



School of Education, Culture and Communication
Division of Applied Mathematics

MASTER THESIS IN MATHEMATICS/APPLIED MATHEMATICS

OPTIMAL LINEAR COMBINATIONS OF PORTFOLIOS SUBJECT TO ESTIMATION RISK

AUTHOR:

ROBIN JONSSON

MASTERUPPSATS I MATEMATIK / TILLÄMPAD MATEMATIK

DIVISION OF APPLIED MATHEMATICS
MÄLARDALEN UNIVERSITY
SE-721 23 VÄSTERÅS, SWEDEN



School of Education, Culture and Communication
Division of Applied Mathematics

DIVISION OF APPLIED MATHEMATICS

Master Thesis in Mathematics / Applied Mathematics

Date:
26th June 2015

Project Name:
Optimal Linear Combinations of Portfolios Subject to Estimation Risk

Author:
Robin Jonsson

Program:
Master's Programme in Financial Engineering, 120 ECTS Credits

Supervisors:
Lars Pettersson, Senior Lecturer
Linus Carlsson, Senior Lecturer

Examiner:
Anatoliy Malyarenko, Professor

Comprising:
30 ECTS credits

Abstract

The combination of two or more portfolio rules is theoretically convex in return-risk space, which provides for a new class of portfolio rules that gives purpose to the Mean-Variance framework out-of-sample. The author investigates the performance loss from estimation risk between the unconstrained Mean-Variance portfolio and the out-of-sample Global Minimum Variance portfolio. A new two-fund rule is developed in a specific class of combined rules, between the equally weighted portfolio and a mean-variance portfolio with the covariance matrix being estimated by linear shrinkage. The study shows that this rule performs well out-of-sample when covariance estimation error and bias are balanced. The rule is performing at least as good as its peer group in this class of combined rules.

ACKNOWLEDGEMENTS

The whole idea of this thesis sort of started where Portfolio Theory I & II ended. I remember an unfulfilled curiosity being how portfolio theory could be applied practically, to help real investors make real decisions.

I was intrigued by the covariance estimator by Ledoit & Wolf, that shrunk estimation error, and after brief e-mail correspondence with Olivier Ledoit, I started doing research. I would like to thank Olivier Ledoit & Michael Wolf for writing the excellent (in my opinion), not-so-technical paper “Honey, I Shrunk The Covariance Matrix”, and urge anyone interested in the mean-variance framework to read it. That paper was the starting point in finding a topic for my thesis.

Anyone who has written an academic paper will probably recognize the feeling when I say that it has been an emotional roller-coaster. There are a number of people I hold in high regard for being able to finish this thesis. My supervisor Lars Pettersson has been a great lecturer and mentor, that has the unique ability to captivate attention. He seldom gives the obvious answer, but plants seeds in your head that lets you do the discoveries, if you nurture them. Lars is sincere and straightforward, and always gets back to you beyond expectation.

My other supervisor, Linus Carlsson, has been a devoted sounding board for generating ideas to problems that had no obvious solution, and for finding my mistakes in well time so I could repair them. I did not have as much previous experience with Linus, but he is a great asset to the division of mathematics and I really recommend future students to apply for his supervision.

I would further like to direct my thanks to two students, Jessica and David Radeschnig, whom I have shared study room with during the semester. We have laughed, sighed, shared ideas, and helped each other even though we worked on different topics. Moreover, Jessica deserves a special thank you for the review and comments on my thesis, before completion.

Another thank you goes to my friend and fellow student Niklas Jørgensen who has supplied me with material, that was important for my thesis. Without his help, the quality would likely have been diminished.

I give a general appreciation to all the students I have met during my five years at Mälardalen University. These people have come from all over the world and widen my views of cultures and ethics, and made me friends for life.

I would also like to thank my parents for always letting me go my own way and explore the things in life that interests me. This has led to a curiosity about scientific research and a desire to learn how different fields in life as well as how scientific and social academia correlate to each other. Moreover I would like give a special thanks to my wonderful girlfriend Jennifer who always supports me even though I had to spend a lot of time with my studies. Without her I would had to make greater sacrifices in my private life to complete this thesis.

Robin Jonsson, June 2015

Contents

INTRODUCTION	6
1 FUNDAMENTALS OF PORTFOLIO SELECTION	10
1.1 The Portfolio Parameters	10
1.1.1 The Sharpe Ratio	11
1.1.2 The Equally Weighted Portfolio	11
1.2 Efficient Portfolios	11
1.2.1 The Unconstrained Mean Variance Portfolio	12
1.2.2 The Global Minimum Variance Portfolio	12
1.2.3 The Target Portfolio	13
2 ESTIMATION RISK	15
2.1 Estimators under Parameter Uncertainty	15
2.1.1 More Important Distribution Results	16
2.2 Estimation Risk in Portfolio Rules	17
2.2.1 Uncertainty in the Unconstrained Portfolio	17
2.2.2 Uncertainty in the Global Minimum Variance Portfolio	18
2.3 A Well-Conditioned Covariance Shrinkage Estimator	22
2.3.1 Consistent Estimators to the Shrinkage Constant	24
2.3.2 Target Matrices	25
3 PORTFOLIO RULES	26
3.1 The Proposed Combined Rule	26
3.1.1 Tractable Solutions for the Expectations	28
3.1.2 The Optimal Combination Constant	30
3.2 Other Combined Rules	31
3.2.1 The Traditional Markowitz Rule	31
3.2.2 The Kan & Zhou (2007) Combined Rule	32
3.2.3 The Jorion (1986) Combined Rule	33
4 PERFORMANCE EVALUATION	35
4.1 The Experimental Design	35
4.1.1 Performance Measures	36
4.1.2 Data Sample	37

4.2	Experimental Results	37
4.2.1	Combination Coefficients	38
4.2.2	Out-of-Sample Results	40
4.2.3	Convexity Plots for the Combined Rules	44
CONCLUSION		47
REFERENCES		49
APPENDICES		52
A MATHEMATICAL PROOFS		52
A.1	Derivation of Equation (1.12)	52
A.2	Proof of Proposition 3	54
B MORE TABLES & FIGURES		55
B.1	Figures	55
B.2	Tables	57
B.3	Convexity Plots	60
C MATLAB FUNCTIONS		62
C.1	Functions for Estimated Combination Coefficients	62
C.1.1	The Combined Proposed Rule	62
C.1.2	The Combined Markowitz Maximum Likelihood Rule	63
C.1.3	The Combined Kan & Zhou Rule	64
C.1.4	The Combined Jorion Rule	66
C.2	Support Functions	67
C.2.1	Function for Estimated ψ^2	67
C.2.2	Function for Estimated θ^2	67
C.2.3	Function for Simulation of M-Variate Normal Matrix	68
C.2.4	Compression of Matrix to Portfolio Size	68

List of Figures

4.1	Ratio of Extreme Combination Coefficients	39
4.2	Convexity Plot for the Proposed Rule, CLW (ccm), $T = 120$	44
4.3	Convexity Plot for the Proposed Rule, CLW (sid), $T = 120$	45
4.4	Convexity Plot for the combined Kan & Zhou rule, CKZ, $T = 120$	46
4.5	Convexity Plot for the Combined Kan & Zhou Rule, CKZ, $T = 150$	46
B.1	Ratio of Extreme Combination Coefficients, Case of $N = 10$	55
B.2	Ratio of Extreme Combination Coefficients, Case of $N = 50$	56
B.3	Ratio of Extreme Combination Coefficients, Case of $N = 75$	56
B.4	Convexity Plot for the Proposed Rule, CLW (ccm), $T = 150$	60
B.5	Convexity Plot for the Proposed Rule, CLW (sid), $T = 150$	60
B.6	Convexity Plot for the Maximum Likelihood Rule, CML, $T = 120$	61
B.7	Convexity Plot for the Maximum Likelihood Rule, CML, $T = 150$	61

List of Tables

1	List of Notations	5
2.1	Estimation Risk for $R(\hat{\mathbf{w}}, \mathbf{w}_*)$	21
4.1	List of Portfolio Rules	37
4.2	Ratio of Sophisticated Weights	40
4.3	Out-of-Sample Sharpe Ratios	41
4.4	Certainty Equivalent Returns, $\gamma = 3$	43
B.1	Certainty Equivalent Returns, $\gamma = 1$	57
B.2	Table of Empirical Returns & Standard Deviations, Panel 1 & 2	58
B.3	Table of Empirical Returns & Standard Deviations, Panel 3 & 4	59

Notations & Conventions

This thesis is written in the field of applied mathematics with specialization in finance. The conventions and notations used herein follow the practice of financial academia, which implies that it might be different from academic papers written in technical mathematics and statistics. This is deliberate due to the fact that readers are likely more used to this convention. Important notations are summarized in Table 1

List of Notations

$\boldsymbol{\mu}$	Bold lower case mu denotes $1 \times N$ mean vector
$\mathbf{1}_N$	$1 \times N$ vector of ones
\mathbf{r}	Bold lower case letter denotes $N \times 1$ real vector
$\boldsymbol{\Sigma}$	Bold captial sigma denotes $N \times N$ covariance matrix
\mathbf{I}	$N \times N$ identity matrix
\mathbf{B}	Bold captial letter denotes matrix of constants
'	Denotes transpose of vector or matrix
$\text{tr}(\mathbf{A})$	Denotes trace of the matrix \mathbf{A}
N	Sample dimension
T	Sample size
$E[\cdot]$	Expectation of a random variable
$\text{Var}[\cdot]$	Variance of a random variable

Table 1: List of Notations

Moreover, for an arbitrary variable ω , its estimator is either denoted by $\tilde{\omega}$ or $\hat{\omega}$. Tilde is reserved for the sample estimators, and the hat is used for scaled estimators, estimators that stems from other articles, and in cases where an estimator with tilde is already defined, to distinguish between the two. The above table is a complement to the text. Most variables and symbols are explained throughout the text as well.

INTRODUCTION

Ever since the seminal work by Markowitz (1952), portfolio managers have utilized the mean-variance (MV) model for allocating assets to portfolios. The model, which is excellent in theory, relies only on asset returns, variances, and covariances to find the optimal portfolio construction. In practice the problem revolves around parameter uncertainty, as the true parameters that optimize the allocation are unknown to the investor and must be estimated. The statistical convention of using the maximum likelihood (ML) estimators in MV optimization á la Markowitz leads to estimation errors and suboptimal decisions by investors with regard to ex ante performance. This is a well documented problem, especially when the number of securities are large relative to the sample size (Jobson & Korkie, 1980). The phenomenon occurs because of extreme observations relative to the sample mean, that latch on when the sample covariance matrix is computed. The MV optimizer then places bets where the rewards are seemingly the greatest, as oppose to most true. Michaud (1989) labels it “estimation-error maximization” and carries a throughout discussion on the associated limitations. For a portfolio manager this is devastating, since it suppresses their true abilities of stock picking. The manager should know that the nuts and bolts to weight stability and realized Sharpe ratio in portfolios lies in controlling asset returns and covariances through reliable estimation.

Various techniques have been proposed as remedies to this estimation problem. This paper will focus on Bayes–Stein estimation, which have shown great improvement on portfolio performance. It originates from nearly half a century ago, when Charles Stein of Stanford University showed that the usual unbiased mean estimator for the multivariate normal distribution was inadmissible (Stein, 1956). Instead Stein proposed an estimator that shrinks extreme data towards a grand mean. James & Stein (1961) refined the proof of the usual estimators’ inadmissibility and suggested two loss functions for computing the spatial distance, or loss, between two estimators and their common unknown true value. The method used is known as Bayesian¹ because the estimator is conditional on some prior information assumed to be known.

The first non *ad hoc* academic attempts at using Bayes-Stein estimation to shrink the parameters in MV portfolio optimization, is made by Jorion (1986) and Frost & Savarino (1986). It is shown that their shrinkage estimators substantially outperform the sample mean. Both papers impose informative priors for the mean using an empirical Bayesian approach. They employ shrinkage towards a grand mean, that is an “all stocks are identical in return” assumption. Jorion (1986) also suggests using a prior for shrinkage intensity, and studies the limits

¹The origin is attained Thomas Bayes, an 18th-century reverend who studied probability theory. His work is published in *An Essay towards solving a Problem in the Doctrine of Chances*(1763).

of the intensity². The trouble in Frost & Savarino (1986) is that the parameters for shrinkage intensity are assumed to be known *a priori* and estimators for them needs to be numerically approximated. Both papers also fails to consider the case in which the dimensionality is greater than the sample size. If this is the case, the sample covariance matrix is rank deficient and singular. It is well known that the sample covariance matrix is well-conditioned (invertible) asymptotically, but performs poorly on finite samples. In a non-technical fashion, Chopra et al (1993) showed that Stein-type shrinkage of the means and correlations toward a global average for each asset class, before computing covariance matrix and portfolio weights, results in higher mean return and lower variance out-of-sample.

Ledoit & Wolf (2004a) proposes a Bayes-Stein estimator for covariance that is both well-conditioned on finite samples and performs better than the ML estimator (being the sample covariance matrix). The estimator is a linear combination between the sample covariance matrix and a structured matrix³. The structured matrix they propose is an identity matrix scaled by the variance of the sample matrix. Ledoit & Wolf (2003) suggests another structured matrix consisting of the covariance structure of a one-factor model. Their result shows that shrinking towards the one-factor model shows the best out-of-sample performance with regard to variance, followed by shrinking towards identity. They also give a consistent estimator for the optimal shrinkage intensity. Ledoit & Wolf (2004b) proposes the constant correlation model as a third structured matrix. It is comparable to the one-factor model, but easier to implement since it does not require computation of regression coefficients for all samples and periods.

When analysing the estimators efficiency on portfolio level, there is yet another rule that is childishly simple but perform extremely well. The rule is known as one-over- N or naive diversification because it simply treats all the portfolio components with equal weight. Demiguel et al (2009) puts 14 sophisticated portfolio rules to the test in a horse race against the naive rule and show that no sophisticated rule consistently outperforms the naive rule out-of-sample, in terms of Sharpe-ratio and certainty equivalent return (CER). This result makes one wonder whether it is worthwhile to invest time (and capital) in sophisticated portfolio rules. Tu & Zhou (2011) employ a different approach and combine the naive portfolio rule with three sophisticated rules. Their positive results re-validate the usefulness of portfolio theory as value adding under errors from parameter uncertainty. The methodology of combining portfolios as a diversification tool can also be found in for example Jorion (1986), Kan & Zhou (2007) and Demiguel et al (2009).

There is a rich literature of estimators from various techniques that is not considered in this thesis. One class considers econometrics and factor models to estimate the portfolio parameters. There is a considerable amount of work done on both mean and variance estimation. Confer for example Grinold & Kahn (2000) and references therein for techniques on factor models. The unappealing characteristic of factor models is that they are very *ad hoc* in construction. One model might work well on a particular data set, but perform poor on others. There is no way of telling *a priori*, how well the model will work.

²When the intensity is zero, the mean estimator equals that of the maximum likelihood.

³Structured matrices will be presented and discussed in Section 2.3.2

Another class of estimators belong to the Bayesian variety. Bayesian estimators⁴ have been considered for portfolio analysis for about half a decade with various results. Kalymon (1971) used conjugate priors and subjective information to derive a posterior distribution for the mean and variance. Jorion (1986) considered an empirical Bayesian approach and defined optimal portfolio choice in terms of a predictive density function, in order to solve the posterior density given prior information. Pástor (2000) suggests a Bayesian informative prior, but lets the likelihood function⁵ follow a multidimensional factor model. The Bayesian approaches have shown to be dominated by shrinkage estimation (Ledoit & Wolf, 2004b) and will therefore not be considered, as shrinkage is a more general approach that also performs better out-of-sample.

A third class of estimators treats the covariance structure with constraints. This approach is known as thresholding or penalizing. The techniques in the literature are either general or commonly applied to biometrics and meteorology, where the dimensionality of the covariance matrix can be very large, while the sample size is small. See for example Bickel & Levina (2008) for thresholding, Huang et al (2006) for normal penalizing likelihood, or Chang & Tsay (2010) for a recent application to portfolio selection.

A final class is known as re-sampling. The portfolio parameters are simulated from their sampling distribution, and an efficient frontier is computed for each simulation. Then the “true” frontier is assumed to be the average of simulations, however Harvey et al (2008) confirmed that a Bayesian approach dominated re-sampling.

There is a central message among the literature reviewed by the author. There exists interior optimums in linear convex combinations of different mean-, covariance- and weight-structures on portfolio level. These optimums have become more appreciated and central in recent literature, which shows that they are indeed important for portfolio formulation. The author intends to study the sources that add value (or minimizes loss) to portfolio performance under parameter uncertainty and get a better understanding of risk structures, value drivers and diversification effects. The approach will have focus on Bayes-Stein shrinkage and combination of portfolio rules, as the main problem is to derive the weighted combination of the naive portfolio rule, and the estimated rule by Ledoit & Wolf (2004b). The rationale behind this is that the methodology of Tu & Zhou (2011) has shown promising results and is an interesting diversification strategy, while the estimator for the weights of the sophisticated rule is considered to be among the best in the literature. A combination of the two would imply that the covariance estimator will balance covariance bias and error, while the portfolio combination coefficient will balance weight bias and error.

The rest of the paper is organized as follows. Chapter 1 gives the basic concepts of portfolio theory to the unfamiliar reader. Those familiar to MV portfolio formulation, quadratic utility and maximum likelihood estimation can without informational loss, skim the equations and skip the chapter. Chapter 2 reviews important results for estimators under parameter uncertainty, discusses distribution results related to estimation risk and how this risk affects holding a portfolio of assets that carry risk. It also contains the derivation of a new proposition

⁴Not to be confused or mixed with Bayes-Stein estimators which are treated as a distinguished class.

⁵The likelihood function multiplied by the predictive distribution is proportional to the posterior distribution, used to compute the moments of the estimators.

on the estimation risk connected to the out-of-sample Global Minimum Variance portfolio. It ends by the presentation of a shrinkage estimator for the sample covariance matrix. Chapter 3 begins by deriving a new linear combination of two portfolio rules based on the shrinkage estimator and gives a proposition on the results. It continues by presenting existing combined rules and gives a proposition that the new rule is a generalization of an existing rule. Chapter 4 describes the empirical methodology and evaluates the results from a study conducted on real stock market data, with a rolling basis approach.

Chapter 1

FUNDAMENTALS OF PORTFOLIO SELECTION

An investor is intuitively interested in maximizing the wealth generated from future returns. Future returns are of unknown quantity, and therefore considered carrying risk. Financial return and risk are modelled through a MV analysis framework pioneered by Markowitz (1952). The risk of these returns is quantified by the dispersion around the expectation of these returns. An investor that accepts this framework seeks to allocate resources to a MV efficient portfolio.

1.1 The Portfolio Parameters

Before even discussing portfolios, a reader should be aware of the components that comprises the MV framework. In theory, the returns are assumed to be independent and identically distributed normally, that is let $\mathbf{R}_T = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_T\}$ be a set of observed securities in t periods, then $\mathbf{R}_T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The population mean $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$, are however unknown and must be estimated. In a traditionally statistical approach, this is achieved by the Maximum Likelihood (ML) estimators. They are the unbiased sample proxies for the true variables under a normal distribution. The sample mean and variance is given by

$$\tilde{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t, \quad (1.1)$$

and

$$\tilde{\boldsymbol{\Sigma}} = \frac{1}{T} \sum_{i=1}^T (\mathbf{r}_i - \tilde{\boldsymbol{\mu}})(\mathbf{r}_i - \tilde{\boldsymbol{\mu}})', \quad (1.2)$$

respectively.

A dense literature has showed that stock returns do not follow a normal distribution, or any other stable distribution for that matter. Returns are typically fat-tailed (confer for example Officer (1972) and references therein) which is problematic because standard statistical measures then fails to account for tail risk. It will be seen in Section 2.1 that (1.1) and (1.2) are inadmissible to other estimators, for portfolio selection.

1.1.1 The Sharpe Ratio

As a measure of risk-adjusted return, Sharpe (1966) developed the now standardized Sharpe ratio. Denote the scalar return on a portfolio by $\mu_p = \mathbf{w}'\boldsymbol{\mu}$ for some weight vector \mathbf{w} and return vector $\boldsymbol{\mu}$. Further denote its scalar variance by $\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ where $\boldsymbol{\Sigma}$ is the associated covariance matrix. The ratio for a portfolio is given by

$$\theta_p = \frac{\mu_p}{\sigma_p}, \quad (1.3)$$

and enables comparison among combinations of assets in the investable universe. A higher ratio is better, as it grows with return and diminishes with volatility. One drawback of the ratio is that it is scalable, and unconstrained portfolios can have extreme weights while the Sharpe ratio remains reasonable. The extreme weights problem will be solved by normalization in Section 4.1.

1.1.2 The Equally Weighted Portfolio

We find many names for those we love, the naive, or one-over- N , or equally weighted, or in the latest fashion; the “Talmudic”¹ portfolio rule. Within this paper it will be referred to as Equally Weighted (EW). Its return and variance is given by

$$r_p = \frac{1}{N}\boldsymbol{\mu}'\mathbf{1}_N \quad \text{and} \quad \sigma_p^2 = \frac{1}{N^2}\mathbf{1}_N'\boldsymbol{\Sigma}\mathbf{1}_N, \quad (1.4)$$

for $\mathbf{w}_{EW} = \mathbf{1}_N/N$ where $\mathbf{1}_N$ is a vector of ones, with N elements. Its simplicity and yet effective performance have given sophisticated rules some hard run for the money (Demiguel et al, 2009). Platen & Rendek (2012) device index strategies based on the EW rule, and show that the EW rules have higher return and less risk versus benchmark indices over time.

The advantage of this portfolio rule is its independence of estimators. It means that the portfolio is free of estimation error, but by construction still carries bias toward an optimal MV portfolio. Moreover as the number of assets increase, the portfolio diversifies away asset specific variances. It will be seen in Chapter 4, how risk is suppressed by diversification.

