# Optimal portfolios – estimation and uncertainty assessment in the high-dimensional setting

Erik Thorsén

February 16, 2022

# Abstract

Financial portfolios and diversification go hand in hand. Diversification is one of, if not the, best risk mitigation strategy there is. If any investment that is part of the portfolio performs poorly, then it will not impact the performance of the portfolio because there are many other investments that may behave differently. Modern Portfolio Theory (MPT) is a framework for constructing diversified portfolios. However, MPT relies on unknown parameters which need to be estimated. These parameters are the two first moments of the return distribution. By replacing them with some estimates, estimation uncertainty is introduced to the allocation problem. If the uncertainty is not well understood or taken into account, it has the potential to dominate any characteristic of the investment strategy. There is no way to determine if the strategy works or not.

This thesis contains five papers which provide results on how to deal with estimation uncertainty in high-dimensional and, in some cases, finite sample portfolios from the MPT framework. These results provide tools to better understand the investment process and the empirical results that can be observed.

Paper 1 explores all of the portfolios that can be placed in the framework of MPT. The paper provide the sampling distribution for all optimal portfolios and their characteristics. Their joint distribution is charactericed in terms of a stochastic representation. This is done by assuming that the returns follow a multivariate normal distribution. Furthermore, the high-dimensional asymptotic joint distribution for the quantities of interest is derived. The high-dimensional distribution can be thought of as an infinitely diversified portfolio together with an infinite amount of data to estimate the parameters. The high-dimensional distribution can be seen to provide a good approximation to the finite sample one in a simulation study.

Paper 2 continues on the idea of paper 1. It considers the quadratic utility allocation problem from paper 1 but with an additional risk-free asset in the portfolio. This portfolio in known as the Tangency Portfolio (TP). The distribution of the sample TP weights is derived but under a more general model in comparison to paper 1. The asset returns are assumed to follow a closed skew-normal distribution. Results show that skewness implies that the sample TP weights is on average biased in finite sample. Although there is a bias in finite sample, the sample TP weights are consistent in high dimensions.

Paper 3 takes on a practical aspect of investing, how to transition from one portfolio to another. A reallocation scheme is developed, which minimizes the out-of-sample variance of the Global Minimum Variance (GMV) portfolio, given a holding portfolio. The holding portfolio can be deterministic or a random GMV portfolio from the previous period. Results show that this reallocation scheme can achieve great results in terms of estimating a relative loss in an extensive simulation study. The relative loss is a simple extension of the portfolio variance. Furthermore, the empirical application shows that the reallocation scheme can provide the smallest out-of-sample variance in comparison to a number of benchmarks. It also provides the smallest change in portfolio weights between reallocation points. The theoretical results from this paper is implemented in the DOSPortfolio R package.

Paper 4 derives properties of two different performance measures for three different high-dimensional GMV portfolio estimators. The performance measures are the out-of-sample variance and the out-of-sample loss. The former is always used as an evaluation metric for an empirical application of the GMV portfolio. The results show that the out-of-sample loss does not need the same stringent assumptions as the out-of-sample variance in the high-dimensional setting. It can therefore cover many more models. Using the out-of-sample loss, an ordering of

the three different portfolios is provided. This ordering is verified in a simulations study and an empirical application.

Paper 5 extends the results of one of the high-dimensional GMV portfolios used in paper 3 and 4. It introduces Thikonov regularization to the portfolio weights which results in a Ridge type estimator for the sample covariance matrix combined with the linear shrinkage from paper 3 and 4. The portfolio solution covers the case where there are more assets in comparison to the number of datapoints. The portfolio performance is investigated in a large simulation study and its performance is studied in an empirical application. A simulation study shows that the method is comparable to a number of benchmarks. Furthermore, an empirical application show that it can provide the lowest out-of-sample variance and provide good characteristics for the portfolio weights.

