean)	13.63	15.44	0.8829	Denoising (Mean)	13.49	15.64	0.8625
Denoising (75th)	13.62	15.46	0.8810	Denoising (75th)	13.57	15.67	0.8661
Denoising (90th)	13.34	15.22	0.8763	Denoising (90th)	12.86	15.06	0.8541
Shrinkage	14.81	16.22	0.9127	Shrinkage	13.83	15.79	0.8759
NL Shrinkage	14.36	16.32	0.8802	NL Shrinkage	11.61	13.43	0.8642
DCC	13.18	13.64	0.9665	DCC	12.90	13.78	0.9368
DCC Denoising	12.89	13.71	0.9405	DCC Denoising	12.73	13.70	0.9290
DCC Shrinkage	13.57	14.07	0.9645	DCC Shrinkage	13.56	13.91	0.9749
DCC-NL	12.25	13.83	0.8858	DCC-NL	11.61	13.43	0.8642
(e) Correlation Distance				(f) Covariance Distance			
Covariance Estimator	Return (%)	Std (%)	Sharpe	Covariance Estimator	Return (%)	Std (%)	Sharpe
Raw Covariance	13.34	15.62	0.8538	Raw Covariance	13.46	15.33	0.8786
Denoising (Mean)	13.05	16.16	0.8075	Denoising (Mean)	13.31	13.96	0.9533
Denoising (75th)	13.16	16.18	0.8132	Denoising (75th)	12.87	15.72	0.8186
Denoising (90th)	13.19	15.38	0.8577	Denoising (90th)	12.97	15.10	0.8592
Shrinkage	13.95	16.06	0.8687	Shrinkage	14.06	15.79	0.8905
NL Shrinkage	13.61	15.74	0.8644	NL Shrinkage	13.29	15.24	0.8717
DCC	13.68	13.85	0.9876	DCC	13.31	13.96	0.9533
DCC Denoising	12.76	13.55	0.9416	DCC Denoising	13.18	14.05	0.9383
DCC Shrinkage	13.68	13.85	0.9876	DCC Shrinkage	13.83	14.11	0.9800
DCC-NL	12.30	13.52	0.9099	DCC-NL	12.51	13.83	0.9048