1.2 Efficient Portfolios

All assets and combinations of assets that constitute a bundle or portfolio lie in the first or fourth quadrant in (σ, μ) -space. The set has a convex upper boundary where all the efficient asset combinations lie. The combinations are known as MV efficient portfolios and collectively represents the efficient frontier. The efficient frontier, which constitutes the efficient part of the set of possible portfolios lie in the first quadrant where $\mu > 0$, as any portfolio where $\mu < 0$ is inferior to holding cash or a risk-free investment with fixed return. A presentation of the most common efficient portfolios will be accounted for in this section.

¹Courtesy of Demiguel et al (2009).

1.2.1 The Unconstrained Mean Variance Portfolio

Consider a universe of n assets with mean return vector $\boldsymbol{\mu} \in \mathbb{R}^N$, weight vector $\mathbf{w} \in \mathbb{R}^N$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{S}_{++}^{N \times N}$, where $\mathbb{S}_{++} \subset \mathbb{R}_{++}^{N \times N}$ is the set of positive definite symmetric matrices. Positive definiteness ensures that the covariance matrix is invertible. The return and variance of a MV portfolio is given by $\mu_p = \mathbf{w}'\boldsymbol{\mu}$ and $\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ respectively. There are infinitely many combinations of \mathbf{w} that the investor can choose from, and a conventional way of distinguishing between them is through the utility the investor receives from a particular combination. The quadratic utility function has the properties for a quantitative solution to this decision problem. To see this, let

$$U(\mathbf{w}) = \mathbf{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}, \quad (1.5)$$

be the utility function. As seen from (1.5), return increases utility whereas variance is penalizing. The constant γ is referred to as the risk aversion coefficient and determines the investor's risk tolerance. Then by differentiating (1.5) with respect to \mathbf{w} and solving for the optimal \mathbf{w}_* gives

$$\mathbf{w}_* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \quad (1.6)$$

The well known solution in (1.6) is the unconstrained optimal rule. It is the true quantity held in each asset if $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ had been known. It lies on the efficient frontier which constitutes all portfolios (combinations of assets) that lie on the boundary of the convex set. There exists no other portfolio which is preferred to an efficient portfolio. A portfolio on the efficient frontier is chosen with regard to the level of return an investor is expecting. The minimum variance portfolio is the least risky one attainable, given a set of assets to choose from. It follows directly from (1.5) and (1.6) that the utility of the true portfolio is

$$U(\mathbf{w}_*) = \frac{1}{2\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \frac{\theta^2}{2\gamma}. \quad (1.7)$$

A result notable in (1.7) is that the utility changes with the square of the Sharpe ratio of the *ex ante* MV portfolio.

1.2.2 The Global Minimum Variance Portfolio

The Global Minimum Variance (GMV) portfolio is the portfolio that minimize variance, constrained by its weights adding up to unity. The portfolio has a well known analytical solution for the weight vector. Consider the same setting as in the previous section, then the problem is

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}'\mathbf{1}_N = 1 \end{aligned} \quad (1.8)$$

with Lagrangian given by

$$\mathcal{L}(\mathbf{w}, \lambda) = \frac{1}{2} \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} - \lambda (\mathbf{w}' \mathbf{1}_N - 1).$$

The solution to the optimization problem is

$$\mathbf{w}_g = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}_N}{\mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \mathbf{1}_N}, \quad (1.9)$$

which can easily be checked. As seen from (1.9), the solution is directly independent of $\boldsymbol{\mu}$ (but not indirectly in estimators of $\boldsymbol{\Sigma}$), contrary to the unconstrained portfolio rule in (1.6). Jagannathan & Ma (2003) showed that when uncertainty about means are large, no loss is made holding the GMV portfolio instead of the true MV portfolio.

Let $U(\mathbf{w}_g)$ be the utility of the GMV portfolio. Then by using (1.5),

$$\begin{aligned} U(\mathbf{w}_g) &= \left(\frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}_N}{\mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \mathbf{1}_N} \right)' \boldsymbol{\mu} - \frac{\gamma}{2} \left(\frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}_N}{\mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \mathbf{1}_N} \right)' \boldsymbol{\Sigma} \left(\frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}_N}{\mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \mathbf{1}_N} \right), \\ &= \mu_g - \frac{\gamma}{2 \mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \mathbf{1}_N}, \end{aligned} \quad (1.10)$$

where μ_g is the expected excess return on the *ex ante* GMV portfolio, as is evident from (1.9).

1.2.3 The Target Portfolio

The target portfolio is related to the GMV portfolio except for the additional constraint of a minimum target return, which is usually the case for portfolio managers seeking to meet or beat a benchmark return. The problem is stated as

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} \\ \text{S.t.} \quad & \mathbf{w}' \mathbf{1} = 1 \\ & \mathbf{w}' \boldsymbol{\mu} = q, \end{aligned} \quad (1.11)$$

where q is a minimum return in basis points. The solution is

$$\mathbf{w}^* = \frac{\theta^2 - qb}{a\theta^2 - b^2} \boldsymbol{\Sigma}^{-1} \mathbf{1} + \frac{qa - b}{a\theta^2 - b^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad (1.12)$$

with $a = \mathbf{1}' \boldsymbol{\Sigma}^{-1} \mathbf{1}$ and $b = \mathbf{1}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ for brevity. The interested reader can confer Appendix A.1 for a derivation of (1.12) since it involves more algebra than the previous rules.

From this rule, the efficient frontier can be traced out as the variance being a function of a return level q . Let the variance of a portfolio be $\sigma_p^2 : \mathbb{R} \ni q \rightarrow \sigma_p^2(q) \in \mathbb{R}^+$ and use the result

from (A.2) to get

$$\begin{aligned}
\sigma_p^2 &= \mathbf{w}'\Sigma\mathbf{w} \\
&= \mathbf{w}'\Sigma(\lambda\Sigma^{-1}\mathbf{1} + \gamma\Sigma^{-1}\boldsymbol{\mu}) \\
&= \lambda\mathbf{w}'\Sigma\Sigma^{-1}\mathbf{1} + \gamma\mathbf{w}'\Sigma\Sigma^{-1}\boldsymbol{\mu} \\
&= \lambda\mathbf{w}'\mathbf{1} + \gamma\mathbf{w}'\boldsymbol{\mu} \\
&= \lambda + \gamma q.
\end{aligned} \tag{1.13}$$

Now, put (A.4) and (A.5) in (1.13) to get

$$\begin{aligned}
\sigma_p^2 &= \lambda + \gamma q \\
&= \frac{\theta^2 - qb}{a\theta^2 - b^2} + \frac{aq - b}{a\theta^2 - b^2} q \\
&= \frac{\theta^2 + aq^2 - 2bq}{a\theta^2 - b^2}.
\end{aligned}$$

The efficient frontier can then be scaled into the appropriate time scale. This is done conventionally by multiplying the variance and returns by the scale t and the standard deviation by \sqrt{t} . The most common factor from daily to yearly scale is using a factor of 252, since it corresponds to the number of business days during a year.

Chapter 2

ESTIMATION RISK

Estimation risk occurs because the true unknown portfolio rule cannot be replicated, as asset returns vary over time. The MV framework is only truly optimal if the decision maker has the parametric information available, that is the ability to foresee the future. Unfortunately, this information is unavailable and must be estimated. Therefore the model output will only be as good as its inputs. The inputs suffer from parameter uncertainty, which can be quantified using risk functions. This section will assess the sources of estimation risk, discuss developed results and provide quantitative analysis so as to understand its impact on portfolio selection.

A risk function is designed to evaluate and analyse the opportunity cost or expected loss between two estimators. The literature distinguish between risk functions for estimators and portfolio rules. Estimators are usually evaluated by quadratic loss or entropy loss, while portfolio rules are evaluated by difference in utility.

2.1 Estimators under Parameter Uncertainty

The sample distribution of a covariance matrix from a multivariate normal random distribution follows a Wishart distribution, denoted $W_N(\tau, \mathbf{V})$ where τ denotes the number of degrees of freedom and $\mathbf{V} \in \mathbb{S}_{++}^{N \times N}$ is a scale matrix. An assertion on how the sample covariance matrix in (1.2) is connected to the Wishart distribution is made, and the maximum likelihood estimators for $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\Sigma}}^{-1}$ are derived.

Let the sample set $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_1 \dots, \mathbf{y}_N\}$ be a $T \times N$ random matrix such that $\mathbf{y}_i \sim N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = 1, \dots, N$, are independent random vectors, and $\boldsymbol{\Sigma}$ is the unknown population covariance matrix. Let the sample proxy be given by $\mathbf{A} = \mathbf{Y}\mathbf{Y}'$, then $\mathbf{A} \sim W_N(T - 1, \boldsymbol{\Sigma})$, with $T - 1$ degrees of freedom¹. Moreover, making the substitution $\mathbf{A}^{-1} = \mathbf{B}$, then $\mathbf{B} \sim W_N^{-1}(T - 1, \boldsymbol{\Sigma}^{-1})$, which is known as the inverted Wishart distribution. From Muirhead (1982, Chapter 3), it follows that

$$\mathbb{E}[\mathbf{A}] = (T - 1)\boldsymbol{\Sigma}, \quad \text{and} \quad \mathbb{E}[\mathbf{A}^{-1}] = \frac{1}{T - N - 2}\boldsymbol{\Sigma}^{-1}. \quad (2.1)$$

Now, recalling (1.2), $\tilde{\boldsymbol{\Sigma}}$ can be expressed as \mathbf{A}/T and thus, $\tilde{\boldsymbol{\Sigma}} \sim W_N(T - 1, \boldsymbol{\Sigma})/T$. It then

¹See for example Pestman (1998), page 434 for theorem and proof.

follows that

$$\mathbb{E}[\tilde{\boldsymbol{\Sigma}}] = \frac{\mathbb{E}[\mathbf{A}]}{T} = \frac{T-1}{T}\boldsymbol{\Sigma}. \quad (2.2)$$

This estimator has an important property when substituting the ML estimator into (2.2). Economically, it invests less aggressively as a portfolio optimization component than the traditional ML estimator.

By following the same reasoning, $\tilde{\boldsymbol{\Sigma}}^{-1}$ can be expressed as $T\mathbf{A}^{-1}$ such that $\tilde{\boldsymbol{\Sigma}}^{-1} \sim W_N(T-1, \boldsymbol{\Sigma}^{-1})/T$. The inverse estimator is then

$$\mathbb{E}[\tilde{\boldsymbol{\Sigma}}^{-1}] = T\mathbb{E}[\mathbf{A}^{-1}] = \frac{T}{T-N-2}\boldsymbol{\Sigma}^{-1}, \quad \text{for } N+2 < T. \quad (2.3)$$

The inverse expectation is unfortunately not unbiased, but is still important for solving risk functions under parameter uncertainty.

2.1.1 More Important Distribution Results

There also exist a useful identity in Haff (1979), which will be necessary when evaluating expectations in Sections 2.2 and 3.1. The identity is $\mathbf{W} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-\frac{1}{2}} \sim W_N(T-1, \mathbf{I}_N)/T$. The moments of \mathbf{W}^{-1} are shown to be

$$\mathbb{E}[\mathbf{W}^{-1}] = \left(\frac{T}{T-N-2} \right) \mathbf{I},$$

which is related to the result in (2.3), and

$$\mathbb{E}[\mathbf{W}^{-2}] = \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right) \mathbf{I}, \quad (2.4)$$

with \mathbf{I} being an $N \times N$ identity matrix. A second necessary result is with regard to the square of $\tilde{\boldsymbol{\mu}}$. It is shown in Pestman (1998, p. 407) that the exact sample distribution is $\tilde{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$. Then if $\boldsymbol{\Sigma}$ is non-singular, it is shown in Muirhead (1982, Theorem 1.4.1) that

$$\tilde{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}} \sim \chi_N^2(T\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})/T, \quad (2.5)$$

where χ_N^2 is the non-central Chi-squared distribution with N degrees of freedom and non-centrality parameter $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. From (2.5) it is easy to see that

$$\mathbb{E}[\tilde{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\mu}}] = \frac{N + T\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{T}, \quad (2.6)$$

from the the definition of expectation for the non-central Chi-Squared distribution.

2.2 Estimation Risk in Portfolio Rules

The first acquaintance with portfolio rules were made in Section 1.2. There it was established that the investor is concerned with the utility received from holding a portfolio rule. The utility of the rule in (1.5) was denoted by $U(\mathbf{w}_*)$, and was the highest utility attainable, strictly under the assumption that the investor has the true information. To compare utilities of portfolio rules, standard decision theory suggests to evaluate the strictly positive loss that occurs from parameter uncertainty. Let $U(\tilde{\mathbf{w}})$ be the utility of a portfolio rule using estimated parameters, then the loss is

$$L(\mathbf{w}_*, \tilde{\mathbf{w}}) = U(\mathbf{w}_*) - U(\tilde{\mathbf{w}}). \quad (2.7)$$

Note that since the second term in (2.7) contains estimated parameters, the utility itself must be estimated and is thus a random variable. Practically, $\tilde{\mathbf{w}}$ depends on the realizations of the dataset \mathbf{R}_T (see Section 1.1), and repeated observations can be seen as the expected out-of-sample performance measure of the rule. The loss can also be analytically quantified by solving $E[U(\tilde{\mathbf{w}})]$. Let $R(\mathbf{w}_*, \tilde{\mathbf{w}})$ be the risk function, such that

$$R(\mathbf{w}_*, \tilde{\mathbf{w}}) = E[L(\mathbf{w}_*, \tilde{\mathbf{w}})] = U(\mathbf{w}_*) - E[U(\tilde{\mathbf{w}})], \quad (2.8)$$

where $E[U(\tilde{\mathbf{w}})]$ denotes the expected utility out-of sample, as in Kan & Zhou (2007). The severity in loss depends highly on how realistic the estimation is, that is, the assumptions made about the parameters. There are three general cases which are distinct from each other in terms of prior knowledge of the future returns and variances. Kalymon (1971) and Barry (1974) specifies the cases as (1) $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ both known², (2) $\boldsymbol{\mu}$ unknown and $\boldsymbol{\Sigma}$ known and (3), $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ both unknown. More recent multivariate statistics³ have enabled the fourth case of $\boldsymbol{\mu}$ known, and $\boldsymbol{\Sigma}$ unknown to be studied.

2.2.1 Uncertainty in the Unconstrained Portfolio

The three cases of uncertainty have all been analytically solved for the unconstrained MV portfolio, in Kan & Zhou (2007). They show that for the first case where $\boldsymbol{\Sigma}$ is assumed known, the expected utility is

$$E[U(\tilde{\mathbf{w}}) | \boldsymbol{\Sigma}] = \frac{\theta^2}{2\gamma} - \frac{N}{2\gamma T},$$

and thus, using (1.7) and (2.8), the loss is $N/2\gamma T$. The loss increases with dimension and decreases with sample size. When the situation is reversed, that is $\boldsymbol{\mu}$ is known, the expected utility is

$$E[U(\tilde{\mathbf{w}}) | \boldsymbol{\mu}] = k_1 \frac{\theta^2}{2\gamma},$$

²This is the trivial case studied in Markowitz (1952) where the ML estimators are accepted as asymptotically true.

³Referring to the derivation of moments for the Wishart and inverted Wishart distribution, see for example Haff (1979).

with

$$k_1 = \frac{T}{T-N-2} \left(2 - \frac{T(T-2)}{(T-N-1)(T-N-4)} \right),$$

using two identities from Haff (1979) that involves the first and second moments of the inverse Wishart distribution. The loss in utility is $(1 - k_1)\theta^2/(2\gamma)$, due to estimation error in $\tilde{\Sigma}$. The third and most realistic case, with both parameters unknown is

$$\mathbb{E}[U(\tilde{\mathbf{w}})] = k_1 \frac{\theta^2}{2\gamma} - \frac{NT(T-2)}{2\gamma(T-N-1)(T-N-2)(T-N-4)},$$

with the estimation risk being the greatest.

2.2.2 Uncertainty in the Global Minimum Variance Portfolio

For the GMV portfolio, the case of Σ being known is redundant because the rule requires no estimation of μ , as discussed in Section 1.2.2. The two other specified cases are analogous for the same reason, and only one needs to be accounted for.

Let the weight of the sample GMV portfolio be given as in (1.9), and recognize that the expected return on the GMV portfolio is $\mu_g = \mathbf{w}'_g \mu$. Then by substituting μ_g into (1.5), the optimal weight can be written as

$$\mathbf{w} = \frac{1}{\gamma} \Sigma^{-1} \mu_g \mathbf{1}_N, \quad (2.9)$$

where both parameters are known. The utility of this portfolio is

$$\begin{aligned} U(\mathbf{w}) &= \mathbf{w}' \mu - \frac{\gamma}{2} \mathbf{w}' \Sigma^{-1} \mathbf{w} \\ &= \frac{\mu_g}{\gamma} \mathbf{1}'_N \Sigma^{-1} \mu - \frac{1}{2\gamma} \mu_g^2 \mathbf{1}'_N \Sigma^{-1} \Sigma \Sigma^{-1} \mathbf{1}_N. \\ &= \frac{\mu_g}{\gamma} \mathbf{1}'_N \Sigma^{-1} \mu - \frac{1}{2\gamma} \frac{\mu_g^2}{\sigma_g^2}, \\ &= \frac{\mu_g^2}{2\gamma \sigma_g^2} \end{aligned} \quad (2.10)$$

with $\sigma_g^2 = 1/\mathbf{1}'_N \Sigma^{-1} \mathbf{1}_N$ being the variance of the GMV portfolio, and μ_g^2/σ_g^2 its square Sharpe ratio. As we are dealing with estimation risk, the case with unknown parameters must be solved. The rule under parameter uncertainty is given by

$$\hat{\mathbf{w}} = \frac{1}{\gamma} \tilde{\Sigma}^{-1} \tilde{\mu}_g \mathbf{1}_N. \quad (2.11)$$

Let $U(\hat{\mathbf{w}})$ denote the utility of this rule, then

$$\begin{aligned} U(\hat{\mathbf{w}}) &= \hat{\mathbf{w}}' \mu - \frac{\gamma}{2} \hat{\mathbf{w}}' \Sigma^{-1} \hat{\mathbf{w}} \\ &= \frac{1}{\gamma} \left(\tilde{\Sigma}^{-1} \tilde{\mu}_g \mathbf{1}_N \right)' \mu - \frac{1}{2\gamma} \left(\tilde{\Sigma}^{-1} \tilde{\mu}_g \mathbf{1}_N \right)' \Sigma \left(\tilde{\Sigma}^{-1} \tilde{\mu}_g \mathbf{1}_N \right) \\ &= \frac{1}{\gamma} \tilde{\mu}_g \mathbf{1}'_N \tilde{\Sigma}^{-1} \mu - \frac{1}{2\gamma} \tilde{\mu}_g^2 \mathbf{1}'_N \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} \mathbf{1}_N. \end{aligned}$$

Using results from Section 2.1.1, the expected utility is

$$\begin{aligned}
E[U(\hat{\mathbf{w}})] &= E \left[\frac{1}{\gamma} \tilde{\mu}_g \mathbf{1}'_N \tilde{\Sigma}^{-1} \boldsymbol{\mu} - \frac{1}{2\gamma} \tilde{\mu}_g^2 \mathbf{1}'_N \tilde{\Sigma}^{-1} \boldsymbol{\Sigma} \tilde{\Sigma}^{-1} \mathbf{1}_N \right] \\
&= \frac{1}{\gamma} E \left[\tilde{\mu}_g \mathbf{1}'_N \tilde{\Sigma}^{-1} \boldsymbol{\mu} \right] - \frac{1}{2\gamma} E \left[\tilde{\mu}_g^2 \mathbf{1}'_N \tilde{\Sigma}^{-1} \boldsymbol{\Sigma} \tilde{\Sigma}^{-1} \mathbf{1}_N \right] \\
&= \frac{1}{\gamma} E \left[\tilde{\mu}_g \mathbf{1}'_N \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{W}^{-1} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\mu} \right] - \frac{1}{2\gamma} E \left[\tilde{\mu}_g^2 \mathbf{1}'_N \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{W}^{-2} \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}_N \right]. \quad (2.12)
\end{aligned}$$

To be able to solve the expectations I must use two partial results. Kan & Smith (2008, Proposition 1) shows that $\tilde{\mu}_g | \tilde{\psi}^2 \sim N(\mu_g, \sigma_g^2(1 + \tilde{\psi}^2)/T)$ where $\tilde{\psi}$ is the sample estimator of $\psi = (\boldsymbol{\mu} - \mu_g \mathbf{1}_N)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mu_g \mathbf{1}_N)$ and $\sigma_g^2 = 1/\mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \mathbf{1}_N$. Then by Kan & Smith (2008, Lemma 2), the moments of the return on the GMV portfolio are

$$E[\tilde{\mu}_g] = \mu_g, \quad \text{and,} \quad \text{Var}[\tilde{\mu}_g] = \frac{[T(1 + \psi^2) - 2] \sigma_g^2}{T(T - N - 1)}.$$

Using their results I can write

$$\begin{aligned}
E[\tilde{\mu}_g^2] &= \text{Var}[\tilde{\mu}_g] + E[\tilde{\mu}_g]^2, \\
&= \frac{[T(1 + \psi^2) - 2] \sigma_g^2}{T(T - N - 1)} + \mu_g^2,
\end{aligned}$$

by the definition of variance. Since $\tilde{\mu}_g$ and $\tilde{\Sigma}$ are independent, the expectations in (2.12) can be solved as the product of expectations. It follows that

$$E[U(\hat{\mathbf{w}})] = \frac{c_1 \mu_g \mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\gamma} - \frac{c_2}{2\gamma} \left(\frac{[T(1 + \psi^2) - 2] \sigma_g^2}{T(T - N - 1)} + \mu_g^2 \right), \quad (2.13)$$

where $c_1 = T/(T - N - 2)$ and $c_2 = T^2(T - 2)/[(T - N - 1)(T - N - 2)(T - N - 4)]$. To make the expression free of vectors and matrix notations we see can that $\mu_g \mathbf{1}'_N \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \mu_g^2 / \sigma_g^2$, and do the substitution. I propose the following for the estimation risk between the MV portfolio and the GMV portfolio from the above results,

Proposition 1. *Assume that $T > N + 4$, then the expected loss of holding the out-of-sample GMV portfolio instead of the MV portfolio under parameter uncertainty is*

$$R(\hat{\mathbf{w}}, \mathbf{w}_*) = \frac{\theta^2}{2\gamma} - \frac{c_1 \mu_g^2}{\gamma \sigma_g^2} + \frac{c_2}{2\gamma} \left(\frac{[T(1 + \psi^2) - 2] \sigma_g^2}{T(T - N - 1)} + \mu_g^2 \right), \quad (2.14)$$

by the risk function in (2.8).