# List of included papers

This thesis is based on the following papers

## Papers included in this thesis

Bodnar, Taras, Holger Dette, Nestor Parolya, and Erik Thorsén (2022a). "Sampling distributions of optimal portfolio weights and characteristics in small and large dimensions". In: *Random Matrices: Theory and Applications.*

Javed, Farrukh, Stepan Mazur, and Erik Thorsén (2021). "Tangency portfolio weights under a skew-normal model in small and large dimensions". In: *Research report in Mathematical Statistics, 2021:8.*

Bodnar, Taras, Nestor Parolya, and Erik Thorsén (2021a). "Dynamic shrinkage estimation of the high-dimensional minimum-variance portfolio". In: *arXiv preprint arXiv:2106.02131.*

Bodnar, Taras, Nestor Parolya, and Erik Thorsén (2021b). "Is the empirical out-of-sample variance an informative risk measure for the high-dimensional portfolios?" In: *arXiv preprint arXiv:2111.12532.*

Bodnar, Taras, Nestor Parolya, and Erik Thorsén (2022). "Two is better than one: Regularized shrinkage of largeminimum variance portfolio". In: *Research report in Mathematical Statistics, 2022:3.*

## Papers & other research results which are not included in this thesis

Bodnar, Taras, Mathias Lindholm, Vilhelm Niklasson, and Erik Thorsén (2022a). "Bayesian quantile-based portfolio selection". In: *Under revision in Applied Mathematics and Computation.*

Bodnar, Taras, Mathias Lindholm, Erik Thorsén, and Joanna Tyrcha (2021a). "Quantile-based optimal portfolio selection". In: *Computational Management Science*, pp. 1–26.

Bodnar, Taras, Vilhelm Niklasson, and Erik Thorsén (2021a). "Volatility sensitive Bayesian estimation of portfolio VaR and CVaR". In: *Research report in Mathematical Statistics, 2021:6.*

Bodnar, Taras, Nestor Parolya, and Erik Thorsén (2021c). *DOSPortfolio: dynamic optimal shrinkage of high-dimensional portfolios.* R package version 0.1. URL: `https://CRAN.R-project.org/package=DOSPortfolio`.

**Author contribution:** E. Thorsén has actively contributed to developing the content of all papers. In the first paper, E. Thorsén contributed with the derivation of the theoretical results and with writing the majority of the paper in collaboration with all other authors. The original idea was presented by T. Bodnar. In the second paper, E. Thorsén contributed with the theoretical derivations and writing the majority of the paper in collaboration of S. Mazur and F. Javed. The original idea was presented by S. Mazur and F. Javed. The original idea of the third paper was born from discussions with T. Bodnar and N. Parolya. In this paper, E. Thorsén contributed with the theoretical derivations and the writing of the paper. The original idea of the fourth paper stemmed from discussions from the third paper together with T. Bodnar and N. Parolya. In this paper, E. Thorsén contributed with the theoretical derivations and the writing of the paper. The fifth paper also sprung from the third and fourth paper. In this paper, E. Thorsén contributed to the theoretical derivations, the writing of the paper and the idea behind it. E. Thorsén carried out all computer implementations in all articles. In all other aspects the authors contributed equally.

**General comment:** An earlier version of Bodnar, Dette, Parolya, and Thorsén (2022b) was included in the Licenciate thesis of Erik Thorsén, Thorsén (2019).

# Acknowledgements

Early in my graduate studies I asked a professor what it takes to complete a PhD and he answered "grit". The purpose of a PhD is to spend a number of years intensively digging deeper into a subject, coming up with ideas that most often have been solved, does not seem to be solvable or cannot be solved at all. It is a very humbling experience and there are many people that I would like to thank for making it a fun one as well!

The first person to thank is my supervisor Taras Bodnar. I am immensely grateful for your guidance, knowledge and discussions over the years. Thank you for giving me the opportunity of doing this. I could not have asked for a better supervisor. I also want to thank my co-supervisor Joanna Tyrcha for your support.