It can be seen that the loss increases with N and decreases with T , precisely as for the MV portfolio in Section 2.2.1. By doing estimation analysis, the magnitude of loss can be studied.

An experiment is presented in Table 2.1 where methodology from Kourtis et al (2012) is mimicked. They remove all parametric errors by letting $\boldsymbol{\mu} = k\mathbf{1}_N$ for some constant k . By doing that, all parameters can be expressed in terms of θ^2 and k . It follows that $\mu_g = k$, $\sigma_g^2 = k^2/\theta^2$ and $\psi^2 = 0$. By choosing k and θ^2 for different N and T , the parametric risk is isolated and can be studied without interference from data.

Estimation Risk for $R(\hat{\mathbf{w}}, \mathbf{w}_*)$

Panel A: $\theta^2 = 0.2$					
N	T	Annual $\mu = 5\%$		Annual $\mu = 10\%$	
		$\gamma = 1$	$\gamma = 3$	$\gamma = 1$	$\gamma = 3$
10	60	10.26	3.42	11.00	3.67
	120	10.17	3.39	10.66	3.55
	240	10.15	3.38	10.54	3.51
25	60	10.85	3.62	13.34	4.45
	120	10.28	3.43	11.06	3.69
	240	10.18	3.39	10.67	3.56
50	60	102.84	34.28	381.08	127.03
	120	10.73	3.58	12.85	4.28
	240	10.26	3.42	11.00	3.67
100	120	58.84	19.61	205.11	68.37
	240	10.67	3.56	12.63	4.21

Panel B: $\theta^2 = 0.4$					
N	T	Annual $\mu = 5\%$		Annual $\mu = 10\%$	
		$\gamma = 1$	$\gamma = 3$	$\gamma = 1$	$\gamma = 3$
10	60	20.25	6.75	20.98	6.99
	120	20.17	6.72	20.67	6.89
	240	20.15	6.72	20.56	6.85
25	60	20.80	6.93	23.17	7.72
	120	20.27	6.76	21.06	7.02
	240	20.18	6.73	20.69	6.90
50	60	96.63	32.21	326.37	108.79
	120	20.71	6.90	22.80	7.60
	240	20.26	6.75	21.02	7.01
100	120	63.83	21.28	195.19	65.06
	240	20.67	6.89	22.63	7.54

Table 2.1: Estimation Risk for $R(\hat{\mathbf{w}}, \mathbf{w}_*)$

The table report the expected losses of holding the out-of-sample GMV portfolio instead of the MV portfolio.

The numbers are annualized and in percentages. The losses are computed using (2.14) for different portfolio sizes (N) and sample lengths (T). Two risk aversion coefficients ($\gamma = 1, 3$) are reported, as well as annual return levels 5% and 10%, for square Sharpe ratios 0.2 and 0.4 respectively.

There are several important observations to be drawn from Table 2.1. The first is the confirmation that estimation risk decreases as T increases, which is evident from all cases studied. The second observation is that N is increasing estimation risk, which is natural because the number of parameters that need estimation increases by order N^2 in the covariance matrix as N increases. The third observation is that the ratio N/T is also an important factor. If (N, T) denotes sample dimensions, then it is seen in both Panel A and Panel B, for dimensions (50, 60) and (100, 120), that the loss is much greater than for the other sample dimensions.

2.3 A Well-Conditioned Covariance Shrinkage Estimator

In a series of papers (Ledoit & Wolf (2003, 2004a,b)), a shrinkage estimator for the covariance matrix was developed. The estimator is a linear convex combination between the sample covariance matrix, which is unbiased under normality but suffers from estimation error, and a structured matrix that requires very little estimation but suffers from bias. The optimum combination is shown to balance bias and estimation error asymptotically, and shrink the spatial distance between the estimator and the true covariance matrix, especially when the dimensionality is much greater relative to the sample size⁴. Even if the sample covariance matrix is ill-conditioned, the structured matrix is always invertible and therefore the estimator is also always invertible.

Define the shrinkage estimator $\boldsymbol{\Sigma}_{LW} \in \mathbb{S}_+^{N \times N}$ and let

$$\boldsymbol{\Sigma}_{LW} = \alpha \mathbf{F} + (1 - \alpha) \tilde{\boldsymbol{\Sigma}}, \quad (2.15)$$

where $\tilde{\boldsymbol{\Sigma}} = \{\tilde{\sigma}_{ij}\}$ is defined as in (1.2) and $\mathbf{F} \in \mathbb{S}_{++}^{N \times N}$ is a positive definite structured matrix. α is the shrinkage constant that must be optimized to minimize the shrinkage estimator. The optimal shrinkage constant is found by solving a quadratic loss function under the Frobenius norm, which for a symmetric matrix $\mathbf{Z}_{N \times N}(\mathbb{R})$ is given by

$$\|\mathbf{Z}\|^2 = \text{Trace}(\mathbf{Z}^2) = \sum_{i=1}^N \sum_{j=1}^N z_{ij}^2.$$

This norm is considered to be the distance between the shrinkage estimator, and the true unknown covariance matrix, and is solved by a quadratic loss function

$$L(\alpha) = \|\alpha \mathbf{F} + (1 - \alpha) \tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|^2. \quad (2.16)$$

⁴For stock market indices, it is not uncommon to have several hundreds of securities while the estimation window is 60 months of monthly data.

Now let a risk function $R(\cdot)$ be the expected loss under repeated simulations, that is

$$\begin{aligned}
R(\alpha) &= E[L(\alpha)] \\
&= \sum_{i=1}^N \sum_{j=1}^N E[(\alpha f_{ij} + (1 - \alpha)\tilde{\sigma}_{ij} - \sigma_{ij})^2] \\
&= \sum_{i=1}^N \sum_{j=1}^N \text{Var}[\alpha f_{ij} + (1 - \alpha)\tilde{\sigma}_{ij}] + (E[\alpha f_{ij} + (1 - \alpha)\tilde{\sigma}_{ij} - \sigma_{ij}])^2 \\
&= \sum_{i=1}^N \sum_{j=1}^N \alpha^2 \text{Var}[f_{ij}] + (1 - \alpha)^2 \text{Var}[\tilde{\sigma}_{ij}] \\
&\quad + 2\alpha(1 - \alpha)\text{Cov}[f_{ij}, \tilde{\sigma}_{ij}] + \alpha^2(\phi_{ij} - \sigma_{ij})^2.
\end{aligned} \tag{2.17}$$

The risk function (2.17) must be minimized with respect to α in order to find a minimum point α^* . The first and second order derivatives are

$$R'(\alpha) = 2 \sum_{i=1}^N \sum_{j=1}^N \left(\alpha \text{Var}[f_{ij}] + (1 - \alpha) \text{Var}[\tilde{\sigma}_{ij}] + (1 - 2\alpha) \text{Cov}[f_{ij}, \tilde{\sigma}_{ij}] + \alpha(\phi_{ij} - \sigma_{ij})^2 \right),$$

and

$$R''(\alpha) = 2 \sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}[f_{ij} - \tilde{\sigma}_{ij}] + 2\text{Cov}[f_{ij}, \tilde{\sigma}_{ij}] + (\phi_{ij} - \sigma_{ij})^2 \right),$$

respectively, and it is easy to see that $R''(\alpha) > 0$. The optimal shrinkage constant is found by solving $R'(\alpha) = 0$ for α . This is done by

$$\begin{aligned}
0 &= \sum_{i=1}^N \sum_{j=1}^N \left(\alpha \text{Var}[f_{ij}] + (1 - \alpha) \text{Var}[\tilde{\sigma}_{ij}] + (1 - 2\alpha) \text{Cov}[f_{ij}, \tilde{\sigma}_{ij}] + \alpha(\phi_{ij} - \sigma_{ij})^2 \right) \\
\alpha^* \sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}[f_{ij} - \tilde{\sigma}_{ij}] + (\phi_{ij} - \sigma_{ij})^2 \right) &= \sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}[\tilde{\sigma}_{ij}] - \text{Cov}[f_{ij}, \tilde{\sigma}_{ij}] \right) \\
\alpha^* &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}[\tilde{\sigma}_{ij}] - \text{Cov}[f_{ij}, \tilde{\sigma}_{ij}] \right)}{\sum_{i=1}^N \sum_{j=1}^N \left(\text{Var}[f_{ij} - \tilde{\sigma}_{ij}] + (\phi_{ij} - \sigma_{ij})^2 \right)}.
\end{aligned} \tag{2.18}$$

Denote the asymptotic variance of the sample covariance matrix by AsyVar , and the asymptotic covariance of the sample covariance matrix, with the target matrix by AsyCov . Ledoit & Wolf (2003, Theorem 1) lets $\pi = \sum_{i=1}^N \sum_{j=1}^N \text{AsyVar}[\sqrt{T}\tilde{\sigma}_{ij}]$, $\rho = \sum_{i=1}^N \sum_{j=1}^N \text{AsyCov}[\sqrt{T}f_{ij}, \tilde{\sigma}_{ij}]$ and $\gamma = \sum_{i=1}^N \sum_{j=1}^N (\phi_{ij} - \sigma_{ij})^2$, and shows that (2.18) can be written as

$$\alpha^* = \frac{1}{T} \frac{\pi - \rho}{\gamma} + O\left(\frac{1}{T^2}\right). \tag{2.19}$$

They show that the shrinkage intensity asymptotically converges in distribution, to a constant $\kappa = (\pi - \rho)/\gamma$.

2.3.1 Consistent Estimators to the Shrinkage Constant

The components of (2.18) must be estimated from the data sample. Starting with π , Ledoit & Wolf (2003, Lemma 1) defines a consistent estimator $\hat{\pi}$ such that

$$\hat{\pi}_{ij} = \frac{1}{T} \sum_{t=1}^T [(x_{it} - \tilde{\mu}_i)(x_{jt} - \tilde{\mu}_j) - \tilde{\sigma}_{ij}]^2, \quad (2.20)$$

where x are returns from sample i and j in observation t and $\tilde{\mu}$ is a mean return over the sample. They prove that (2.20) converges in probability to π .

Next, Ledoit & Wolf (2003, Lemma 2) define the sample correlation as

$$r_{ij} = \frac{\tilde{\sigma}_{ij}}{\sqrt{\tilde{\sigma}_{ii}\tilde{\sigma}_{jj}}}, \quad \text{and} \quad \bar{r} = \frac{2}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N r_{ij}, \quad (2.21)$$

using (2.21), and give a consistent estimator for ρ in the form of

$$\hat{\rho} = \sum_{i=1}^N \hat{\pi}_{ii} + \sum_{i=1}^N \sum_{j \neq i}^N \frac{\bar{r}}{2} \left(\sqrt{\frac{\tilde{\sigma}_{ii}}{\tilde{\sigma}_{jj}}} \hat{\theta}_{ii,ij} + \sqrt{\frac{\tilde{\sigma}_{jj}}{\tilde{\sigma}_{ii}}} \hat{\theta}_{jj,ij} \right), \quad (2.22)$$

where

$$\hat{\theta}_{ii,ij} = \frac{1}{T} \sum_{t=1}^T ((y_{it} - \bar{y}_i)^2 - \tilde{\sigma}_{ii}) ((y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - \tilde{\sigma}_{ij})$$

$$\hat{\theta}_{jj,ij} = \frac{1}{T} \sum_{t=1}^T ((y_{jt} - \bar{y}_j)^2 - \tilde{\sigma}_{jj}) ((y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j) - \tilde{\sigma}_{ij}).$$

The third and final consistent estimator, the sample analogous to γ , is

$$\hat{\gamma} = \sum_{i=1}^N \sum_{j \neq i}^N (f_{ij} - \tilde{\sigma}_{ij})^2, \quad (2.23)$$

as proved in Ledoit & Wolf (2003, Lemma 3). By substituting the consistent estimators, $\hat{\alpha}^* = \hat{\kappa}/T = (\hat{\pi} - \hat{\rho})/(T\hat{\gamma})$, and a practical estimator to (2.19) exists. To ensure that the estimator is well-behaved a truncation is made such that

$$\hat{\alpha}^* = \max \{0, \min \{ \hat{\alpha}^*, 1 \} \}, \quad (2.24)$$

to ensure that the shrinkage constant is bound by the interval $[0, 1]$.

2.3.2 Target Matrices

In the previous section, the optimal linear combination of two estimators was derived without any information about the target matrix. The target matrix $\mathbf{F} = \{f_{ij}\}$ is chosen so as to balance estimation error with bias. \mathbf{F} is on purpose heavily biased, as it is composed with a low number of degrees of freedom. The properties required to maintain validity of Equation (2.18) is that (1) it must be positive definite symmetric (to be strictly non-singular), (2) highly structured, that is to contain very little estimation error, and (3) be an asymptotically biased estimator of the sample covariance matrix.

The structured matrix is referred to as the shrinkage target. Three targets have been evaluated in the literature. The first is the identity matrix scaled by the average sample variance. Let the target be $\mathbf{F} = v\mathbf{I}_{N \times N}$ with $v = \text{tr}(\hat{\Sigma}, \mathbf{I})/N$ being the trace of the sample covariance matrix with the identity matrix, divided by the dimension. The scale factor can be seen as a “grand variance”, a purposely naive prior assumption that all assets have the same variance and no covariance. The estimator will be shrunk towards this target with an estimated optimal ratio. For this target, the asymptotic covariance term ρ in (2.19) is zero since all off-diagonal elements of the target matrix is zero. It follows that the constant of the estimated shrinkage intensity is reduced to $\kappa = \pi/\gamma$.

The second target is the one-factor model covariance matrix. Using linear regression, where $\mathbf{b} = \{b_i\}$ is the vector of slopes, $\tilde{\sigma}_m^2$ is the market variance and $\mathbf{D} = \{d_{ii}\}$ is the residual variance matrix⁵ with target matrix is given by $\mathbf{F} = \tilde{\sigma}_m^2 \mathbf{b}\mathbf{b}' + \mathbf{D}$. The number of parameters estimated in this matrix are far less than the sample covariance matrix, especially in large dimensions.

The third investigated target matrix is the covariance matrix of the constant correlation model. Denote the sample covariance matrix $\Sigma = \{\sigma_{ij}\}$, and denote the correlation matrix $\Phi = \{\phi_{ij}\}$ where $\phi_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. Then the constant correlation factor is given by

$$\bar{r} = \frac{2}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \phi_{ij},$$

and $\mathbf{F} = \{f_{ij}\}$ where $f_{ij} = \bar{r}\sqrt{\sigma_{ii}\sigma_{jj}}$, and $f_{ii} = \sigma_{ii}$.

⁵The one-factor model assumes that $\text{Cov}(d_i, d_j) = 0 \forall i, j$, hence the matrix is diagonal.

Chapter 3

PORTFOLIO RULES

A central part of Modern Portfolio Theory is the *Two-Fund Separation Theorem*, which suggests that the investor should divide wealth between risky funds and fixed income assets to balance the total risk exposure. A recent academic development in portfolio selection has greatly enriched the combination of two or more portfolios which adds dimensionality to the traditional approach. Kan & Zhou (2007) propose a three-fund portfolio consisting of the class of risk-free¹ assets, the sample GMV portfolio and the tangency portfolio. The intuition is that while both portfolios are prone to estimation error, the errors of respective portfolio are not perfectly correlated and thus, a combination of the two portfolios (or perhaps two other portfolios) diversifies estimation error. Demiguel et al (2009) are the first to suggest the equally weighted portfolio combined with the minimum variance portfolio, which belongs to the class of combined rules related to this thesis. Tu & Zhou (2011) extend the work on rules which combines the equally weighted rule with four different sophisticated portfolios. Their results show that combined portfolios improve greatly in relation to their respective components alone.

This chapter will start off by the proposition of a new combination rule between two distinct portfolio rules in Section 3.1, where the estimated optimal weight between them is analytically derived after an approximation is made. Section 3.2 will discuss already developed combined rules viable for comparison in performance tests.

3.1 The Proposed Combined Rule

Let the portfolio rule be a combination of the EW rule and a sophisticated rule, as given by Tu & Zhou (2011). The sophisticated rule is estimated from the shrinkage method of the covariance matrix, proposed by Ledoit & Wolf (2004b). That gives the combined rule

$$\tilde{\mathbf{w}}_C = (1 - \delta)\mathbf{w}_{EW} + \delta\tilde{\mathbf{w}}_{LW}, \quad (3.1)$$

where $\delta \in [0, 1]$ is the weight assigned to the sophisticated rule. Tu & Zhou (2011, Proposition 1) show that portfolio rules from the class that (3.1) belongs to, have convex optimums, that

¹Supposedly risk-free, as even governments fall these days.

is there exist a unique δ that maximize utility. The goal is to find a unique optimal δ^* such that the combined rule $\tilde{\mathbf{w}}_C$ dominates both \mathbf{w}_{EW} and the sophisticated rule . Let the estimated covariance matrix given by Ledoit & Wolf (2004b) be denoted by $\tilde{\Sigma}_{LW}$, then the unconstrained portfolio rule $\tilde{\mathbf{w}}_{LW}$ associated with the covariance matrix is

$$\tilde{\mathbf{w}}_{LW} = \frac{1}{\gamma} \tilde{\Sigma}_{LW}^{-1} \tilde{\boldsymbol{\mu}}, \quad (3.2)$$

where $\tilde{\boldsymbol{\mu}}$ is the ML estimator, and $\tilde{\Sigma}_{LW}^{-1}$ is the inverse of (2.15), provided that it exists.

Using the rule in (3.1), the utility in (1.7), and applying the proof to Tu & Zhou (2011, Proposition 3), the risk function of the proposed rule is given by

$$\begin{aligned} R(\mathbf{w}_*, \tilde{\mathbf{w}}_C) &= \frac{\gamma}{2} \mathbb{E} \left[\left\{ (1 - \delta)(\mathbf{w}_{EW} - \mathbf{w}_*) + \delta(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*) \right\}' \Sigma \left\{ (1 - \delta)(\mathbf{w}_{EW} - \mathbf{w}_*) + \delta(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*) \right\} \right] \\ &= \frac{\gamma}{2} \mathbb{E} \left[\left\{ (1 - \delta)\mathbf{a} + \delta\mathbf{b} \right\}' \Sigma \left\{ (1 - \delta)\mathbf{a} + \delta\mathbf{b} \right\} \right], \end{aligned} \quad (3.3)$$

where $\mathbf{a} = \mathbf{w}_{EW} - \mathbf{w}_*$ and $\mathbf{b} = \tilde{\mathbf{w}}_{LW} - \mathbf{w}_*$. The expression inside the expectation in (3.3) can be further manipulated. Let $f(\delta)$ denote the differentiable function of the expression inside the expectation, and expand it by

$$\begin{aligned} f(\delta) &= \left\{ (1 - \delta)\mathbf{a} + \delta\mathbf{b} \right\}' \Sigma \left\{ (1 - \delta)\mathbf{a} + \delta\mathbf{b} \right\} \\ &= \left\{ (1 - \delta)\mathbf{a}'\Sigma + \delta\mathbf{b}'\Sigma \right\} \left\{ (1 - \delta)\mathbf{a} + \delta\mathbf{b} \right\} \\ &= (1 - \delta)^2 \mathbf{a}'\Sigma\mathbf{a} + 2(1 - \delta)\delta \mathbf{a}'\Sigma\mathbf{b} + \delta^2 \mathbf{b}'\Sigma\mathbf{b} \\ &= \mathbf{a}'\Sigma\mathbf{a} - 2\delta \mathbf{a}'\Sigma\mathbf{a} + \delta^2 \mathbf{a}'\Sigma\mathbf{a} + 2\delta \mathbf{a}'\Sigma\mathbf{b} - 2\delta^2 \mathbf{a}'\Sigma\mathbf{b} + \delta^2 \mathbf{b}'\Sigma\mathbf{b}. \end{aligned} \quad (3.4)$$

In order to find the interior optimum (that is δ_*) between the EW rule and the LW rule, the first order condition is set equal to zero and solved for δ as follows,

$$\begin{aligned} \frac{df(\delta)}{d\delta} &= -2\mathbf{a}'\Sigma\mathbf{a} + 2\delta \mathbf{a}'\Sigma\mathbf{a} + 2\mathbf{a}'\Sigma\mathbf{b} - 4\delta \mathbf{a}'\Sigma\mathbf{b} + 2\delta \mathbf{b}'\Sigma\mathbf{b}, \\ &\implies -\mathbf{a}'\Sigma\mathbf{a} + \delta \mathbf{a}'\Sigma\mathbf{a} + \mathbf{a}'\Sigma\mathbf{b} - 2\delta \mathbf{a}'\Sigma\mathbf{b} + \delta \mathbf{b}'\Sigma\mathbf{b} = 0, \\ &\iff \delta (\mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'\Sigma\mathbf{b} + \mathbf{b}'\Sigma\mathbf{b}) = \mathbf{a}'\Sigma\mathbf{a} - \mathbf{a}'\Sigma\mathbf{b}, \\ &\iff \delta_* = \frac{\mathbf{a}'\Sigma\mathbf{a} - \mathbf{a}'\Sigma\mathbf{b}}{\mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'\Sigma\mathbf{b} + \mathbf{b}'\Sigma\mathbf{b}}, \end{aligned} \quad (3.5)$$

Note that (3.5) has three unique terms, those who contain only \mathbf{a} , those which contain both \mathbf{a} and \mathbf{b} , and a term containing only \mathbf{b} . After substituting back, the first term $(\mathbf{w}_{EW} - \mathbf{w}_*)'\Sigma(\mathbf{w}_{EW} - \mathbf{w}_*)$ measures the impact of bias from the EW rule, imposed on the combined rule. The second term, $(\mathbf{w}_{EW} - \mathbf{w}_*)'\Sigma(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)$ measures the misspecification between the true rule and the LW rule, and the third term, $(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)'\Sigma(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)$ measures the impact of variance from the LW rule.

Even though δ_* is designed to balance bias and variance from two rules, the expression in (3.5) is not of much use in its current form. The first involves no expectations or estimators

and is thus straight forward to evaluate. Being the only term in common for all rules the result by Tu & Zhou (2011) is verified here, thus it follows that

$$\begin{aligned}
\mathbf{a}'\Sigma\mathbf{a} &= \eta_1 = (\mathbf{w}_{EW} - \mathbf{w}_*)'\Sigma(\mathbf{w}_{EW} - \mathbf{w}_*) \\
&= (\Sigma\mathbf{w}_{EW} - \frac{1}{\gamma}\boldsymbol{\mu})'(\mathbf{w}_{EW} - \mathbf{w}_*) \\
&= \mathbf{w}_{EW}'\Sigma\mathbf{w}_{EW} - \frac{2}{\gamma}\mathbf{w}_{EW}'\boldsymbol{\mu} + \frac{\theta^2}{2\gamma}.
\end{aligned} \tag{3.6}$$

A consistent estimator is the sample equivalent $\tilde{\eta}_1$, that is

$$\tilde{\eta}_1 = \mathbf{w}_{EW}'\tilde{\Sigma}\mathbf{w}_{EW} - \frac{2}{\gamma}\mathbf{w}_{EW}'\tilde{\boldsymbol{\mu}} + \frac{\tilde{\theta}^2}{2\gamma}. \tag{3.7}$$

The estimator $\tilde{\theta}^2$ is given as by Kan & Zhou (2007). The other two terms contain expectations when they are substituted back into (3.5), that must be solved. Starting with the mixed term, using (3.2) yields

$$\begin{aligned}
\mathbf{a}'\Sigma\mathbf{E}[\mathbf{b}] &= \eta_{13} = (\mathbf{w}_{EW} - \mathbf{w}_*)'\Sigma\mathbf{E}[(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)] \\
&= (\mathbf{w}_{EW}\Sigma - \frac{1}{\gamma}\boldsymbol{\mu})'\mathbf{E}[(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)] \\
&= \mathbf{w}_{EW}'\Sigma\mathbf{E}[\tilde{\mathbf{w}}_{LW}] - \mathbf{w}_{EW}'\Sigma\mathbf{w}_* - \frac{1}{\gamma^2}\boldsymbol{\mu}'\mathbf{E}[\tilde{\mathbf{w}}_{LW}] + \frac{1}{\gamma^2}\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} \\
&= \frac{1}{\gamma} \left(\mathbf{w}_{EW}'\Sigma\mathbf{E}[\Sigma_{LW}^{-1}]\boldsymbol{\mu} - \mathbf{w}_{EW}'\boldsymbol{\mu} - \frac{1}{\gamma}\mathbf{E}[\boldsymbol{\mu}'\Sigma_{LW}^{-1}\tilde{\boldsymbol{\mu}}] + \frac{\theta^2}{\gamma} \right).
\end{aligned} \tag{3.8}$$

Note that if the inverse ML estimator in (1.2) scaled by $(T - N - 2)/T$, would be substituted for the LW covariance estimator, the whole expression would be equal to zero. The second term that balance the variance from the LW rule, is evaluated as

$$\begin{aligned}
\mathbf{E}[\mathbf{b}'\Sigma\mathbf{b}] &= \eta_3 = \mathbf{E}[(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)'\Sigma(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)] \\
&= \mathbf{E}[(\tilde{\mathbf{w}}_{LW}\Sigma - \frac{1}{\gamma}\boldsymbol{\mu})'(\tilde{\mathbf{w}}_{LW} - \mathbf{w}_*)] \\
&= \mathbf{E}[\tilde{\mathbf{w}}_{LW}'\Sigma\tilde{\mathbf{w}}_{LW} - \tilde{\mathbf{w}}_{LW}'\Sigma\mathbf{w}_* - \frac{1}{\gamma}\boldsymbol{\mu}'\tilde{\mathbf{w}}_{LW} + \frac{1}{\gamma}\boldsymbol{\mu}'\mathbf{w}_*] \\
&= \mathbf{E}[\tilde{\mathbf{w}}_{LW}'\Sigma\tilde{\mathbf{w}}_{LW}] - \frac{2}{\gamma^2}\mathbf{E}[\boldsymbol{\mu}'\Sigma_{LW}^{-1}\tilde{\boldsymbol{\mu}}] + \frac{\theta^2}{\gamma^2} \\
&= \frac{1}{\gamma^2}\mathbf{E}[\tilde{\boldsymbol{\mu}}'\Sigma_{LW}^{-1}\Sigma\Sigma_{LW}^{-1}\tilde{\boldsymbol{\mu}}] - \frac{2}{\gamma^2}\mathbf{E}[\boldsymbol{\mu}'\Sigma_{LW}^{-1}\tilde{\boldsymbol{\mu}}] + \frac{\theta^2}{\gamma^2}.
\end{aligned} \tag{3.9}$$

3.1.1 Tractable Solutions for the Expectations

The expectations in (3.8) and (3.9) are intractable to solve analytically due to the complexity of the estimator. Luckily there is a lemma known as the *Matrix Inversion Lemma* (see for example Boyd & Vandenberghe (2004, Appendix C.4.3)), that allows for an approximation which solves the expectations. A special version of the lemma states that

Lemma 3.1 (The Matrix Inversion Lemma). *Given a non-singular matrix $\mathbf{A} \in \mathbb{R}_{n \times n}$, and a real matrix $\mathbf{B} \in \mathbb{R}_{n \times n}$, the identity for the inverse of a sum of two matrices is*

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{B} + \mathbf{B} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{B} \mathbf{A}^{-1},$$

or in reduced form by

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} (\mathbf{I} + \mathbf{B} \mathbf{A}^{-1})^{-1} \mathbf{B} \mathbf{A}^{-1},$$

using the identity $(\mathbf{A} \mathbf{B})^{-1} = \mathbf{A}^{-1} \mathbf{B}^{-1}$ if both \mathbf{A} and \mathbf{B} are invertible.

Now using the reduced form in Lemma 3.1, let $\mathbf{A} = (1 - \alpha) \tilde{\Sigma}^{-1}$ and $\mathbf{B} = \alpha \mathbf{F}$. \mathbf{F} is always invertible by definition and assume that $\tilde{\Sigma}$ is non-singular. Then the inverse can be expressed as

$$\begin{aligned} (\alpha \mathbf{F} + (1 - \alpha) \tilde{\Sigma})^{-1} &= \frac{\tilde{\Sigma}^{-1}}{1 - \alpha} - \frac{\tilde{\Sigma}^{-1}}{1 - \alpha} \left(\mathbf{I} + \alpha \mathbf{F} \frac{\tilde{\Sigma}^{-1}}{1 - \alpha} \right)^{-1} \alpha \mathbf{F} \frac{\tilde{\Sigma}^{-1}}{1 - \alpha}, \\ &= \frac{\tilde{\Sigma}^{-1}}{1 - \alpha} - \mathbf{P}. \end{aligned} \quad (3.10)$$

Simulations made in preparation for this section showed that $\|\mathbf{P}\| \ll \|\frac{\tilde{\Sigma}^{-1}}{1 - \alpha}\|$ for all $\alpha \in [0, 1]$.² This enables the same approximation trick as used by Tu & Zhou (2011, Appendix B), that is to let a proportion of the estimator be constant, which in this case is \mathbf{P} .

The result in (3.10) is enough to solve the first approximate expectation. First,

$$\begin{aligned} \mathbb{E}[\boldsymbol{\mu} \tilde{\Sigma}_{LW}^{-1} \tilde{\boldsymbol{\mu}}] &= \mathbb{E} \left[\boldsymbol{\mu}' \left(\frac{\tilde{\Sigma}^{-1}}{1 - \alpha} - \mathbf{P} \right) \tilde{\boldsymbol{\mu}} \right] \\ &\approx \frac{c_1 \theta^2}{(1 - \alpha)} - \boldsymbol{\mu}' \mathbf{P} \boldsymbol{\mu}, \quad T > N + 2, \end{aligned} \quad (3.11)$$

by using (2.3) and $c_1 = T / (T - N - 2)$.

For the second expectation, a partial result using (3.10) is,

$$\begin{aligned} \tilde{\Sigma}_{LW}^{-1} \boldsymbol{\Sigma} \tilde{\Sigma}_{LW}^{-1} &= \left(\frac{\tilde{\Sigma}^{-1}}{1 - \alpha} - \mathbf{P} \right) \boldsymbol{\Sigma} \left(\frac{\tilde{\Sigma}^{-1}}{1 - \alpha} - \mathbf{P} \right) \\ &= \frac{\tilde{\Sigma}^{-1} \boldsymbol{\Sigma} \tilde{\Sigma}^{-1}}{(1 - \alpha)^2} - \frac{\tilde{\Sigma}^{-1} \boldsymbol{\Sigma} \mathbf{P}}{1 - \alpha} - \frac{\mathbf{P} \boldsymbol{\Sigma} \tilde{\Sigma}^{-1}}{1 - \alpha} + \mathbf{P} \boldsymbol{\Sigma} \mathbf{P}. \end{aligned}$$

²More precisely, \mathbf{P} is about one millionth the size of the first term in the norm.

The first term requires the inverse Wishart identity from Section 2.1.1. Using the identity,

$$\begin{aligned} \mathbb{E} \left[\tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} \right] &= \mathbb{E} \left[\left(\Sigma^{-\frac{1}{2}} \mathbf{W}^{-1} \Sigma^{-\frac{1}{2}} \right) \Sigma \left(\Sigma^{-\frac{1}{2}} \mathbf{W}^{-1} \Sigma^{-\frac{1}{2}} \right) \right], \\ &= \Sigma^{-\frac{1}{2}} \mathbb{E} \left[\mathbf{W}^{-2} \right] \Sigma^{-\frac{1}{2}}, \\ &= c_2 \Sigma^{-1}, \end{aligned}$$

with $c_2 = T^2(T-2)/[(T-N-1)(T-N-2)(T-N-4)]$ from (2.4). The second term is solved using (2.3) and the third term is trivial. Now, all the results required to solve the second expectation in η_3 are attained, and by using the assumption of independence between $\tilde{\mu}$ and $\tilde{\Sigma}$, it equals

$$\mathbb{E} \left[\tilde{\mu}' \tilde{\Sigma}_{LW}^{-1} \Sigma \tilde{\Sigma}_{LW}^{-1} \tilde{\mu} \right] = c_2 \frac{N+T\theta^2}{T(1-\alpha)^2} - c_1 2 \frac{N+T\mu' \mathbf{P} \mu}{T(1-\alpha)} + \frac{N+T\mu' \mathbf{P} \Sigma \mathbf{P} \mu}{T}. \quad (3.12)$$

3.1.2 The Optimal Combination Constant

Now that all the partial results have been derived, the estimated optimal constant for the linear convex combination in (3.1) can be solved. The first term, η_1 , was evaluated in the previous section. The second term, η_{13} , is constructed by substituting (3.10) and (3.11) into (3.8). It results in

$$\eta_{13} = \frac{1}{\gamma} \left(\frac{c_1}{1-\alpha} \left(\mathbf{w}_{EW}' \mu - \mathbf{w}_{EW}' \Sigma \mathbf{P} \mu - \frac{\theta^2}{\gamma} \right) + \frac{1}{\gamma} \mu' \mathbf{P} \mu - \mathbf{w}_{EW}' \mu + \frac{\theta^2}{\gamma} \right), \quad (3.13)$$

with c_1 given as in the previous section. A consistent estimator for (3.13) is its sample analogy,

$$\tilde{\eta}_{13} = \frac{1}{\gamma} \left(\frac{c_1}{1-\alpha} \left(\mathbf{w}_{EW}' \tilde{\mu} - \mathbf{w}_{EW}' \tilde{\Sigma} \mathbf{P} \tilde{\mu} - \frac{\tilde{\theta}^2}{\gamma} \right) + \frac{1}{\gamma} \tilde{\mu}' \mathbf{P} \tilde{\mu} - \mathbf{w}_{EW}' \tilde{\mu} + \frac{\tilde{\theta}^2}{\gamma} \right). \quad (3.14)$$

The last term, η_3 , is evaluated using the results in (3.11) and (3.12), which substitutes into (3.9) as

$$\eta_3 = \frac{c_2}{\gamma^2} \frac{N+T\theta^2}{T(1-\alpha)^2} - c_1 \frac{N+T\mu' \mathbf{P} \mu}{T(1-\alpha)} + \frac{N+T\mu' \mathbf{P} \Sigma \mathbf{P} \mu}{T} - \frac{2c_1\theta^2}{\gamma^2(1-\alpha)} + \mu' \mathbf{P} \mu + \frac{\theta^2}{\gamma^2}. \quad (3.15)$$

Similarly to η_{13} , the consistent estimator to η_3 is the sample counterpart,

$$\tilde{\eta}_3 = \frac{c_2}{\gamma^2} \frac{N+T\tilde{\theta}^2}{T(1-\alpha)^2} - c_1 \frac{N+T\tilde{\mu}' \mathbf{P} \tilde{\mu}}{T(1-\alpha)} + \frac{N+T\tilde{\mu}' \mathbf{P} \tilde{\Sigma} \mathbf{P} \tilde{\mu}}{T} - \frac{2c_1\tilde{\theta}^2}{\gamma^2(1-\alpha)} + \tilde{\mu}' \mathbf{P} \tilde{\mu} + \frac{\tilde{\theta}^2}{\gamma^2}. \quad (3.16)$$

It follows from equations (3.7), (3.14) and (3.16) that an approximation for (3.5) is given by

$$\tilde{\delta}_* = \frac{\tilde{\eta}_1 - \tilde{\eta}_{13}}{\tilde{\eta}_1 - 2\tilde{\eta}_{13} + \tilde{\eta}_3}, \quad (3.17)$$

and the result is summarized in the following proposition.

Proposition 2. *Assume that $T > N + 4$ such that the second moment of the inverse Wishart distribution exists. Then there exist a combination of \mathbf{w}_{EW} and \mathbf{w}_{LW} , given by $\mathbf{w}_C = (1 - \delta)\mathbf{w}_{EW} + \delta\mathbf{w}_{LW}$, such that the estimated optimum is*

$$\mathbf{w}_C = (1 - \tilde{\delta}_*)\mathbf{w}_{EW} + \tilde{\delta}_*\mathbf{w}_{LW},$$

where $\tilde{\delta}_* = (\tilde{\eta}_1 - \tilde{\eta}_{13}) / (\tilde{\eta}_1 - 2\tilde{\eta}_{13} + \tilde{\eta}_3)$ and $\tilde{\eta}_1$, $\tilde{\eta}_{13}$ and $\tilde{\eta}_3$ are given by (3.7), (3.14) and (3.16) respectively.

The rule proposed in this paper will be compared to other established portfolio rules that utilize combination with the EW rule. Such rules will be presented in the following section.

3.2 Other Combined Rules

This section will provide details for the portfolio rules mentioned briefly in the introduction to this chapter. The discussion will be limited to rules related to this thesis, that is rules that combine the equally weighted portfolio with a theoretically sophisticated portfolio. The combination of the equally weighted rule and that of MacKinlay & Pástor (2000) is not considered due to large estimation errors for the optimal coefficient between the rules, as discussed by Tu & Zhou (2011). They instead propose to weigh the coefficient equally at 50% but no attempt at doing so will be made in thesis.

3.2.1 The Traditional Markowitz Rule

The most trivial rule in the class is the traditional rule mentioned³ by Demiguel et al (2009) and analytically derived by Tu & Zhou (2011). The rule is

$$\mathbf{w}_1 = (1 - \delta)\mathbf{w}_{EW} + \delta\hat{\mathbf{w}},$$

where $\hat{\mathbf{w}}$ is a scaled version of the one given in (1.6). This rule is considered the most trivial because when taking the expectation of the risk function associated with this rule, the mixed term equals zero. The reason is a scaled version of the ML estimator which cancels the terms, as discussed briefly in Section 3.1 when the mixed term was derived. Since the mixed term is zero by construction, the estimated delta reduces to

$$\tilde{\delta}_{TM} = \frac{\tilde{\eta}_1}{\tilde{\eta}_1 + \tilde{\pi}_2},$$

where the notation $\tilde{\pi}$ is kept from the original paper, if not stated elsewhere in this thesis. $\tilde{\eta}_1$ is given by (3.7), and

$$\tilde{\pi}_2 = (c_3 - 1)\tilde{\theta}^2 + \frac{c_3 N}{\gamma^2 T},$$

where the constant $c_3 = (T - 2)(T - N - 2) / ((T - N - 1)(T - N - 4))$. An important conclusion was drawn about this rule by Tu & Zhou (2011, Proposition 1). If $\eta_1 > 0$, which it

³Demiguel et al (2009) considered a normalized version for relevant comparison with other rules.

practically is because the equally weighted portfolio is positively biased relative to the true portfolio, the loss in holding \mathbf{w}_1 is strictly less than holding any of the two components alone. It also follows from the proposition that because $0 < \tilde{\delta}_{TM} < 1$, the rule diversifies between estimation error and bias. Moreover, this rule is a special case of the proposed rule in Proposition 2. It can be seen by scaling the LW estimator by c_1 and setting $\alpha = 0$.

Proposition 3. *Assume that $T > N + 4$ such that the moments of the inverse Wishart distribution exist. Denote $\hat{\Sigma} = c_1 \tilde{\Sigma}$ such that $\tilde{\Sigma}$ is an unbiased estimator of Σ and let,*

$$\hat{\Sigma}_{LW}^{-1} = \left(\alpha \mathbf{F} + (1 - \alpha) \hat{\Sigma} \right)^{-1}.$$

Further let

$$\hat{\mathbf{w}}_C = (1 - \tilde{\delta}_*) \mathbf{w}_{EW} + \tilde{\delta}_* \hat{\mathbf{w}}_{LW},$$

be the combined rule between the equally weighted and adjusted LW rule, then $\hat{\mathbf{w}}_{LW} = \hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_C = \mathbf{w}_1$ when $\alpha = 0$.

The proof of proposition 3 is found in Appendix A.2.

3.2.2 The Kan & Zhou (2007) Combined Rule

This rule is more interesting because it serves as a combination of three rules, the equally weighted, the sample GMV portfolio and the tangency portfolio. This three-fund rule follows the linear combination

$$\mathbf{w}_2 = (1 - \delta) \mathbf{w}_{EW} + \delta \hat{\mathbf{w}}_{KZ},$$

where

$$\hat{\mathbf{w}}_{KZ} = \hat{\mathbf{w}}(\alpha_{KZ}) = \frac{1}{\gamma} \left(\alpha_{KZ} \hat{\Sigma}^{-1} \tilde{\boldsymbol{\mu}} + (1 - \alpha_{KZ}) \tilde{\mu}_g \hat{\Sigma}^{-1} \mathbf{1}_N \right).$$

The estimated weight of α_{KZ} is given by

$$\alpha_{KZ}^* = \frac{1}{c_3} \left(\frac{\hat{\psi}^2}{\hat{\psi}^2 + \frac{N}{T}} \right),$$

with $\hat{\psi}^2$ being an estimator of the square of the variable in Section 2.2.2, given by Kan & Zhou (2007, Equation (66)). The estimated optimal combination that maximizes the expected utility is

$$\tilde{\delta}_{KZ} = \frac{\tilde{\eta}_1 - \hat{\pi}_{13}}{\tilde{\eta}_1 - 2\hat{\pi}_{13} + \hat{\pi}_3},$$

with

$$\hat{\pi}_{13} = \frac{\tilde{\theta}^2}{\gamma^2} - \frac{1}{\gamma} \mathbf{w}'_{EW} \tilde{\boldsymbol{\mu}} + \frac{1}{c_2 \gamma} \left[\left(\alpha_{KZ}^* \mathbf{w}'_{EW} \tilde{\boldsymbol{\mu}} + (1 - \alpha_{KZ}^*) \tilde{\mu}_g \mathbf{w}'_{EW} \mathbf{1}_N \right) - \frac{1}{\gamma} \left(\alpha_{KZ}^* \tilde{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \tilde{\boldsymbol{\mu}} + (1 - \alpha_{KZ}^*) \tilde{\mu}_g \tilde{\boldsymbol{\mu}}' \hat{\Sigma}^{-1} \mathbf{1}_N \right) \right],$$

and,

$$\hat{\pi}_3 = \frac{\tilde{\theta}^2}{\gamma^2} - \frac{1}{c_2 \gamma^2} \left(\tilde{\theta}^2 - \frac{N}{T} \alpha_{KZ}^* \right).$$

The covariance estimator is the scaled ML estimator and $\tilde{\mu}_g$ is the expected return on the sample GMV portfolio. Confer primarily Tu & Zhou (2011), but also Kan & Zhou (2007) for a more detailed treatment of this rule.