I also want to give special thanks to my co-authors that also made this an exciting and great learning experience. Nestor Parolya, whom I have gotten the pleasure to work extensively with over the last few years. You are a co-author of almost all the papers in this thesis and it has been a joy. I have learned a lot from our discussions. Thank you, Stepan Mazur and Farrukh Javed for the fun discussions and great collaborations! Furthermore, thank you Vilhelm Niklasson, Mathias Lindholm and Holger Dette for the (very) long and rewarding discussions!

Thank you Ola Hössjer for letting me be a part in developing and teaching the course Statistical Data Processing. I have learned so much from the experience and grown as a person from it!

I also want to thank all my colleagues at the Department of Mathematics. It has been a incredibly fun place to work and I think that the environment has been, and will continue to be, great. There are a number of individuals that have made my experience even better. Thank you Vilhelm, Gustav and Stanislas for being wonderful people to share an office with. I hope to continue our brewing adventures Vilhelm and the wonderful banter Gustav! I also want to thank Felix, Måns, Sebastian and Hampus whom i have gotten to spend a lot of time with during covid. It made everything so much easier.

Last but not least, I cannot express enough love and gratitude towards my old (and new!) family. Mom, dad and my beloved sister, I cannot thank you enough for all your support. Johanna, you make me a much happier person and I look forward to many more years with you and our son.

# Contents

# Part I

# Introduction

# 1. Decisions, uncertainty and asset allocations

A portfolio is any combination of assets. An asset might be a house, a stock, the value of a contract or any object with a price that can potentially be bought or sold. The price of an asset develops over time. The value of a house in Stockholm, Sweden, has increased a lot in the past years[1]. However, the future price for a house is uncertain given that house prices are close to all-time highs. Will the price continue to increase or will there be a decrease? If there is a change, how big will that be? Depending on the answer, the optimal choice might be to sell the house today. However, to make a decision on whether or not the house should be sold today, there is a need to specify its future value. The future is unknown or random, in some sense. There are many possible changes in the assets value, both in direction and magnitude. In the context of a model there is a possibility to estimate the future price through the use of historical prices. The decision process now includes even more uncertainty. There is uncertainty from the model, the perception of past and future prices, and the uncertainty from using estimates instead of the "true" parameters for the model. If the house is one asset among many then there can be as many sources of uncertainty as there are assets, if not more. The allocation itself, in this case selling the house, is uncertain. Uncertainty is everywhere in decisions.

The market price of a house is rarely observed. As a matter of fact, it is only observed when it is sold. Even though prices are rarely observed there might be houses that are similar. There are many different observations from the same population. That might not be the case for the second asset that is listed above, a stock. The frequency of which the price of a stock is observed can be much higher. On the other hand, there is only one stock with one accompanying timeseries. This makes the analysis slightly different. There can be a lot of historical data, but no identical copies of the stock that can be used to compare and extract information from. Prices can only be observed as they develop.

In this thesis, the aim is not to predict tomorrows price of an asset but to answer what amount of a budget should be spend on the different assets in the long run. The methods in this thesis build upon the seminal work of Markowitz (1959), what is known as Modern Portfolio Theory (MPT). Markowitz argued that any portfolio which simply maximizes its profit will result in a naive solution. Investing all the capital in the asset with the highest return is not sensible if the future is not known. Such an investment will be extremely risky. As a consequence, Markowitz argued that any well diversified portfolio should be preferred to any non-diversified portfolio. A well-diversified portfolio can be obtained through many different procedures. Markowitz proposed the use of the mean and variance of the asset return distribution for the allocation problem. If an asset has high return on average, it would make sense to invest in it if return is all that matters. If an asset has a large variance then it is risky to invest in this asset. A high return might be at the cost of a large amount of risk. If an asset is not risky, then it would always make sense to invest in it.