3.2.3 The Jorion (1986) Combined Rule

The Bayes–Stein rule considered by Philippe Jorion, based on the statistically superior shrinkage approach by James & Stein (1961), avalanched a new class of portfolio rules that considered shrinking estimators toward targets with fewer to no free parameters. Tu & Zhou (2011) derived an approximation for the true combined rule, since both the mean and covariance estimator of the Jorion rule are linear combinations of estimators, and thus impose a problem for the tractability of the analytical expected values. However the published version of their article lacks details necessary for replication of their study and thus, this section will be richer as a remedy. The problem is treated the same for the covariance estimator in Section 3.1.1 of this thesis.

The rule follows the linear combination

$$\mathbf{w}_3 = (1 - \delta) \mathbf{w}_{EW} + \delta \hat{\mathbf{w}}_{PJ}, \quad \text{with} \quad \hat{\mathbf{w}}_{PJ} = \frac{1}{\gamma} \hat{\Sigma}_{PJ}^{-1} \boldsymbol{\mu}_{PJ},$$

where

$$E[\boldsymbol{\mu}_{PJ}] = (1 - w) \boldsymbol{\mu} + w \mu_g \mathbf{1}_N, \quad \text{and} \quad w = \frac{\lambda}{\lambda + T}.$$

The variable $\lambda \in [0, 1]$ is an empirical Bayes estimator balancing the estimator and its probability density function, $P(\lambda | \boldsymbol{\mu}, \mu_g, \boldsymbol{\Sigma}) \sim \Gamma(N + 2, d)$, follows a gamma distribution with shape parameter $N + 2$ and rate parameter d . The sample estimator for $\boldsymbol{\mu}_{PJ}$ is given by

$$\hat{\boldsymbol{\mu}}_{PJ} = (1 - \hat{w}) \tilde{\boldsymbol{\mu}} + \hat{w} \tilde{\mu}_g \mathbf{1}_N, \quad \text{where} \quad \hat{w} = \frac{N + 2}{(N + 2) + d}.$$

The mean estimator is shrunk toward a grand mean, the sample mean of the ex ante global minimum variance portfolio with a shrinkage factor $\hat{w} \in [0, 1]$, where $d = (\tilde{\boldsymbol{\mu}} - \mu_g \mathbf{1}_N)' T \hat{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{\mu}} - \mu_g \mathbf{1}_N)$. The covariance estimator is given by

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{PJ} &= \hat{\boldsymbol{\Sigma}} \left(1 + \frac{1}{T} \right) + \frac{\lambda}{T + 1 + \lambda} \frac{\mathbf{1}_N \mathbf{1}_N'}{\mathbf{1}_N' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_N} \\ &= \hat{\boldsymbol{\Sigma}} \left(1 + \frac{1}{T} \right) + \mathbf{B}, \end{aligned}$$

with $\mathbf{B} = \{b_{ij}\}$ being the second term. When evaluating the inverse covariance estimator, Tu & Zhou (2011) treat the second term as a matrix of constants $\mathbf{C} = \{c_{ij}\}$, such that

$$\hat{\boldsymbol{\Sigma}}_{PJ}^{-1} = \hat{\boldsymbol{\Sigma}}^{-1} c_4 + \mathbf{C},$$

where $c_4 = T/(T + 1)$, using Lemma 3.1. In analogy to (3.13), the mixed term of the Jorion rule is given by

$$\begin{aligned}\eta_{13}^{PJ} &= \frac{1}{\gamma} \left(\mathbf{w}_{EW}' \boldsymbol{\Sigma} E \left[\widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} \boldsymbol{\mu}_{PJ} \right] - \mathbf{w}_{EW}' \boldsymbol{\mu} - \frac{1}{\gamma} E \left[\boldsymbol{\mu} \widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} \boldsymbol{\mu}_{PJ} \right] + \theta^2 \right) \\ &= \frac{1}{\gamma} \left(\mathbf{w}_{EW}' (\mathbf{I}c_4 + \mathbf{C}) \boldsymbol{\mu}_{PJ} - \mathbf{w}_{EW}' \boldsymbol{\mu} - \frac{1}{\gamma} \boldsymbol{\mu}' (\boldsymbol{\Sigma}^{-1}c_4 + \mathbf{C}) \boldsymbol{\mu}_{PJ} + \theta^2 \right),\end{aligned}$$

and its practical counterpart analogous to (3.14) by,

$$\tilde{\eta}_{13}^{PJ} = \frac{1}{\gamma} \left(\mathbf{w}_{EW}' (\mathbf{I}c_4 + \mathbf{C}) \hat{\boldsymbol{\mu}}_{PJ} - \mathbf{w}_{EW}' \tilde{\boldsymbol{\mu}} - \frac{1}{\gamma} \tilde{\boldsymbol{\mu}}' (\tilde{\boldsymbol{\Sigma}}^{-1}c_4 + \mathbf{C}) \hat{\boldsymbol{\mu}}_{PJ} + \tilde{\theta}^2 \right),$$

where $E \left[\widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} \boldsymbol{\mu}_{PJ} \right]$ is treated as the product of expectations since the variables are independent.

Tu & Zhou (2011, Appendix B) show that

$$\widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} = \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1} c_4^2 - 2\widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \mathbf{C} c_4 + \mathbf{C} \boldsymbol{\Sigma} \mathbf{C},$$

and since $\boldsymbol{\mu}_{PJ}$ is multivariate normal (Jorion (1986)), $E \left[\boldsymbol{\mu}_{PJ} \widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} \boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}_{PJ}^{-1} \boldsymbol{\mu}_{PJ} \right]$ can be evaluated as (3.12) in Section 3.1.1, with $\boldsymbol{\mu}$ substituted by $\boldsymbol{\mu}_{PJ}$ and \mathbf{P} for \mathbf{C} . It then follows that

$$\begin{aligned}\eta_3^{PJ} &= \frac{c_3 c_4^2 N + T \boldsymbol{\mu}_{PJ}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{PJ}}{\gamma^2} - c_4 \frac{N + T \boldsymbol{\mu}_{PJ}' \mathbf{C} \boldsymbol{\mu}_{PJ}}{T} \\ &\quad + c_4 \frac{N + T \boldsymbol{\mu}_{PJ}' \mathbf{C} \boldsymbol{\Sigma} \mathbf{C} \boldsymbol{\mu}_{PJ}}{T} - \frac{2}{\gamma^2} \boldsymbol{\mu}' (\boldsymbol{\Sigma}^{-1}c_4 + \mathbf{C}) \hat{\boldsymbol{\mu}}_{PJ} + \frac{\theta^2}{\gamma^2},\end{aligned}$$

and its practical estimator by,

$$\begin{aligned}\tilde{\eta}_3^{PJ} &= \frac{c_3 c_4^2 N + T \hat{\boldsymbol{\mu}}_{PJ}' \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\mu}}_{PJ}}{\gamma^2} - c_4 \frac{N + T \hat{\boldsymbol{\mu}}_{PJ}' \mathbf{C} \hat{\boldsymbol{\mu}}_{PJ}}{T} \\ &\quad + c_4 \frac{N + T \hat{\boldsymbol{\mu}}_{PJ}' \mathbf{C} \boldsymbol{\Sigma} \mathbf{C} \hat{\boldsymbol{\mu}}_{PJ}}{T} - \frac{2}{\gamma^2} \boldsymbol{\mu}' (\boldsymbol{\Sigma}^{-1}c_4 + \mathbf{C}) \hat{\boldsymbol{\mu}}_{PJ} + \frac{\theta^2}{\gamma^2},\end{aligned}$$

The estimated constant for the combined Jorion rule is

$$\tilde{\delta}_*^{PJ} = \frac{\tilde{\eta}_1 - \tilde{\eta}_{13}^{PJ}}{\tilde{\eta}_1 - 2\tilde{\eta}_{13}^{PJ} + \tilde{\eta}_3^{PJ}},$$

and the rule can now practically be compared to the other rules.

Chapter 4

PERFORMANCE EVALUATION

In order to establish whether a rule is practically sufficient for portfolio management it must be evaluated under circumstances mimicking such environment. Therefore this empirical section will be branched towards both in-sample results, as well as out-of-sample results using real data from major market indices constituting reasonably sized portfolios. The in-sample study carries very little practical inferences, but are important for technical reasons. The out-of-sample study is more targeted toward the practitioner, who is concerned with accumulating wealth most efficiently over time.

4.1 The Experimental Design

The unconstrained portfolios in the theoretical sections of type (1.6), are practically tedious to work with because they allow for very extreme weights in most cases, which in turn translates to unrealistic returns from period to period on a rolling basis. Therefore all portfolio rules are normalized such that

$$\mathbf{w}_{\text{Norm}} = \frac{\mathbf{w}_i}{|\mathbf{1}'_N \mathbf{w}_i|},$$

where i denotes the specific portfolio rule under consideration. The denominator ensures that each portfolio is scaled to be fully invested into risky assets without any asset being leveraged. The experiment is conducted on a “rolling-sample” basis, which means that for a total sample of T monthly returns, a subset M is chosen to be the *in-sample* period. During the in-sample period, parameters are estimated for the *out-of-sample* period $t = M + 1$. The experiment continues by rolling one month forward and repeat until month T is reached. For each portfolio rule and time period, 100 simulations will be iterated by constructing portfolios of size $N = [10, 25, 50]$ where the components are randomly and uniformly selected from each sample set. The average result of the 100 simulations is reported. The experiment is conducted in Matlab and functions are available upon request, as well as being included in Appendix C.

4.1.1 Performance Measures

This study is structured around several relevant performance measures. Without particular order, the first measures are the *out-of-sample* return, r_i and standard deviation, $\sigma_{r,i}$ of each portfolio rule i . Directly as a result of the measures follows the *out-of-sample* Sharpe-ratio of rule i , denoted for each period by

$$\text{SR}_{t,i} = \frac{r_{t,i}}{\sigma_{r,i}}, \quad M < t \leq T,$$

where $r_{t,i} = \hat{\mathbf{w}}'_{i,M} \mathbf{r}_t$, $\sigma_{r,i}^2 = \hat{\mathbf{w}}'_{i,M} \mathbf{S}_t \hat{\mathbf{w}}_{i,M}$ and $\mathbf{S}_{t,i} = (1/T)(\mathbf{r}_t - \bar{r}_t \mathbf{1}_N)'(\mathbf{r}_t - \bar{r}_t \mathbf{1}_N)$, the sample covariance matrix of realized returns. \mathbf{r}_t is the $N \times 1$ return vector in period t and \bar{r}_t is the arithmetic mean of returns. The reported Sharpe-ratio is the average of all periods ratios for each rule.

Second, the *certainty equivalent return* (CER) is evaluated for each rule. The CER is interpreted as the risk-free return level required for the investor to be indifferent to a portfolio of risky assets. Over $T - M$ periods, the quantity measured and reported is

$$\text{CER}_i = r_{t,i} - \frac{\gamma}{2} \sigma_{r,i}^2.$$

Moreover, the efficiency in estimation of the combination constants are important to determine whether a particular rule behaves stably over time or not. For example Tu & Zhou (2011) reported that the combined MacKinlay-Pástor rule suffered large estimation errors. The estimated coefficients are studied for size, because a very small coefficient means that the sophisticated rule is not contributing to potential performance, that is the performance is attributed in large to the equally weighted portfolio alone. The coefficients are also studied for variability. High variability translates to instability in estimation and makes the rule less reliable. Tu & Zhou (2011) find that the combined Jorion rule fails to improve even as the sample size T increases, and suspect that it is because the rule itself is an estimation and not analytically tractable. The same applies to the proposed rule in Section 3.1. Therefore the variance point estimate for each rule is

$$\tilde{\sigma}_{\delta,i}^2 = \frac{1}{T-1} \sum_{t=M}^T (\delta_{i,t} - \bar{\delta}_i)^2,$$

so as to make observations about variability. Moreover, for practical reasons, each rule is truncated to ensure that $\tilde{\delta}_i \in [0, 1]$. The truncation is

$$\tilde{\delta}_i = \max \{0, \min\{\tilde{\delta}_i, 1\}\},$$

because seldom, but occasionally, the estimated coefficients land outside the desired interval which causes problems. This problem decreases as the sample size increases and each estimate is more accurate, as will be seen in Section 4.2.1.

4.1.2 Data Sample

The real market data used in this thesis consists of the Standard & Poors 500 constituent list from the Thomson Reuters Datastream database. The raw data comprise the prices of the last trading day in each month from 1999-12-31 to 2015-01-31. This gives a sample of $T = 181$ monthly returns. The refined data that made it into analysis required to have the full time-series length, that is stocks that were introduced later or discontinued trading during the period were discarded. Moreover stocks that traded thinly, even on monthly basis were also discarded. This results in a refined dataset of 414 stocks eligible for portfolio formation.

The portfolios are generated in sets of 10, 25, 50 and 75 assets and are evaluated on a “rolling” basis using four different lengths of estimation windows. Since each portfolio only constitutes a fraction of the dataset, 50 portfolios are randomly generated for each T and portfolio size, and then averaged.

4.2 Experimental Results

This section is entirely devoted to the presentation, discussion and reporting of results from the evaluation of the measures and quantities given in the previous section. The results are considered for the portfolio rules presented in Table 4.1.

List of portfolio rules

Abbreviation	Description
Panel A: Portfolio Rules	
EW	The equally-weighted rule. Divides all assets by $1/N$
LW (ccm)	Portfolio using the Ledoit & Wolf (2004b) covariance estimator
LW (sid)	Portfolio using the Ledoit & Wolf (2004a) covariance estimator
ML	MV Portfolio with maximum likelihood estimators
KZ	Kan & Zhou (2007) linear combination of the MV and GMV portfolios
PJ	Portfolio with mean and covariance estimators from Jorion (1986)
Panel B: Combined Portfolio Rules	
CLW (ccm)	Proposed rule. Linear combination of EW and LW (ccm)
CLW (sid)	Proposed rule. Linear combination of EW and LW (sid)
CML	Linear combination of EW and ML
CKZ	Linear combination of EW and KZ
CPJ	Linear combination of EW and PJ

Table 4.1: List of Portfolio Rules

This table reports all portfolios considered in the empirical study. Panel A gives the individual portfolio rules (with exception for **KZ** which is a linear combination), while Panel B provides the class of combined rules studied in this thesis. The abbreviations (ccm) and (sid) stand for the constant correlation model and the scaled identity respectively.

As seen, there are six rules considered from the literature. All rules are carefully chosen so as to be either benchmark portfolios or strategies with proven track record in the literature, with relevant design for the experiment. The study will report all the individual rules, and the addition of three combined rules already considered in previous research. The contribution by the author comprises two more rules in the class of linear combinations of portfolio rules.

4.2.1 Combination Coefficients

In Chapter 3 there are four combined portfolio rules presented. All rules have special characteristics making them very different from one another in terms of estimating the optimal combination between the equally weighted rule and each respective sophisticated rule. It is important to study the behaviour of the coefficient under different criteria to see if a particular rule is practically feasible. In addition to the study by Tu & Zhou (2011, Table 5), this experiment also considers the ratio of extreme coefficients in several cases. This is indeed important because it reveals whether the estimated coefficients are well behaved in small samples, which is more likely to an investor than having thousands of sample observations. Since the data sample is constrained by $T = 181$ months, the experiment is constructed such that for portfolios of size $N = [10, 25, 50]$ ¹ randomly selected from the dataset, combination coefficients are computed for $T = [60, 65, 70, \dots, 170, 175, 179]$ in increments of 5 (except for the last). At each T , 50 portfolios are simulated, which results in a $50 \times (181 - T)$ size matrix of combination coefficients. The elements equal to 1 or 0 are truncated (see Section (4.1.1)) and the ratio between the sum of truncated values and the total elements of the matrix constitutes the ratio of extreme coefficients. From Figure 4.1 it can be seen that the proposed rule is most well-behaved when $T = 120$. When T is less, the estimation error increases because of α^* (see (2.24)) which is more frequently also truncated to 1, and as a result the covariance estimator in (3.10) becomes singular. At $T = 120$ the covariance estimator is so well-behaved that less than 0.3% of the α^* estimates gets truncated, which means that most of the extreme coefficients (the $\sim 5\%$) occur simply because of variability in the estimation of δ^* .

¹This section only considers $N = 25$, confer Appendix B for the other cases.

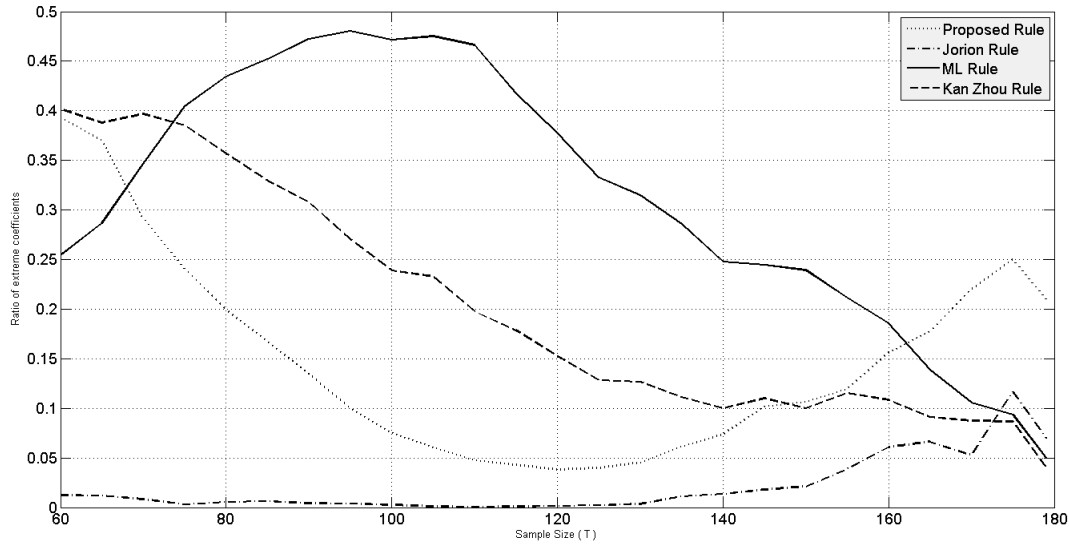


Figure 4.1: Ratio of Extreme Combination Coefficients

This figure displays the ratio of extreme coefficients for portfolios of size $N = 25$ and T varying between 60 and 179. At each instance, 50 portfolios are simulated and the ratio represents the percentage of truncated coefficients.

As for the ML rule which has the highest ratio of extreme coefficients, the explanation can be found in Tu & Zhou (2011, Table 5) where they show that for $T = 120$, the estimated coefficient is relatively small at ~ 0.20 with a standard error of 0.10. This means that for all $T \in [60, 120]$, many realizations are truncated from below 0. This conclusion is strengthened in the following experiment, carried out and presented in Table 4.2, where datasets have been simulated from a multivariate normal distribution for portfolios of $N = 25$ synthetic assets with mean and variances generated on a uniformly distributed interval. For each T , 1000 portfolios are simulated and mean-summary statistics is presented.

The estimated coefficients of the proposed rule are given in Panel 1. The star symbolises that the specific target matrix used is the one in Ledoit & Wolf (2004a) (that is, shrinking toward identity scaled by a “grand variance”). The reason is that the rule behaved strangely on simulated data when the target matrix was the constant correlation covariance matrix. The source of the problem lied in the way $\hat{\gamma}$ (see (2.23)) behaved. The square of the Frobenius norm became extremely small since the synthetic data were independently drawn, and thus had no correlation in the first place. Since this is a parameter problem rather than a model problem, a more well-behaved target was chosen to remedy the occurrence. The coefficient in Panel 1 is estimated with low variability and high stability for all $T > 120$.

The results in Table 4.2 deviates somewhat from the study in Tu & Zhou (2011). In this study, the Jorion rule (Panel 2) seems to converge to ~ 0.19 instead of increasing towards 1 for large samples. Moreover, this study shows much smaller standard errors when the sample moves into the hundreds. The ML rule (Panel 3) and the Kan & Zhou rule (Panel 4) coefficients behave approximately the same as in the Tu & Zhou (2011) study. It can be seen especially, as was discussed in Figure 4.1, that the estimated coefficients of the ML rule are small with

Ratio of Sophisticated Weights

Parameters	T						
	60	120	240	480	960	3000	6000
Panel 1: \mathbf{w}_{LW}^*							
$\tilde{\delta}_C$	0.227 (0.354)	0.287 (0.182)	0.191 (0.065)	0.215 (0.023)	0.239 (0.013)	0.262 (0.006)	0.268 (0.005)
Panel 2: \mathbf{w}_{PJ}							
$\tilde{\delta}_{PJ}$	0.142 (0.058)	0.399 (0.278)	0.121 (0.206)	0.139 (0.039)	0.173 (0.018)	0.190 (0.013)	0.193 (0.012)
Panel 3: \mathbf{w}_{ML}							
$\tilde{\delta}_{ML}$	0.142 (0.155)	0.275 (0.061)	0.481 (0.043)	0.668 (0.025)	0.807 (0.014)	0.930 (0.005)	0.964 (0.003)
Panel 4: \mathbf{w}_{KZ}							
$\tilde{\delta}_{KZ}$	0.932 (0.158)	0.923 (0.080)	0.915 (0.043)	0.933 (0.025)	0.958 (0.014)	0.984 (0.005)	0.992 (0.003)

Table 4.2: Ratio of sophisticated weights

This table reports the ratios of average estimated combination coefficients and their standard error in parenthesis, for each rule respectively. For each sample size, 1000 portfolios of 25 synthetic assets are sampled from an m -variate normal distribution. Each assets mean and variance parameter is sampled from a uniform interval.

variability enough to generate some realizations outside the interval for $\tilde{\delta}_{ML}$.