To practically use a portfolio from the MPT framework the parameters of the model, the mean vector and the covariance matrix, need to be estimated. As previously said, this introduces estimation uncertainty. If the amount of data is small, then there is a large risk that the estimated

[1]As Metallica says "Sad but true", for the author of this thesis `https://www.maklarstatistik.se/omrade/riket/stockholms-lan/#/villor/48m-prisutveckling`

portfolio will not be accurate and volatile as new data is observed. The sampling distribution of the portfolio will have a large variance. One important portfolio in the MPT framework is the Global Minimum Variance (GMV) portfolio. This portfolio provides the smallest variance in the MPT framework. Okhrin and Schmid (2006) derived the exact sampling distribution of the GMV portfolio. In the paper, the authors showed that the portfolio can have very fat tails depending on the sample size and the number of assets in the portfolio. This is the first sign of a tradeof between diversification and the increasing influence of estimation uncertainty. The portfolio has been extensively researched in many different contexts to cope with this feature. Frahm and Memmel (2010) extended the portfolio weights and constructed an estimator which combines the GMV portfolio with a target portfolio, much like Stein (1956) derived a regularization method for the sample mean vector. Frahm and Memmel (2010) used properties of the sample covariance matrix to develop a regularization method for the GMV portfolio. The method provides a way to combine the target and the GMV portfolio. Furthermore, Kempf and Memmel (2006) showed that an estimate of the GMV portfolio can also be obtained through a regression approach. Due to the vast literature in regression models, this approach has also spurred much research of the MPT framework (see, e.g., Maillet, Tokpavi, and Vaucher (2015) and the references therein).

High-dimensional portfolios can be thought of as a consequence of an infinite amount of diversification. Diversification is one, if not the best, risk management tool. It should, in theory, decrease the risk of the portfolio. That is not always the case. By introducing one new asset to the portfolio, more parameters than the number of assets in the portfolio need to be estimated. The covariance matrix suffers from the curse of dimensionality. In terms of estimation uncertainty, this does not scale well. There are many ways to solve this. Ledoit and Wolf (2020) and Bodnar, Parolya, and Schmid (2018) consider estimating high-dimensional GMV portfolios using two different approaches. The former assumes that the eigenvectors are known and consider a nonlinear (rotational-invariant) estimation method for the covariance matrix. The latter develops a regularization method much like Frahm and Memmel (2010) but does so for high-dimensional portfolios. The mentioned papers do not assume any specific structure on the asset return distribution. Another common model for asset returns, which impose some strucutre, is the factor model (see, e.g., Ross (1976)). A factor can be many things, such as the area in which the house is located. Ding, Y. Li, and Zheng (2021) use such a model to describe the implications of estimation to the GMV portfolio. The high-dimensional GMV portfolio appears in other fields as well, such as signal processing where it goes under the name beamformer (see, e.g., J. Li, Stoica, and Wang (2004)).

The largest argument for using the GMV portfolio in the MPT framework is that it does not depend on the mean vector. Estimating the mean is hard (see, e.g., Merton (1980), Best and Grauer (1991)). However, it is only one out of many portfolios in the MPT framework. Including a target for the portfolio mean results in the mean-variance portfolio. El Karoui (2010) investigated the weights of the mean-variance portfolio when linear constraints are introduced to the portfolio allocation problem. That can include position restrictions as well as an investors desire for a given return. Bodnar, Okhrin, and Parolya (2021) considered high-dimensional portfolios but use another portfolio allocation problem, namely the quadratic utility. The setting is similar to Bodnar, Parolya, and Schmid (2018) where the authors regularize the quadratic utility portfolio towards a target portfolio in the high-dimensional setting. Karlsson, Mazur, and Muhinyuza (2021) used an extension to the quadratic utility, where a risk-free asset is introduced. There they derived different statistical properties of its solution, known as the tangency portfolio.

The rest of this thesis is structured as follows. Chapter 2 presents the framework that this thesis will work within, namely MPT. As previously described, MPT use information which is not available. It relies on parameters which are not known. In the subsequent chapter, Chapter 3, the models used in this thesis are introducd. It also introduces the more formal concept of estimation uncertainty and what it means for MPT. Chapter 4 presents what happens in the MPT framework when the investor diviersifies as much as possible. The portfolio will contain many, if not an infinite number, of assets. The last two chapters, Chapter 5 and 6 provide a summary of the papers presented in this thesis as well as a summary of future possible research.

All code for this thesis is available on Github at https://github.com/Ethorsn/Phd-thesis.

# 2. Modern portfolio theory

Let the asset return $\mathbf{y}$ be a $p$-dimensional random vector with mean vector $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$ and covariance matrix $\mathrm{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$. The matrix $\boldsymbol{\Sigma}$ is of dimensions $(p \times p)$. Although there are usually little restriction on $\boldsymbol{\mu}$ there are very specific restrictions on the covariance matrix $\boldsymbol{\Sigma}$. Since the covariance matrix is a subject of its own there will be a whole section dedicated to it. Its restrictions are disregarded for now. Using the two moments for the asset returns, the portfolio return is defined as $y = \mathbf{w}^\top \mathbf{y}$. It has mean $\mathrm{E}(y) = \mathbf{w}^\top \boldsymbol{\mu}$ and variance $\mathrm{Var}(y) = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$. Let $\mu_0$ be the target return that the investor would like to achieve from his/her portfolio and $\mathbf{1}$ column vector of ones with appropriate dimension. Markowitz (1959) considered the following optimization problem

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^\top \mathbf{1} = 1, \\ & \mathbf{w}^\top \boldsymbol{\mu} \geq \mu_0 \\ & w_i \geq 0, i = 1, 2, .., p. \end{aligned} \tag{2.1}$$

The objective is to minimize the variance of the portfolio. The constraint $\mathbf{w}^\top \mathbf{1} = 1$ states that the investor must invest all available money. The weights are scaled according to the amount of cash invested. The disposition is very different whenever an inequality is used rather than equality. As Hult, Lindskog, Hammarlid, and Rehn (2012) states, if $\mathbf{w}^\top \mathbf{1} \leq 1$, then the investor could be throwing money away since there is a lot of opportunity left in the market when investing. The second constraint describes the investors expectation on the portfolio. As $\mu_0$ grows, the return of the portfolio will grow. However, increasing $\mu_0$ will change the amount of variance the portfolio can achieve, there is a risk-return trade-off. The last constraint is rather simple though it can have quite large implications. It states that the investor can only invest with money he/she has. A negative value of $w_i$ in the $i$th asset is called a short position. The asset is borrowed from someone who owns it and then sold, hoping that it will be cheaper in the future. For certain types of investors this constraint can be limiting and for others it is a must. In this thesis, it is excluded altogether. That is, this thesis considers

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^\top \mathbf{1} = 1, \\ & \mathbf{w}^\top \boldsymbol{\mu} \geq \mu_0 \end{aligned} \tag{2.2}$$

which is referred to the Mean-Variance (MV) optimization problem. The solution to this problem is very often stated in terms of another famous portfolio, namely the Global Minimum Variance (GMV) portfolio and its related quantities (see, e.g., Bodnar and A. Gupta (2009)). Let $\boldsymbol{\Sigma}^{-1}$ denote the inverse matrix of $\boldsymbol{\Sigma}$, e.g., $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} = \mathbf{I}$, and

$$\mathbf{w}_{GMV} := \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}, \; R_{GMV} := \mathrm{E}(\mathbf{w}_{GMV}^\top \mathbf{y}) = \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}, \; V_{GMV} := \mathrm{Var}(\mathbf{w}_{GMV}^\top \mathbf{y}) = \frac{1}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{2.3}$$

The GMV portfolio can be obtained by letting $\mu_0 = R_{GMV}$ or by removing the constraint $\mathbf{w}^\top \boldsymbol{\mu} \geq \mu_0$ from (2.2). The solution to (2.2) is equal to

$$\mathbf{w}_{MV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} + \frac{\mu_0 - R_{GMV}}{V_{GMV}} \mathbf{Q}\boldsymbol{\mu}, \; \mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{2.4}$$