4.2.2 Out-of-Sample Results

The empirical results important to investors are presented in this section. It should be stressed that parameter uncertainty clouds the investment horizon and greatly impacts on the portfolio selection by sophisticated rules. Tu & Zhou (2011) show that linear combinations of portfolios are theoretically convex and argue that they improve empirically over selecting an individual rule. The results in this section agree, at least partially. The contribution of the study in this thesis, in contrast to the previous study, is to test the combined rules for different dimensions and sample sizes. The first exhibit was presented in Figure 4.1 (and the figures in Appendix B) where it was clearly shown that the behaviour of the combined rules are highly affected by the mean and covariance estimators' dependence of the sampling dimensions. In Table 4.3, the annualized out-of-sample Sharpe ratios are reported for each portfolio rule by averaging the performance over 50 realizations, for each portfolio size and sample length.

In the first row of each panel, the **EW** portfolio is reported. This portfolio is by construction not depending on estimators, and thus is well-behaved throughout all panels. As the portfolio dimension N increases, the risk goes down asymptotically towards market risk. In Panel 1 and Panel 2, there are no cases where sophisticated rules perform as good as the **EW** rule. In fact the only rule with decent performance is the **KZ** rule, and only as long as the ra-

Out-of-Sample Sharpe Ratios

Rules	Portfolio Size (N)							
	10	25	50	75	10	25	50	75
	Panel 1: T = 60				Panel 2: T = 90			
EW	0.75	0.77	0.77	–	0.60	0.65	0.67	0.65
LW (ccm)	0.04	0.23	–0.08	–	0.18	0.16	0.02	0.03
LW (sid)	0.03	0.23	–0.08	–	0.17	0.17	0.02	0.03
ML	0.19	–0.09	–0.08	–	0.16	0.22	–0.08	–0.36
KZ	0.62	0.51	0.11	–	0.51	0.51	0.42	0.08
PJ	–0.17	0.12	0.14	–	0.12	0.22	0.24	0.03
CLW (ccm)	0.32	0.31	0.17	–	0.33	0.53	0.67	0.35
CLW (sid)	0.43	0.44	0.45	–	0.45	0.65	0.65	0.48
CML	0.30	0.30	0.38	–	0.52	0.60	0.46	–0.34
CKZ	0.69	0.63	0.16	–	0.56	0.62	0.56	0.18
CPJ	–0.06	0.42	0.56	–	0.40	0.62	0.57	0.58
	Panel 3: T = 120				Panel 4: T = 150			
EW	1.19	1.18	1.18	1.21	2.01	2.09	2.24	2.25
LW (ccm)	0.86	0.97	0.85	0.78	1.04	0.69	0.61	0.41
LW (sid)	0.88	0.98	0.85	0.78	1.04	0.69	0.61	0.43
ML	0.87	0.92	0.72	0.50	1.03	0.67	0.65	0.48
KZ	1.32	1.43	1.29	1.17	1.66	1.53	1.53	1.47
PJ	0.77	0.90	0.72	0.51	0.91	0.64	0.64	0.48
CLW (ccm)	1.13	1.27	1.28	1.29	1.78	1.70	1.88	2.05
CLW (sid)	1.17	1.29	1.27	1.36	1.83	1.83	1.81	1.84
CML	1.19	1.24	1.17	1.10	1.89	1.93	2.18	2.14
CKZ	1.28	1.42	1.30	1.27	1.84	1.78	1.96	2.03
CPJ	1.12	1.29	1.29	1.27	1.68	1.55	2.06	2.15

Table 4.3: Out-of-Sample Sharpe Ratios

This table reports the out-of-sample Sharpe ratios for all the rules considered. Each number represents average-summary statistics of 50 simulated portfolios, with assets randomly selected from the dataset.

tio N/T is significantly smaller than 1. Even though the linear combinations greatly improve across Panel 1 and Panel 2, over the individual rules, they do not outperform the **EW** rule. The performance of the combined portfolios are merely helped by the **EW** rule. The evidence is explained partially by Table 4.2, where the sophisticated portions in Panel 1, 2, and 3 are low and volatile for $T = 60$ and $T = 120$.

In Table 4.3, Panel 3, the individual rules are less affected by estimation error and hence, more well-behaved. All sophisticated rules except the **KZ** rule (which is a diversified two-fund rule) are still performing below the **EW** rule in terms of risk adjusted returns. However, the linear combinations of the same rules show enough convexity to outperform the **EW** rule in many cases. While the **KZ** rule already performs well as an individual rule, the proposed rules improves greatly in combination with the **EW** rule.

The individual rules fail to improve from Panel 3 to Panel 4, while the **EW** rule does not. This still improves greatly on the combined rules, but not enough to outperform the **EW** rule.

On a general basis, Table 4.2 shows that the sophisticated rules most dependent on the **ML** estimators perform poorly on realized returns, with exception of the **KZ** rule which performs better because it already is a two-fund rule. This clearly affects the out-of-sample performance of the combined rules. It is therefore important to show explicitly in which cases the estimated rules are well-behaved, a study missing in Tu & Zhou (2011). On that result it is also clear from all Panels that the individual rules are much affected by the N/T ratio as the Sharpe ratios decrease while the ratio increases. The reader interested in the realized returns and standard deviations behind the Sharpe ratios is referred to Appendix B, Tables B.2 and B.3.

Moving on to the certainty equivalent returns, they strengthen the results given by the Sharpe ratios. A look at Table 4.4 shows that the **EW** rule is well-behaved across all panels. It is also evident that the sophisticated rules (with the exception of **KZ**) are highly unstable for the shorter estimation windows in Panel 1 and Panel 2. The performance of the combined rules are strengthened only by the size of the ratio invested in the **EW** rule.

The results in Panel 3 are interesting because the CER's are the highest for **CLW**, for all portfolio sizes. The likely explanation is that $T = 120$ is the estimation window that gives the best balance between the target matrix (bias) and the sample covariance matrix (estimation error) of the four windows. Ledoit & Wolf (2004b) showed that as T increases, the shrinkage coefficient of the covariance estimator in (2.15) converges to the **ML** estimator, which is sub-optimal for studies on realized returns as it is perfectly clear that the **ML** estimator is a poor choice out-of-sample. This coincides with Proposition 3, and explains why the performance of the **CLW** rules are poor for small T (because of covariance estimator bias from the target matrix), and poor for larger T 's (because of covariance estimator estimation error from the sample covariance matrix). The conclusion is that longer estimation windows than $T = 150$ would surely just result in the **CLW** rules converging to the **CML** rule, which is undesired. For the case of CER when $\gamma = 1$, confer Appendix B, Table B.1.

Certainty Equivalent Returns, $\gamma = 3$

Rules	Portfolio Size							
	10	25	50	75	10	25	50	75
	Panel 1: T = 60				Panel 2: T = 90			
EW	8.55	9.26	8.97	-	6.67	6.68	6.95	7.19
LW (ccm)	< -5000	-258.06	-662.55	-	< -5000	-186.71	-467.82	-24.27
LW (sid)	-419.60	-3385.47	-493.55	-	-58.22	-7.60	-8.14	-8.04
ML	-509.85	-166.62	-220.86	-	-272.55	-17.54	-80.90	< -5000
KZ	5.28	-61.10	6.49	-	5.56	4.12	0.61	-14.11
PJ	-2033.73	-50.79	< -5000	-	-1086.61	-17.57	-98.12	< -5000
CLW (ccm)	-758.89	-106.84	-68.71	-	< -5000	-12.90	6.75	0.67
CLW (sid)	-21.33	6.36	-28.56	-	-2.88	4.58	5.52	4.07
CML	-78.87	-17.06	-18.41	-	-41.10	2.48	-22.68	-677.74
CKZ	6.71	-27.20	8.08	-	6.86	5.28	3.25	-8.14
CPJ	-252.93	7.24	-1009.60	-	-393.30	5.46	5.05	-644.81
	Panel 3: T = 120				Panel 4: T = 150			
EW	13.95	14.29	14.55	14.21	18.59	19.38	19.17	19.17
LW (ccm)	11.60	11.40	10.03	9.29	12.05	11.71	6.14	3.22
LW (sid)	12.51	12.94	12.24	12.50	13.30	13.72	9.90	7.96
ML	11.80	10.50	0.13	-7.12	11.75	11.71	6.23	2.60
KZ	13.11	13.31	13.18	12.55	16.51	15.93	14.19	14.53
PJ	10.65	10.16	4.24	-4.07	10.38	11.40	6.20	2.76
CLW (ccm)	13.65	14.61	15.05	14.70	17.32	17.45	17.45	17.25
CLW (sid)	14.14	14.71	13.95	14.19	17.78	17.78	16.01	13.98
CML	13.55	14.03	14.03	14.01	17.18	18.56	18.29	18.43
CKZ	13.31	14.01	13.96	13.94	17.37	17.17	16.48	16.69
CPJ	13.92	14.42	14.59	14.56	16.80	17.28	17.22	17.83

Table 4.4: Certainty Equivalent Returns, $\gamma = 3$

This table reports the certainty equivalent returns from respective portfolio rule under different criteria. The highest CER in each experiment is highlighted in bold. The experiment is conducted out-of-sample in four panels of different estimation length T , on a rolling basis. In each panel, four portfolio sizes are evaluated and the reported numbers are the averages over 50 simulations. “ccm” stands for the constant correlation model and “sid” is short for scaled identity, and they are distinguished by which target matrix the sample matrix is shrunk towards. The risk aversion coefficient $\gamma = 3$.

4.2.3 Convexity Plots for the Combined Rules

This section provides knowledge about the convexity of the combined portfolio rules. In this experiment, δ_i is not estimated but incremented in steps of 0.05 on the closed interval $[0, 1]$. The plots reveal evidence of where the true coefficients lie, and hence shows how good or poor each estimation method is. The experiment is constructed such that for each plot, a period of T estimates is chosen for an N -sized portfolio. 1000 out-of-sample portfolios are then simulated for each δ increment and the average return and standard deviation is plotted in (σ, μ) space. Selected plots are discussed in this section, and more plots can be found in Appendix B.3.

Figure 4.2 shows the behaviour of the convexity for the proposed combined rule, for various portfolio sizes. $T = 120$ is chosen because it was considered to have the right balance between estimation error and bias for the covariance estimator (see Appendix B.3 for the $T = 150$ case).

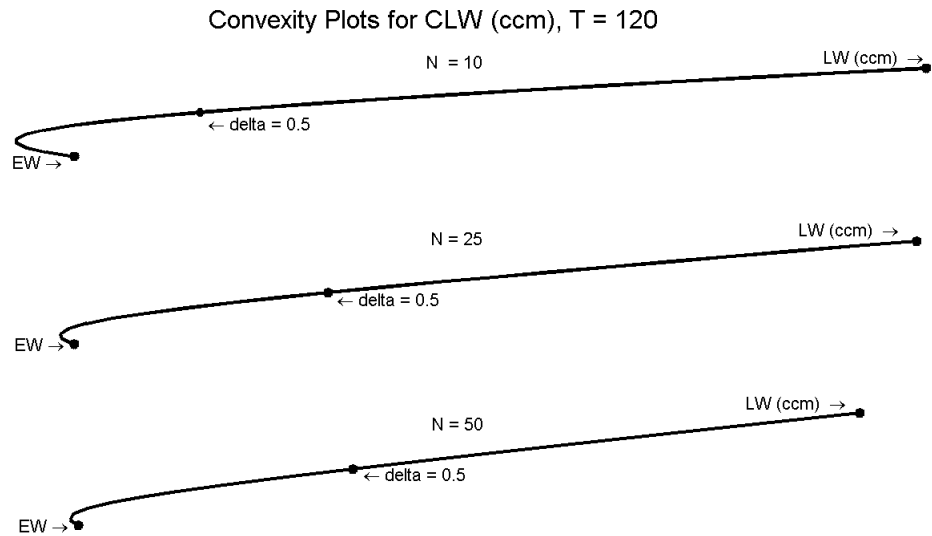


Figure 4.2: Convexity Plot for the Proposed Rule, CLW (ccm), $T = 120$

This curve is plotted in (σ, μ) -space for the proposed rule using the Ledoit & Wolf covariance estimator with the constant correlation model being the target matrix.

It is clear that there exist combined portfolios with out-of-sample performance higher than respective individual rule. The factor that limits the use of this class of rules certainly lies in the estimation of the combination coefficients. The efficient set also becomes more linear as the portfolio size (N) increases. The same trend is seen for the **CLW (sid)** rule in Figure 4.3. For comparison, the $T = 150$ case is found in Appendix B.3.

The **CKZ** rule shows most convexity of all rules, which should not be a surprise given the previous results where the rule was well-behaved across all samples. It possesses greater opportunity as it is a three-fund rule, already diversified by the two sophisticated components. Figure 4.4 shows the curvature of the **CKZ** rule. The same behaviour of diminishing convexity is displayed here for increased portfolio sizes. Figure 4.5 shows the **CKZ** plot for $T = 150$.

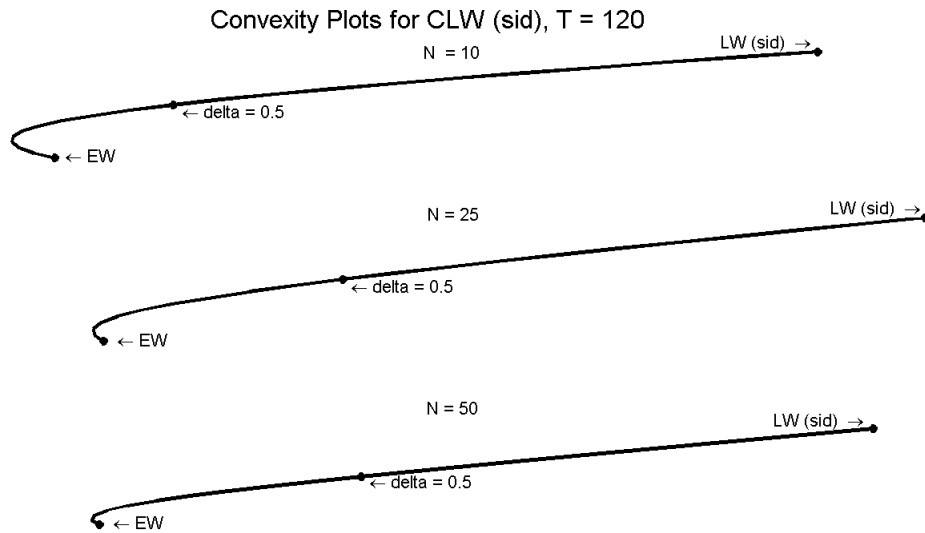


Figure 4.3: Convexity Plot for the Proposed Rule, CLW (sid), T = 120

This curve is plotted in (σ, μ) -space for the proposed rule using the Ledoit & Wolf covariance estimator with the scaled identity matrix being the target matrix.

The previous experiment in Table 4.2 indicated that the **CKZ** rule is highly weighted towards **KZ** already in small estimation windows. It follows that for both Figure 4.4 and Figure 4.5, the estimated coefficient lies to the right of $\delta = 0.9$.

The plots for the **CML** rule is left in Appendix B.3 because it is a lesser version of the **CLW** rules, that is less convex in all samples. The plots for the **CPJ** rule is more or less linear and therefore left out completely.

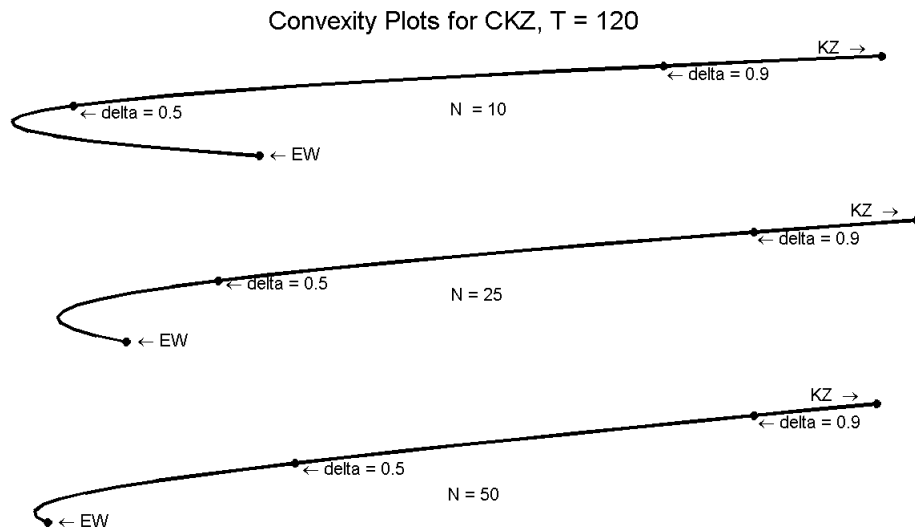


Figure 4.4: Convexity Plot for the Combined Kan & Zhou Rule, CKZ, T = 120
 This curve is plotted in (σ, μ) -space for the linear combination of the equally weighted rule and the two-fund Kan & Zhou rule composed of a normalized combination of the MV rule and GMV rule. The result is a three-fund rule with $\delta = 0.5$ being the equally weighted combination between the rules.

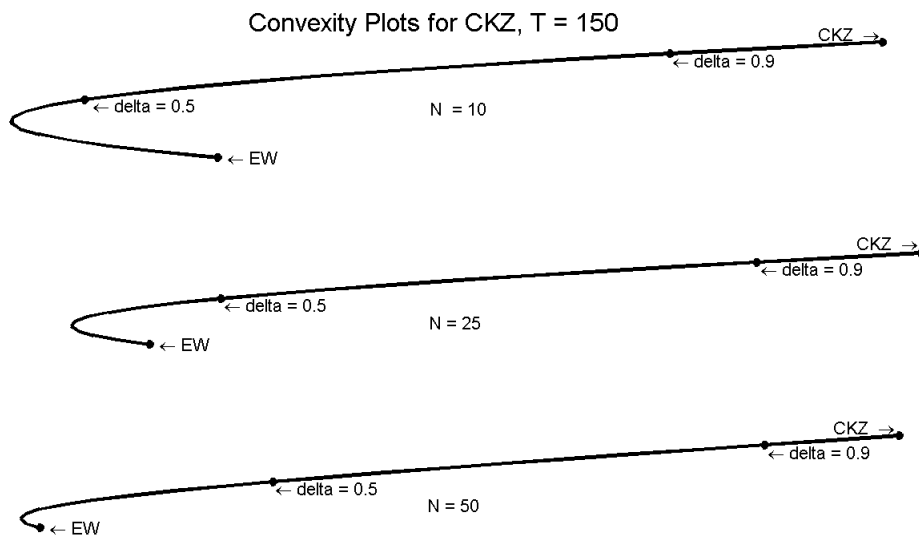


Figure 4.5: Convexity Plot for the combined Kan & Zhou, CKZ, T = 150
 This curve is plotted in (σ, μ) -space for the linear combination of the equally weighted rule and the two-fund Kan & Zhou rule composed of a normalized combination of the MV rule and GMV rule. The result is a three-fund rule with $\delta = 0.5$ being the equally weighted combination between the rules.

CONCLUSION

The existence of parameter uncertainty questions the usefulness of theoretical allocation models for practical investing. A well-documented literature suggests that simply dividing the assets in equally sized proportions is at least as good. Recent research shows that the linear combination of an equally weighted allocation and a sophisticated MV model in some cases increase out-of-sample portfolio performance significantly.

This thesis has enriched the current research by the proposal of a new combined rule based on a shrinkage estimator for the covariance matrix. The new rule performs at least as good as its peer group on realized returns. Moreover it has been shown that not only the chosen rule is important for performance, but equally important is the size of the portfolio dimension and estimation window, for generating well-behaved weight vectors. The developed **CLW** rules performs best when the sample size balances estimation error from the sample covariance matrix with bias from the target matrix. When the estimation window is too short, the shrinkage coefficient is volatile and the rule suffers because many realizations are estimated with a near-singular sample covariance matrix. When the estimation window is longer than preferred, the rule performs too close to the **CML** rule which is inferior out-of-sample due to estimation error.

All the sophisticated rules give highly volatile estimations when the estimation window (T) is short, and perform poorly when the N/T ratio is close to 1 because the sample covariance matrix is ill-conditioned. The exception is the Kan & Zhou two-fund rule which diversifies and performs marginally better. When the estimation window increases, all combined rules become more well-behaved due to the sophisticated components being more accurate. The **CKZ**, and both **CLW** rules do perform better than the **EW** portfolio when $T = 120$ in terms of Sharpe ratios. True for all combined rules is that they either improve over one of its components, or both. In most cases the **EW** rule helps the sophisticated rules by trading estimation error for bias.

The shrinkage estimator used for the covariance matrix in this thesis belongs to a class of estimators which have got a lot of attention in the last ten years. Most likely there exists more estimators (see for example Kourtis et al (2012) and Kubokawa & Inoue (2014)) in this class that can be used in a combined rule. Ledoit & Wolf (2014) have also introduced a new class of estimators of non-linear shrinkage of the covariance matrix, that are interesting for future research. Moreover, no effort has been made to decrease the estimation error of the mean estimator. It is thereby possible that another mean estimator together with the covariance matrix of the **CLW** rules, yields a better portfolio rule.

From the graphical figures it is clear that the convexity exists not just theoretically, but also

in the empirical studies of out-of-sample performance. The restriction that limits the usefulness is currently that the estimated combination coefficients are not optimal. Further research lies in finding more accurate estimates for the combination coefficients. Perhaps numerical methods gives more accurate results. Moreover, since diversification between portfolio rules has the same convex properties as diversification among assets, the key perhaps lies in the diversification among more portfolio rules, that is to create N-fund rules. The Kan & Zhou rule clearly hints toward that direction.

The end statement deserves a caveat before making recommendations. Modern portfolio theory is a relatively young science with pitfalls for the practical investor, due to the inherent estimation errors of portfolio optimization. All investment decisions are done under uncertainty and the golden rule will most likely never be found. Parametric solutions are only as good as the inputs and while they become better, and in certain cases outperform the thousand year old diversification strategy advised in the Talmud, it is certainly more reassuring to an investor to show asymptotic converge of risk from an equally-weighted portfolio, than to live with the realized tail risks an optimization model generates from estimation error.

Upon that assessment, the retail investor is most likely better off with focus on simple diversification rules and stock picking instead of employing sophisticated rules for three reasons: it is tedious to collect data for estimation, it takes quantitative knowledge to set up a portfolio optimizer, and the retail investor is not bound by rules and constraints, as institutional investors are.

Bibliography

- Barry, C. B. (1974). Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, **29**(2), 515–522.
- Bickel, P. J., & Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 2577–2604.
- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Chang, C., & Tsay, R. S. (2010). Estimation of covariance matrix via the sparse Cholesky factor with lasso. *Journal of Statistical Planning and Inference*, **140**(12), 3858–3873.
- Chopra, V. K., Hensel, C. R., & Turner, A. L. (1993). Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s. *Management Science*, **39**(7), 845–855.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. *Review of Financial Studies*, **22**(5), 1915–1953.
- Frost, P. A., & Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, **21**(3), 293–305.
- Grinold, R. C., & Kahn, R. N. (2000). *Active portfolio management*. McGraw-Hill Professional.
- Haff, L. R. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, **9**(4), 531–544.
- Harvey, C. R., Liechty, J. C., & Liechty, M. W. (2008). Bayes vs. resampling: a rematch, *Journal of Investment Management*, **6**(1), 1–17.
- Huang, J. Z., Liu, N., Pourahmadi, M., & Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, **93**(1), 85–98.
- Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, **58**(4), 1651–1684.
- James, W., & Stein, C. (1961, June). Estimation with quadratic loss. *In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, **1** (1961), 361–379.

- Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, **75**(371), 544–554.
- Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, **21**(03), 279–292.
- Kalymon, B. A. (1971). Estimation risk in the portfolio selection model. *Journal of Financial and Quantitative Analysis*, **6**(01), 559–582.
- Kan, R., & Smith, D. R. (2008). The distribution of the sample minimum-variance frontier. *Management Science*, **54**(7), 1364–1380.
- Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, **42**(03), 621–656.
- Kourtis, A., Dotsis, G., & Markellos, R. N. (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking & Finance*, **36**(9), 2522–2531.
- Kubokawa, T., & Inoue, A. (2014). Estimation of covariance and precision matrices under scale-invariant quadratic loss in high dimension. *Electronic Journal of Statistics*, **8**, 130–158.
- Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, **10**(5), 603–621.
- Ledoit, O., & Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, **88**(2), 365–411.
- Ledoit, O., & Wolf, M. (2004b). Honey, I shrunk the sample covariance matrix, *The Journal of Portfolio Management*, **30**(4), 110–119.
- Ledoit, O., & Wolf, M. (2014). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Available at SSRN 2383361.
- MacKinlay, A. C., & Pástor, L. (2000). Asset pricing models: implications for expected returns and portfolio selection. *Review of Financial Studies*, **13**(4), 883–916.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, **7**(1), 77–91.
- Michaud, R. O. (1989). The Markowitz optimization enigma: is “optimized” optimal?. *Financial Analysts Journal*, **45**(1), 31–42.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley Series In Probability and Mathematical Statistics, John Wiley & Sons.
- Officer, R. R. (1972). The distribution of stock returns. *Journal of the American Statistical Association*, **67**(340), 807–812.

- Pástor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, **55**(1), 179–223.
- Pestman, W. R. (1998). *Mathematical statistics: an introduction*. Walter de Gruyter.
- Platen, E., & Rendek, R. (2012). Approximating the numéraire portfolio by naive diversification. *Journal of Asset Management*, **13**(1), 34–50.
- Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 119–138.
- Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, **1**(399), 197–206).
- Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, **99**(1), 204–215.

Appendix A

MATHEMATICAL PROOFS

This appendix holds the proofs of some mathematical formulas presented in the report. The ordering is the same as in the main document.

A.1 Derivation of Equation (1.12)

The quadratic problem problem is

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} \\ \text{Subject to} \quad & \mathbf{w}' \mathbf{1} = 1 \\ & \mathbf{w}' \boldsymbol{\mu} = q. \end{aligned} \tag{A.1}$$

Start with re-writing it as a Lagrangian function

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \lambda, \gamma) &= \frac{1}{2} \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} + \lambda(1 - \mathbf{1}' \mathbf{w}) + \gamma(q - \mathbf{w}' \boldsymbol{\mu}) \\ \nabla_{\mathbf{w}} \mathcal{L} &= \boldsymbol{\Sigma} \mathbf{w} - \lambda \mathbf{1} - \gamma \boldsymbol{\mu} = 0 \\ \mathbf{w}^* &= \lambda \boldsymbol{\Sigma}^{-1} \mathbf{1} + \gamma \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}. \end{aligned} \tag{A.2}$$

The second constraint $\mathbf{w}' \boldsymbol{\mu} = q$ is used for \mathbf{w}^* to find λ and γ . Use (A.2) and let

$$\begin{aligned} \mathbf{w}' \boldsymbol{\mu} &= q \\ (\lambda \boldsymbol{\Sigma}^{-1} \mathbf{1} + \gamma \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})' \boldsymbol{\mu} &= q \\ \lambda \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{1} + \gamma \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} &= q \end{aligned} \tag{A.3}$$

Define $a \equiv \mathbf{1}' \boldsymbol{\Sigma}^{-1} \mathbf{1}$, $b \equiv \mathbf{1}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, and $\theta^2 \equiv \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, then

$$\lambda = \frac{q - \gamma \theta^2}{b} \tag{A.4}$$

and

$$\begin{aligned}
\mathbf{w}'\mathbf{1} &= 1 \\
\left(\frac{q - \gamma\theta^2}{b}\boldsymbol{\Sigma}^{-1}\mathbf{1} + \gamma\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)' \mathbf{1} &= 1 \\
\frac{q - \gamma\theta^2}{b}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} + \gamma\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} &= 1 \\
\frac{aq}{b} - \frac{\gamma a\theta^2}{b} + \gamma b &= 1 \\
aq - \gamma(a\theta^2 + b^2) &= b \\
\frac{aq - b}{a\theta^2 - b^2} &= \gamma.
\end{aligned} \tag{A.5}$$

When substituting (A.5) back into (A.4), one can solve the final component

$$\begin{aligned}
\lambda &= \frac{q - \gamma\theta^2}{b} \\
&= \frac{q - \left(\frac{aq - b}{a\theta^2 - b^2}\right)\theta^2}{b} \\
&= \frac{1}{b} \left(q - \frac{qa\theta^2 - b\theta^2}{a\theta^2 - b^2} \right) \\
&= \frac{1}{b} \left(\frac{qa\theta^2 - qa\theta^2 + b\theta^2 - qb^2}{a\theta^2 + b^2} \right) \\
&= \frac{\theta^2 - qb}{a\theta^2 - b^2}.
\end{aligned} \tag{A.6}$$

A.2 Proof of Proposition 3

To prove Proposition 3 it is enough to show that the mixed term η_{13} vanishes and that $\eta_3 = \pi_2$ as the substitution is made. Let $\alpha = 0$ and substitute $\widehat{\Sigma}$ for $\tilde{\Sigma}$, then

$$\begin{aligned}
\eta_{13} &= \frac{1}{\gamma} (\mathbf{w}_{EW}' \Sigma \mathbb{E} [\widehat{\Sigma}_{LW}^{-1}] \boldsymbol{\mu} - \mathbf{w}_{EW}' \boldsymbol{\mu} - \mathbb{E} [\boldsymbol{\mu}' \widehat{\Sigma}_{LW}^{-1} \tilde{\boldsymbol{\mu}}] + \theta^2), \\
&= \frac{1}{\gamma} \left(\mathbf{w}_{EW}' \Sigma \mathbb{E} \left[\frac{\widehat{\Sigma}^{-1}}{1-\alpha} - \mathbf{P} \right] \boldsymbol{\mu} - \mathbf{w}_{EW}' \boldsymbol{\mu} - \mathbb{E} \left[\boldsymbol{\mu}' \left(\frac{\widehat{\Sigma}^{-1}}{1-\alpha} - \mathbf{P} \right) \tilde{\boldsymbol{\mu}} \right] + \theta^2 \right), \Big|_{\mathbf{P}=0_{n \times n}, \alpha=0}, \\
&= \frac{1}{\gamma} \left(c_1^{-1} \mathbf{w}_{EW}' \Sigma \mathbb{E} [\tilde{\Sigma}^{-1}] \boldsymbol{\mu} - \mathbf{w}_{EW}' \boldsymbol{\mu} - c_1^{-1} \mathbb{E} [\boldsymbol{\mu}' \tilde{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}] + \theta^2 \right), \\
&= \frac{1}{\gamma} (c_1^{-1} c_1 \mathbf{w}_{EW}' \Sigma \Sigma^{-1} \boldsymbol{\mu} - \mathbf{w}_{EW}' \boldsymbol{\mu} - c_1^{-1} c_1 \theta^2 + \theta^2), \\
&= \frac{1}{\gamma} (\mathbf{w}_{EW}' \boldsymbol{\mu} - \mathbf{w}_{EW}' \boldsymbol{\mu} - \theta^2 + \theta^2), \\
&= 0.
\end{aligned}$$

The second term is

$$\begin{aligned}
\eta_3 &= \mathbb{E} [\tilde{\boldsymbol{\mu}}' \widehat{\Sigma}_{LW}^{-1} \Sigma \widehat{\Sigma}_{LW}^{-1} \tilde{\boldsymbol{\mu}}] - \frac{2}{\gamma^2} \mathbb{E} [\boldsymbol{\mu}' \widehat{\Sigma}_{LW}^{-1} \tilde{\boldsymbol{\mu}}] + \frac{\theta^2}{\gamma^2}, \\
&= \mathbb{E} \left[\tilde{\boldsymbol{\mu}}' \left(\frac{\widehat{\Sigma}^{-1} \Sigma \widehat{\Sigma}^{-1}}{(1-\delta)^2} - \frac{2\widehat{\Sigma}^{-1} \Sigma \mathbf{P}}{1-\delta} + \mathbf{P} \Sigma \mathbf{P} \right) \tilde{\boldsymbol{\mu}} \right] - \frac{2}{\gamma^2} \mathbb{E} \left[\boldsymbol{\mu}' \left(\frac{\widehat{\Sigma}^{-1}}{1-\alpha} - \mathbf{P} \right) \tilde{\boldsymbol{\mu}} \right] + \frac{\theta^2}{\gamma^2}, \Big|_{\mathbf{P}=0_{n \times n}, \alpha=0}, \\
&= \mathbb{E} [\tilde{\boldsymbol{\mu}}' \widehat{\Sigma}^{-1} \Sigma \widehat{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}] - \frac{2}{\gamma^2} \mathbb{E} [\boldsymbol{\mu}' \widehat{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}] + \frac{\theta^2}{\gamma^2}, \\
&= \frac{c_1^{-2}}{\gamma^2} \mathbb{E} [\tilde{\boldsymbol{\mu}}' \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}] - \frac{2c_1^{-1}}{\gamma^2} \mathbb{E} [\boldsymbol{\mu}' \tilde{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}] + \frac{\theta^2}{\gamma^2}, \\
&= \frac{c_1^{-2} c_2 N + T \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu}}{\gamma^2 T} - \frac{2c_1^{-1} c_1}{\gamma^2} \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu} + \frac{\theta^2}{\gamma^2}, \\
&= \left(\frac{T-N-2}{T} \right)^2 \left(\frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \right) \left(\frac{N}{\gamma^2 T} + \frac{\theta^2}{\gamma^2} \right) - \frac{2\theta^2}{\gamma^2} + \frac{\theta^2}{\gamma^2}, \\
&= \frac{(T-N-2)(T-2)}{(T-N-1)(T-N-4)} \left(\frac{N}{\gamma^2 T} + \frac{\theta^2}{\gamma^2} \right) - \frac{\theta^2}{\gamma^2}, \\
&= (c_3 - 1) \frac{\theta^2}{\gamma^2} + \frac{c_3 N}{\gamma^2 T}, \\
&= \pi_2,
\end{aligned}$$

which is exactly the expression by Tu & Zhou (2011), and the proof is complete. \square

Appendix B

MORE TABLES & FIGURES

B.1 Figures

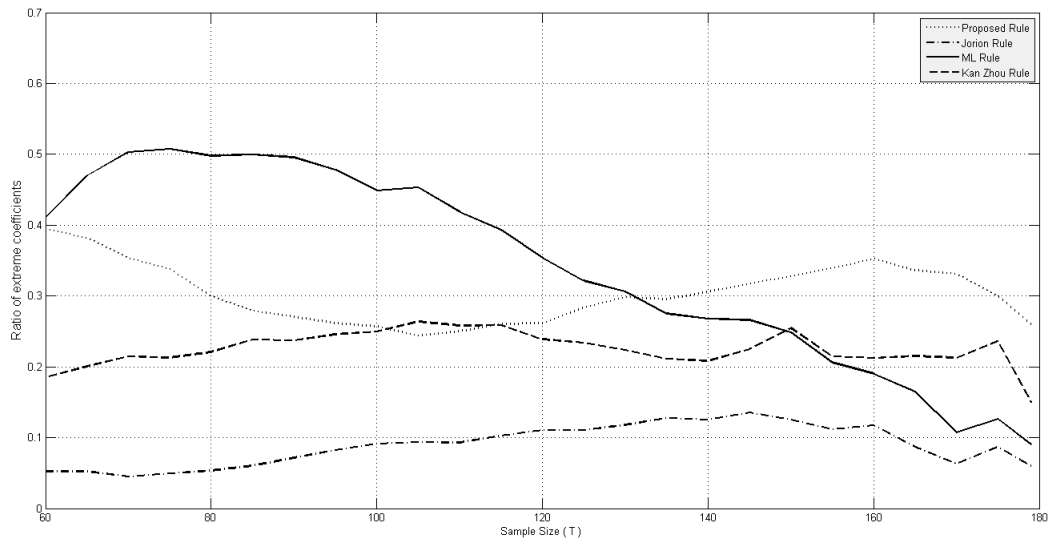


Figure B.1: Ratio of extreme combination coefficients, case of $N = 10$.

This figure displays the ratio of extreme coefficients for portfolios of size $N = 10$ and T varying between 60 and 179. At each instance, 50 portfolios are simulated and the ratio represents the percentage of truncated coefficients.

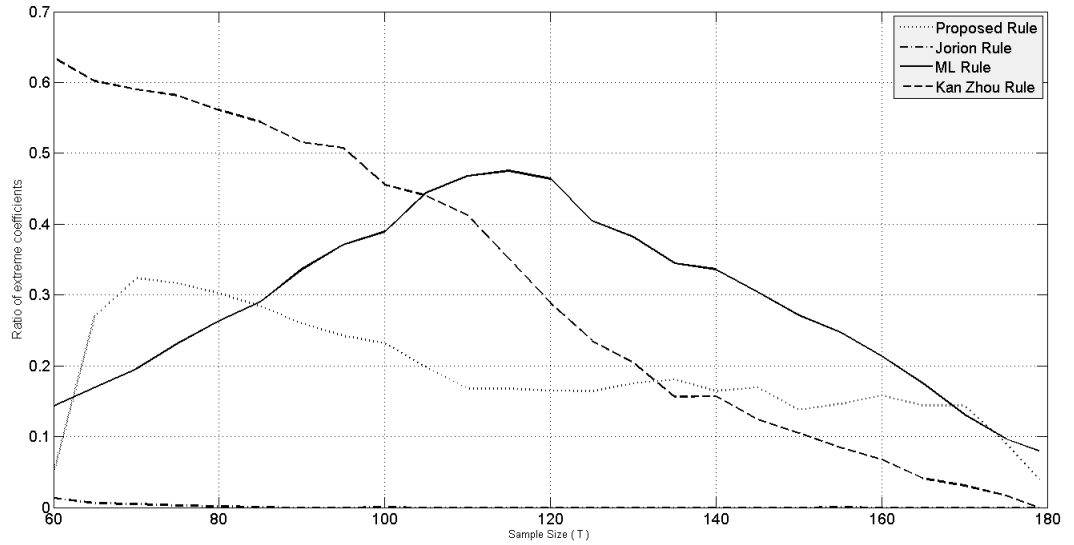


Figure B.2: Ratio of extreme combination coefficients, case of $N = 50$. This figure displays the ratio of extreme coefficients for portfolios of size $N = 50$ and T varying between 60 and 179. At each instance, 50 portfolios are simulated and the ratio represents the percentage of truncated coefficients.

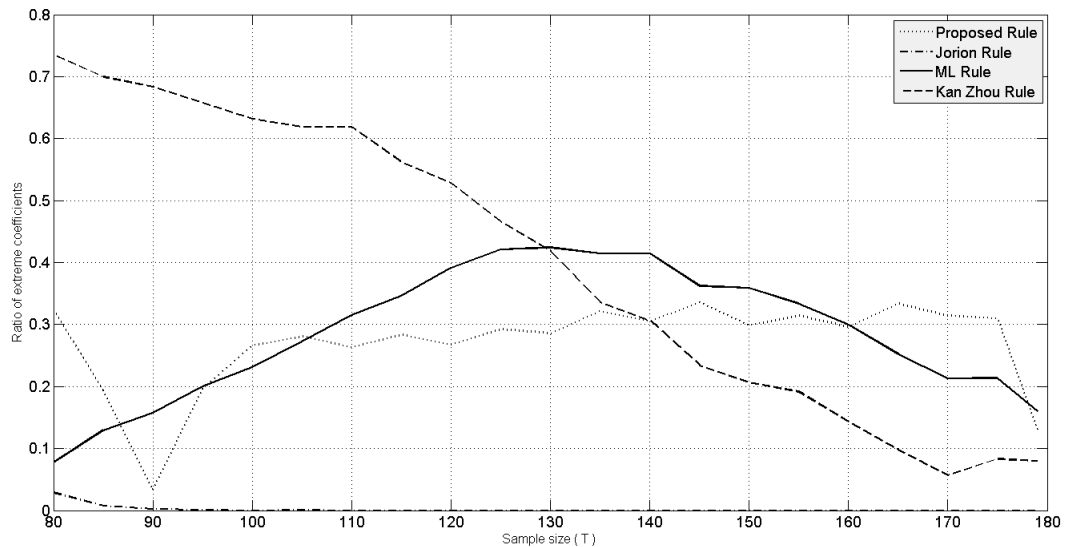


Figure B.3: Ratio of extreme combination coefficients, case of $N = 75$. This figure displays the ratio of extreme coefficients for portfolios of size $N = 75$ and T varying between 60 and 179. At each instance, 50 portfolios are simulated and the ratio represents the percentage of truncated coefficients.

B.2 Tables

Certainty equivalent returns, $\gamma = 1$

Rules	Portfolio Size (N)							
	10	25	50	75	10	25	50	75
	Panel 1: T = 60				Panel 2: T = 90			
EW	11.55	12.04	12.41	-	10.36	10.70	10.78	10.74
LW (ccm)	-98.19	-101.25	-31.83	-	-284.98	-86.58	-7.12	-5.69
LW (sid)	-112.73	-3.60	-4.31	-	-10.33	1.74	0.58	-0.16
ML	< -5000	-259.22	< -5000	-	-781.84	-0.69	-16.99	< -5000
KZ	8.21	7.69	< -5000	-	7.61	5.87	4.11	-1.69
PJ	-1866.20	-234.47	< -5000	-	-47.67	-1.12	-23.05	-3002.86
CLW (ccm)	-55.55	4.57	11.34	-	-96.94	9.43	10.63	10.42
CLW (sid)	-10.35	6.65	8.77	-	8.13	8.70	7.48	8.28
CML	-66.08	12.23	12.16	-	10.45	10.88	10.84	10.86
CKZ	9.81	9.00	-5000	-	9.95	7.99	7.18	-0.79
CPJ	-179.10	-0.42	-56.23	-	0.59	8.94	9.39	9.31
	Panel 3: T = 120				Panel 4: T = 150			
EW	16.71	16.74	16.58	16.58	19.35	20.03	20.39	19.81
LW (ccm)	13.79	14.70	13.84	13.35	14.06	9.09	9.24	6.28
LW (sid)	14.76	15.55	15.60	15.60	15.22	11.19	11.89	10.53
ML	14.34	14.35	12.94	-15.65	14.35	9.43	9.11	8.04
KZ	16.18	15.01	14.32	12.24	17.92	16.30	14.96	15.17
PJ	13.67	14.22	12.94	3.10	13.43	9.07	9.09	8.11
CLW (ccm)	16.68	16.69	16.39	16.22	19.16	19.38	20.29	19.49
CLW (sid)	16.79	16.88	16.95	16.85	19.09	19.03	20.04	19.31
CML	16.48	16.55	16.41	16.57	18.73	19.09	19.87	19.50
CKZ	16.36	16.72	16.17	13.15	18.69	19.37	18.85	17.72
CPJ	15.94	16.34	16.38	16.45	16.92	16.26	17.82	18.26

Table B.1: Certainty Equivalent Returns, $\gamma = 1$

This table reports the certainty equivalent returns from respective portfolio rule under different criteria. The highest CER in each experiment is highlighted in bold. The experiment is conducted out-of-sample in four panels of different estimation length T , on a rolling basis. In each panel, four portfolio sizes are evaluated and the reported numbers are the averages over 50 simulations. “ccm” stands for the constant correlation model and “sid” is short for scaled identity, and they are distinguished by which target matrix the sample matrix is shrunk towards. The risk aversion coefficient $\gamma = 1$.