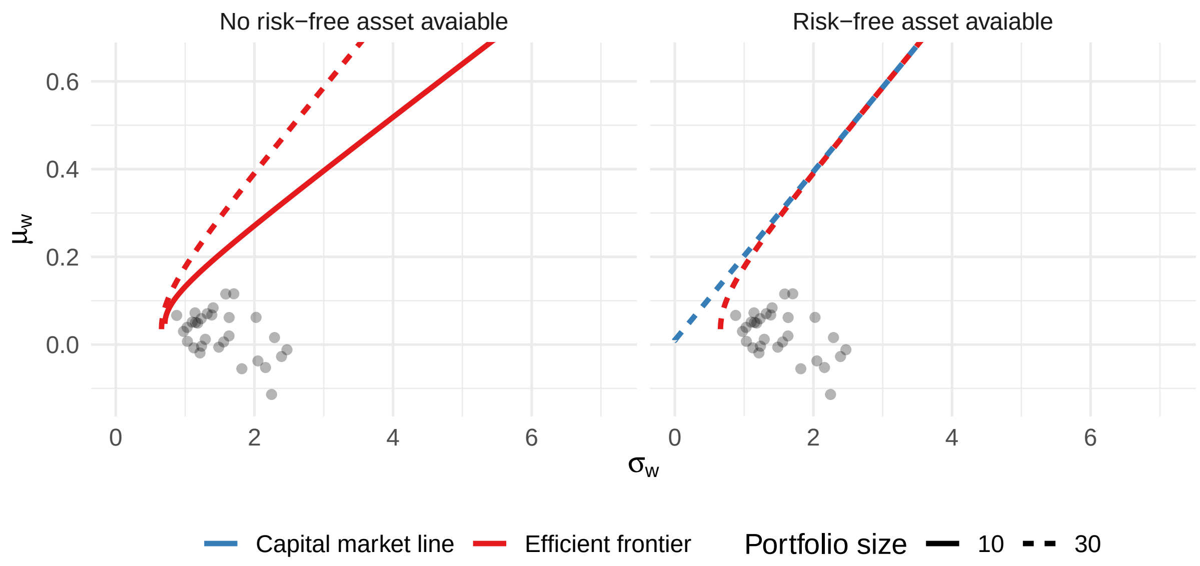


Figure 2.1: Efficient frontiers without and with a risk-free asset. The left plot illustrates two different efficient frontiers for different portfolio sizes. The right plot illustrates the efficient frontier and the capital market line which appears when a risk-free asset is available. The stocks are randomly selected from the S&P500. The individual means and standard deviations are displayed as points.

The solution is a combination of two different portfolios, the GMV portfolio and the self-financing portfolio $\mathbf{Q}\boldsymbol{\mu}$. The ratio $(\mu_0 - R_{GMV})/V_{GMV}$ acts as a weight to how much is allocated in the self-financing portfolio. Since $\mathbf{Q}\boldsymbol{\mu}$ is self-financing, increasing $\mu_0$ does not cost anything, it will simply increase the position in the self-financing portfolio.

The moments of the MV portfolio are equal to

$$\mathrm{E}(\mathbf{w}_{MV}^\top \mathbf{y}) = R_{GMV} + \frac{\mu_0 - R_{GMV}}{V_{GMV}} \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu}, \ \mathrm{Var}(\mathbf{w}_{MV}^\top \mathbf{y}) = V_{GMV} + \left(\frac{\mu_0 - R_{GMV}}{V_{GMV}}\right)^2 \boldsymbol{\mu}^\top \mathbf{Q}\boldsymbol{\mu}. \tag{2.5}$$

From equation (2.5) it can be observed that all values $\mu_0$ are rescaled according to the moments of the GMV portfolio. The two expressions in (2.5) constitute a famous relationship. As $\mu_0$ increases, the risk grows quadratic in comparison to the mean which is linear. This relationship was discovered by Merton (1972) which coined the expression "the efficient frontier". In Figure 2.1 we display the efficient frontier and the capital market line, which is yet to be described, for two different portfolio sizes. As seen in the left figure, increasing the portfolio size shifts the location of the parabola, e.g., moves it to the left, which serves as an illustration of the diversification effect. There is no guarantee that an increase in the portfolio dimension increases the return.