Out-of-sample returns & standard deviations, Panel 1 & 2

Rules	Portfolio Size							
	10	25	50	75	10	25	50	75
	Panel 1: T = 60				Panel 2: T = 90			
EW	1.1%	1.2%	1.2%	-	1.0%	1.1%	1.1%	1.0%
	5.3%	5.2%	5.0%	-	6.1%	5.7%	5.6%	5.6%
LW (ccm)	0.3%	7.3%	-0.5%	-	0.6%	0.5%	0.1%	0.1%
	24.0%	112.2%	21.8%	-	11.5%	10.4%	11.3%	10.4%
LW (sid)	0.2%	7.4%	-0.5%	-	0.6%	0.5%	0.1%	0.1%
	24.4%	112.4%	22.1%	-	11.7%	10.4%	11.4%	10.6%
ML	2.3%	-2.5%	-5.5%	-	0.4%	0.6%	-0.5%	-1080.8%
	41.4%	90.9%	243.4%	-	8.2%	9.1%	19.3%	10379.5%
KZ	0.8%	0.6%	0.3%	-	0.7%	0.6%	0.6%	0.2%
	4.3%	4.4%	10.9%	-	4.8%	4.2%	4.8%	10.9%
PJ	-3.2%	1.2%	10.4%	-	0.3%	0.6%	1.9%	1.2%
	66.5%	37.1%	252.8%	-	9.8%	9.5%	28.2%	121.8%
CLW (ccm)	1.0%	8.2%	0.5%	-	0.9%	0.9%	1.0%	0.5%
	10.7%	93.3%	10.7%	-	9.3%	5.9%	5.4%	5.3%
CLW (sid)	1.1%	1.6%	0.8%	-	0.8%	1.0%	1.0%	0.7%
	8.5%	12.5%	6.3%	-	6.4%	5.1%	5.3%	4.9%
CML	1.4%	1.3%	1.1%	-	0.9%	1.0%	0.9%	-22.8%
	17.0%	15.8%	10.3%	-	5.9%	5.6%	6.8%	233.9%
CKZ	0.9%	0.8%	0.4%	-	0.8%	0.8%	0.8%	0.5%
	4.4%	4.3%	8.4%	-	5.1%	4.5%	4.8%	8.6%
CPJ	-0.4%	1.2%	1.3%	-	0.7%	0.9%	1.2%	1.0%
	24.3%	9.6%	8.4%	-	6.2%	5.0%	7.0%	6.2%

Table B.2: Table of Empirical Returns & Standard Deviations, Panel 1 & 2

Out-of-sample returns & standard deviations, Panel 3 & 4

Rules	Portfolio Size							
	10	25	50	75	10	25	50	75
	Panel 3: T = 120				Panel 4: T = 150			
EW	1.5%	1.5%	1.5%	1.5%	1.7%	1.7%	1.7%	1.7%
	4.4%	4.3%	4.3%	4.2%	3.0%	2.8%	2.7%	2.6%
LW (ccm)	1.3%	1.4%	1.2%	1.3%	1.2%	0.9%	0.8%	0.6%
	5.2%	4.9%	5.0%	5.7%	4.0%	4.5%	4.3%	4.7%
LW (sid)	1.4%	1.4%	1.2%	1.3%	1.2%	0.9%	0.8%	0.6%
	5.4%	5.0%	5.0%	5.8%	4.0%	4.6%	4.3%	4.8%
ML	1.3%	1.4%	1.2%	1.2%	1.2%	0.9%	0.9%	0.8%
	5.2%	5.2%	6.0%	8.6%	4.1%	4.7%	4.8%	5.7%
KZ	1.3%	1.3%	1.2%	1.3%	1.5%	1.4%	1.3%	1.4%
	3.5%	3.1%	3.1%	3.8%	3.2%	3.2%	3.0%	3.2%
PJ	1.3%	1.4%	1.2%	1.2%	1.2%	0.9%	0.9%	0.8%
	5.8%	5.3%	6.0%	8.5%	4.4%	4.8%	4.8%	5.7%
CLW (ccm)	1.4%	1.5%	1.5%	1.5%	1.6%	1.4%	1.5%	1.6%
	4.4%	4.0%	3.9%	3.9%	3.1%	2.9%	2.8%	2.6%
CLW (sid)	1.5%	1.5%	1.4%	1.5%	1.6%	1.5%	1.4%	1.5%
	4.4%	4.0%	3.8%	3.8%	3.1%	2.9%	2.7%	2.8%
CML	1.5%	1.5%	1.4%	1.4%	1.6%	1.5%	1.7%	1.7%
	4.3%	4.1%	4.3%	4.5%	3.0%	2.7%	2.6%	2.7%
CKZ	1.4%	1.3%	1.3%	1.4%	1.6%	1.5%	1.5%	1.6%
	3.7%	3.3%	3.5%	3.7%	3.0%	2.8%	2.6%	2.7%
CPJ	1.4%	1.4%	1.4%	1.5%	1.6%	1.3%	1.6%	1.6%
	4.4%	3.9%	3.8%	4.0%	3.2%	3.0%	2.6%	2.6%

Table B.3: Table of Empirical Returns & Standard Deviations, Panel 3 & 4

B.3 Convexity Plots

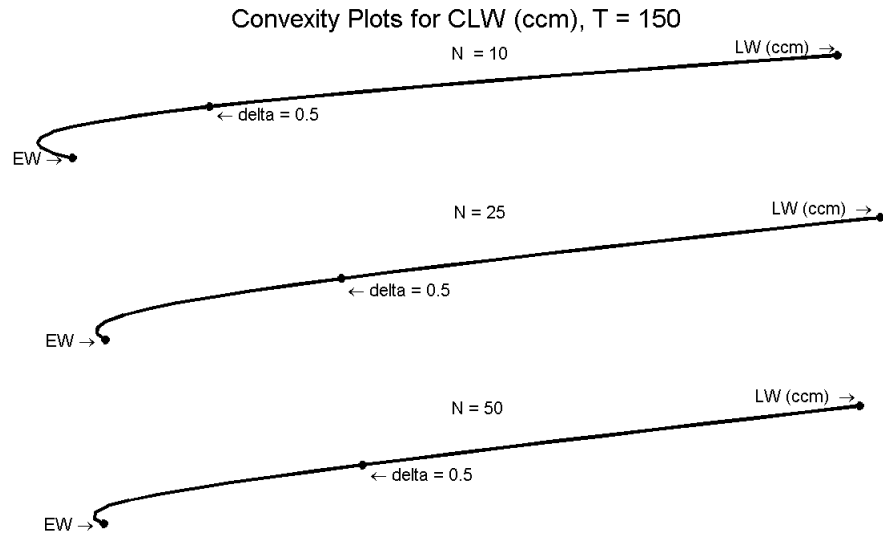


Figure B.4: Convexity Plot for the Proposed Rule, CLW (ccm).
 This curve is plotted in (σ, μ) -space for the proposed rule using the Ledoit & Wolf covariance estimator with the constant correlation model being the target matrix.

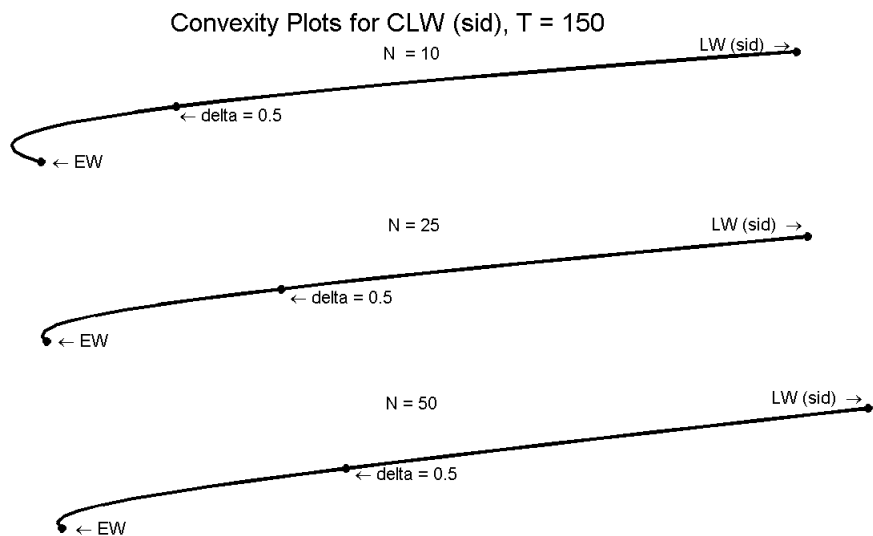


Figure B.5: Convexity Plot for the Proposed Rule, CLW (sid).
 This curve is plotted in (σ, μ) -space for the proposed rule using the Ledoit & Wolf covariance estimator with the scaled identity matrix being the target matrix.

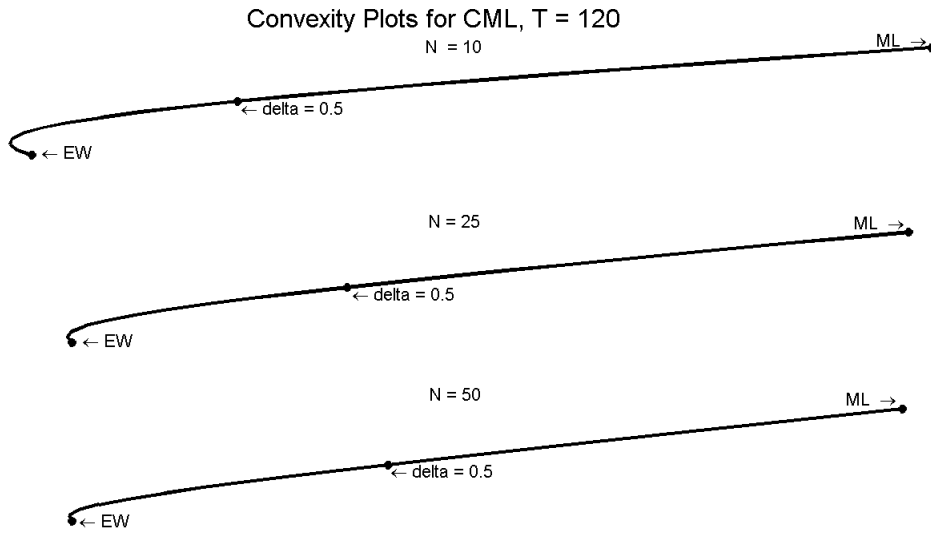


Figure B.6: Convexity Plot for the Maximum Likelihood rule, CML.
 This curve is plotted in (σ, μ) -space for the Markowitz rule in combination the equally weighted portfolio.

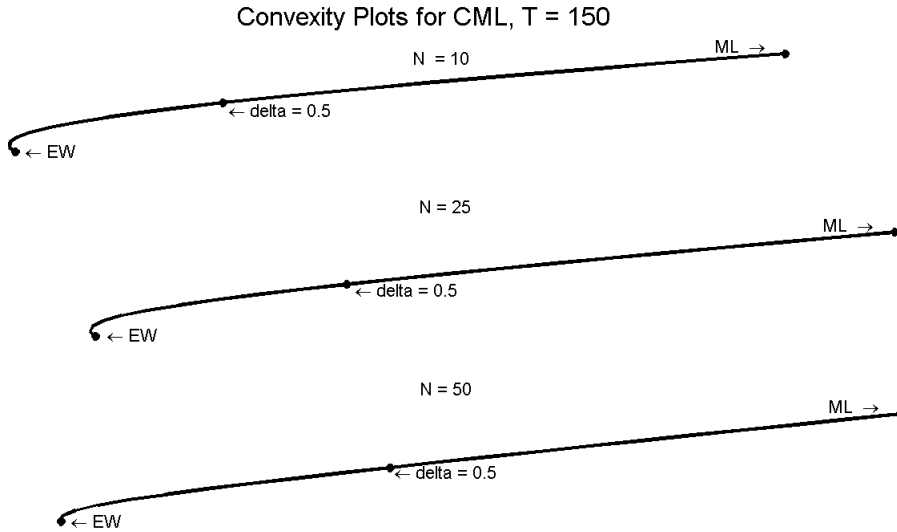


Figure B.7: Convexity Plot for the Maximum Likelihood rule, CML.
 This curve is plotted in (σ, μ) -space for the Markowitz rule in combination the equally weighted portfolio.

Appendix C

MATLAB FUNCTIONS

C.1 Functions for Estimated Combination Coefficients

C.1.1 The Combined Proposed Rule

```
function deltaStar =
propRule(mu, sampleMat, prior, theta, gamma, shrinkage, wEW, N, T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:
% This function estimates the optimal combination coefficient
% for the % proposed rule. The user inputs are
% propRule(mu, sampleMat, prior, theta, gamma, shrinkage, wEW, N, T)
% mu - 1 x N mean vector
% sampleMat - N x N sample covariance matrix
% prior - N x N target matrix
% theta - Estimator for the square Sharpe ratio,
% see function "thetaEst"
% gamma - risk aversion coefficient
% shrinkage - Shrinkage coefficient from covariance estimator
% wEW - vector of weights assigned 1/N
% N - Portfolio dimension
% T - Sample dimension

% Preliminaries
c1 = T/(T-N-2);
c2 = T^2*(T-2)/((T-N-1)*(T-N-2)*(T-N-4));
A = (1-shrinkage)*sampleMat;
B = prior*shrinkage;
```

```

I = eye(N);
P = A\inv(I + B/A)*B/A;

% Components
eta1 = wEW*sampleMat*wEW' - (2/gamma)*wEW*mu' + (theta/(2*gamma));

eta13 = (1/gamma)*((c2/(1 - shrinkage))*(wEW*mu'...
- wEW*sampleMat*P*mu' - theta) + mu*P*mu' - wEW*mu' + theta);

eta3 = (c2/gamma^2)*(N + T*theta)/(T*(1 - shrinkage)^2) -...
c1*(N + T.*mu*P*mu')/(T*(1 - shrinkage)) +...
(N + T.*mu*P*sampleMat*P*mu')/(T) +...
(2*c1*theta)/(gamma^2*(1 - shrinkage)) + mu*P*mu' + theta/gamma^2;

% Results
deltaStar = (eta1 - eta13)/(eta1 - 2*eta13 + eta3);
deltaStar = max(0, min(1, deltaStar));
end

```

C.1.2 The Combined Markowitz Maximum Likelihood Rule

```

function deltaStarM = markRule(mu, scaledMat, theta, gamma, wEW, N, T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:
% This function estimates the optimal combination coefficient for the
% traditional Markowitz Maximum Likelihood rule. The user inputs are
% markRule(mu, scaledMat, theta, gamma, wEW, N, T)
% mu - 1 x N mean vector
% scaledMat - N x N sample covariance matrix scaled to be unbiased
% theta - Estimator for the square Sharpe ratio, see function "thetaEst"
% gamma - risk aversion coefficient
% wEW - vector of weights assigned 1/N
% N - Portfolio dimension
% T - Sample dimension

%Components
c3 = ((T-2)*(T-N-2))/((T-N-1)*(T-N-4));

% Weight parts
eta1 = wEW*scaledMat*wEW' - (2/gamma)*wEW*mu' + (theta/(2*gamma));

```

```

eta2 = (c3 - 1)*theta + (c3*N)/(gamma^2*T);

% Results
deltaStarM = eta1/(eta1 + eta2);
deltaStarM = max(0,min(1,deltaStarM));
end

```

C.1.3 The Combined Kan & Zhou Rule

```

function [deltaStarKZ alpha] = kanzhouRule(...
    mu,mu_g,sampleMat,theta,psi,gamma,wEW,N,T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:
% This function estimates the optimal combination coefficient
% for the Kan & Zhou rule. The user inputs are
% kanzhouRule(mu,mu_g,sampleMat,theta,psi,gamma,wEW,N,T)
% mu - 1 x N mean vector
% mu_g - mean return on global minimum variance portfolio
% sampleMat - N x N sample covariance matrix
% theta - Estimator for the square Sharpe ratio,
% see function "thetaEst"
% psi - Estimator for the square Sharpe ratio with mean
% of GMV portfolio as benchmark, see function "psiEst"
% gamma - risk aversion coefficient
% wEW - vector of weights assigned 1/N
% N - Portfolio dimension
% T - Sample dimension

%preliminaries
ones_n = ones(1,N);
c2 = T^2*(T-2)/((T-N-1)*(T-N-2)*(T-N-4));
c3inv = ((T-N-1)*(T-N-4))/(T*(T-2));

%components
alpha = c3inv*(psi/(psi + (N/T)));
eta1 = wEW*sampleMat*wEW' - (2/gamma)*wEW*mu' + (theta/(2*gamma));
eta13 = (theta/gamma^2) - (1/gamma)*wEW*mu' + (1/(c2*gamma))*...
    ((alpha*wEW*mu' + (1 - alpha)*mu_g*wEW*ones_n') - (1/gamma)*...
    (alpha*mu/sampleMat*mu' + (1 - alpha)*mu_g*mu/sampleMat*ones_n'))

```

```
eta3 = (theta/gamma^2) - (1/(c2*gamma^2))*(theta - (N*alpha)/T);  
  
% results  
deltaStarKZ = (eta1 - eta13)/(eta1 - 2*eta13 + eta3);  
deltaStarKZ = max(0,min(1,deltaStarKZ));  
end
```

C.1.4 The Combined Jorion Rule

```
function [jorionMat muJ deltaStarJor] =
jorionRule(mu, sampleMat, theta, gamma, wEW, N, T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:
% This function estimates the optimal combination coefficient for the
% Phillipe Jorion rule. The user inputs are
% propRule(mu, sampleMat, theta, gamma, wEW, N, T)
% mu - 1 x N mean vector
% sampleMat - N x N sample covariance matrix
% theta - Estimator for the square Sharpe ratio,
% see function "thetaEst"
% gamma - risk aversion coefficient
% wEW - vector of weights assigned 1/N
% N - Portfolio dimension
% T - Sample dimension

% preliminaries
c2 = T/(T - N - 2);
c3 = ((T-2)*(T-N-2))/((T-N-1)*(T-N-4));
c4 = T/(T + 1);
ones_n = ones(1,N);
scaledMat = c2.*sampleMat; % scaled covariance matrix
I = eye(N);
mu_g = (mu*sampleMat*ones_n')/(mu*sampleMat*ones_n'); % grand mean
d = (mu - mu_g.*ones_n)/sampleMat*(mu - mu_g.*ones_n)';
w1 = (N + 2)/((N + 2) + T*d); % mean estimator weight
w2 = (N + 2)/((N + 2) + d + T*d); % covariance estimator weight
muJ = (1 - w1).*mu + w1.*mu_g*ones_n; % James-Stein mean estimator
% Jorion covariance estimator
jorionMat = scaledMat.*(1 + (1/T)) +...
w2*(ones_n'*ones_n)/(ones_n*scaledMat*ones_n');
A = scaledMat.*(1 + (1/T));
B = w2*(ones_n'*ones_n)/(ones_n/scaledMat*ones_n');
C = A\inv(I + B/A)*B/A;

% components
eta1 = wEW*sampleMat*wEW' - (2/gamma)*wEW*mu' + (theta/(2*gamma));
eta13 = (1/gamma)*(wEW*(c4.*I + C)*muJ' -...
```

```

wEW*mu' - (1/gamma)*mu*(inv(A) + C)*muJ' + theta);
eta3 = ((c3*c4^2)/gamma^2)*((N + T*(muJ/sampleMat*muJ'))/T) - ...
c4*((N + T*(muJ*C*muJ'))/T) ...
+ c4*((N + T*(muJ*C*sampleMat*C*muJ'))/T) - ...
(2/gamma^2)*mu*(inv(A) + C)*muJ' + (theta/gamma^2);

%results
deltaStarJor = (eta1 - eta13)/(eta1 - 2*eta13 + eta3);
deltaStarJor = max(0,min(1,deltaStarJor));
end

```

C.2 Support Functions

C.2.1 Function for Estimated ψ^2

```

function psiSquare_a = psiEst(psiSquare,N,T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:
% This function estimates  $\psi^2$  as given in Kan & Zhou (2007), Equation
% (67). It is approximated using the incomplete Beta distribution.
%
a = (N-1)/2;
b = (T-N + 1)/2;
x = psiSquare/(1 + psiSquare);
Bi = betainc(x,a,b);
psiSquare_a = ((T-N-1)*psiSquare - N + 1)/T + ...
(2*psiSquare^((N-1)/2)*(1 + psiSquare)^(-((T-2)/2)))/(T*Bi);
end

```

C.2.2 Function for Estimated θ^2

```

function thetaSquare_a = thetaEst(thetaSquare,N,T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:

```

```

% This function estimates theta^2 as given in Kan & Zhou (2007),
% Equation (51).
% It is approximated using the incomplete Beta distribution.
%
a = N/2;
b = (T-N)/2;
x = thetaSquare/(1 + thetaSquare);
Bi = betainc(x, a, b);
thetaSquare_a = ((T-N-2)*thetaSquare - N)/T + ...
    (2*thetaSquare^(N/2)*(1 + thetaSquare)^(-((T-2)/2)))/(T*Bi);
end

```

C.2.3 Function for Simulation of M-Variate Normal Matrix

```

function data = simMat(a,b,c,d,N,T)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015
%
% Description:
% This function simulates a multivariate normal random sample of
% dimension N and size T, for parameters mu and sigma drawn uniformly
% on the intervals (a,b) and (c,d) respectively.
%
data = zeros(T,N);
rMu = (a + (b-a).*rand(1,N))/12;
rSig = (c + (d-c).*rand(1,N))/sqrt(12);
for i = 1:N;
    r = rMu(:,i) + rSig(:,i).*randn(T,1);
    data(:,i) = r;
end
end

```

C.2.4 Compression of Matrix to Portfolio Size

```

function outputx = compressToPort(x,s)
% Author: Robin Jonsson
% Email: Rjn10005@student.mdh.se
% Course: Degree Project in Applied Mathematics, MMA891
% Master Thesis in Financial Engineering, Spring 2015

```



```

%
% Description:
% The function takes an N x T matrix "x" and compresses it do a
% desired sized matrix of column-dimension "s".

[t,n] = size(x);
% generate random binary vector of "s" ones and the rest zeros.
portSize = randOnes(n,s);
temp = portSize(ones(t,1),:); % Create binary matrix from vector
outputx = temp.*x; % Zero all undesired columns
outputx( :, ~any(outputx,1) ) = []; % shrink matrix to size
end

```