Any point on any of the two lines in the left hand-side plot of Figure 2.1 corresponds to a certain efficient and optimal portfolio with a specific value of $\mu_0$. The points in the two plots depict the expected returns and the standard deviations of a single-stock portfolio. Diversification will always decrease the risk of the portfolio which can be seen in the illustration. No single-stock portfolio can outperform any of the portfolios on the efficient frontier. That cannot happen.

The right hand-side plot of Figure 2.1, displays an efficient frontier and the capital market line. It is the solution to an extension to the mean-variance problem. It displays what happens when a risk-free asset is included in the portfolio allocation problem. The risk-free asset is added to the portfolio as any other asset, with the exception that it is deterministic. With the option to invest in a risk-free asset, the return of the portfolio is $w_0 r_f + \mathbf{w}^\top \mathbf{y}$. The MV problem from

(2.2) is then equal to

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & w_0 + \mathbf{w}^\top \mathbf{1} = 1 \\ & w_0 r_f + \mathbf{w}^\top \boldsymbol{\mu} = \tilde{\mu}_0. \end{aligned} \tag{2.6}$$

However, since $w_0 + \mathbf{w}^\top \mathbf{1} = 1$ the amount invested in the risk-free asset can be substituted by $w_0 = 1 - \mathbf{w}^\top \mathbf{1}$. The problem in (2.6) reduces to an unconstrained optimization problem. Its solution is given by

$$\mathbf{w}_{TP} = \frac{(\tilde{\mu}_0 - r_f)}{(\boldsymbol{\mu} - r_f \mathbf{1})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1})} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}). \tag{2.7}$$

The collection of portfolios given by (2.6) defines the whole capital market line which is shown in Figure 2.1. The portfolio has many interesting properties. If there is a risk-free asset, then there is a possibility to increase the return and decrease the risk of the portfolio. This is most easily explained by the efficient frontier, displayed in Figure 2.1. For a given level of risk a portfolio can always get the same, and sometimes larger, return!

The same solution can be obtained from the optimization problem with the quadratic utility, defined as

$$\min_{\mathbf{w}} \mathbf{w}^\top (\boldsymbol{\mu} - r_f \mathbf{1}) - \frac{\gamma}{2} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$$

for some $\gamma > 0$. The solution is given by

$$\frac{1}{\gamma} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1})$$

which coincide with (2.7) if

$$\frac{1}{\gamma} = \frac{(\tilde{\mu}_0 - r_f)}{(\boldsymbol{\mu} - r_f \mathbf{1})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1})}.$$

The difference is that (2.7) depends on a number of parameters while $\gamma$ is a fixed constant. This is quite common in MPT and there are many portfolio allocation problems which result in the same solution, see, e.g., Bodnar, Parolya, and Schmid (2013).

Ever since the end of 2014, there has been a lack of a risk-free asset in Sweden.[1] The risk-free rate has been less than or equal to zero. Assuming that it is true for a hypothetical investor, (2.7) reduces to $\mathbf{w}_{TP} = \tilde{\mu}_0 \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} / \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. The term $\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ is also present in (2.4), although hidden in the matrix $\mathbf{Q}$. With a little work, (2.4) can be rewritten as

$$\left(1 - \frac{\mu_0 - R_{GMV}}{V_{GMV}^2} R_{GMV}\right) \mathbf{w}_{GMV} + \frac{\mu_0 - R_{GMV}}{V_{GMV}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}.$$

There are two insights to be drawn from this equation. The first is that the weights on the efficient frontier is a combination of two portfolios, in this case the GMV and the tangency portfolio. This result is usually known as the Mutual fund theorem, see Tobin (1958). All portfolios on the efficient frontier can be studied through these two portfolios. The second is that the tangent portfolio, where the efficient frontier and the capital market line meet, is given by (2.7) with

$$\tilde{\mu}_0 = R_{GMV} + \frac{\mu_0 - R_{GMV}}{V_{GMV}} s.$$

Any portfolio with $\tilde{\mu}_0 \leq R_{GMV} + \frac{\mu_0 - R_{GMV}}{V_{GMV}} s$ will be "more efficient" than the efficient frontier if there is a risk-free rate. However, its not obvious that the amount of cash allocated in a risk-free asset can be optimized. Given that cash is free any investor should most likely borrow as much as possible to invest in the market. That might be a risk in itself which is not covered by the model.

All of MPT use the inverse covariance matrix. The next section is devoted to the assumption and restrictions made on the covariance matrix.

[1] See this visualization from Riksbanken

## 2.1 Relationship between assets and the (inverse) covariance matrix

The covariance matrix $\mathbf{\Sigma}$ and the precision matrix $\mathbf{\Sigma}^{-1}$ are fundamental to mean-variance portfolio theory. This section goes into some depth on the assumptions and restrictions that are placed on the covariance matrix.

For a vector $\mathbf{y}$ with finite second moment, the covariance matrix is defined as $\mathbf{\Sigma} = \mathrm{E}((\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^\top)$. It contains the variances of each individual element of $\mathbf{y}$ on the diagonal as well as the covariances between every pair of elements on the off-diagonal. That is, each diagonal element corresponds to the univariate case with variance equal to $\mathrm{E}((y_i - \mu_i)^2)$. In the univariate case, a distribution is usually called degenerate or singular if the variance is equal to zero. In the multivariate case, the covariance matrix can be singular on a number of occasions. It is not limited to the diagonal elements. This is due to the fact that covariances are involved on the off-diagonal. From Harville (1997, ch 14.2) a real symmetric $p \times p$ matrix $\boldsymbol{A}$ is called

- positive definite if $\mathbf{z}^\top \boldsymbol{A}\mathbf{z} > 0$
- positive semi-definite if $\mathbf{z}^\top \boldsymbol{A}\mathbf{z} \geq 0$

for all nonzero vectors $\mathbf{z} \in \mathbb{R}^p$. Let $\boldsymbol{A} > 0$ or $\boldsymbol{A} \geq 0$ denote a positive or semi-positive definite matrix $\boldsymbol{A}$. The concept of a singular or degenerate distribution is replaced by a quadratic form. It involves the covariance matrix of a multi- or matrixvariate distribution. The definition of positive or semi-positive definiteness can be quite cumbersome to work with. The conditions need to hold for all vectors $\mathbf{z}$. A necessary condition for a matrix to be positive definite can be derived using the eigenvalues of a matrix and its eigenvalue decomposition, as described in Harville (1997, ch. 21).

**Definition 2.1.1.** Let $\boldsymbol{A}$ be a $p \times p$ matrix. The characteristic roots (with multiplicity) are given by the solutions to

$$|\boldsymbol{A} - \lambda \mathbf{I}| = 0$$

where $|\cdot|$ is the determinant of a matrix.

Let $\lambda_i$, $i = 1, 2, ..., p$, denote the *ordered* eigenvalues of the matrix $\boldsymbol{A}$ such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. Given an eigenvalue $\lambda_i$, the eigenvectors $\mathbf{u}_i$ are defined by $\boldsymbol{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$, $i = 1, 2, ..., p$. Let $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p)$ and $\mathbf{U} = (\mathbf{u}_1^\top, \mathbf{u}_2^\top, ..., \mathbf{u}_p^\top)^\top$. The relation between the matrix $\boldsymbol{A}$ and its eigenvalues and eigenvectors can also be written as

$$\boldsymbol{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \tag{2.8}$$

which is called the eigen or spectral decomposition. A necessary condition for a matrix to be positive definite can be directly obtained from the eigenvalue decomposition. Let $\mathbf{z} \in \mathbb{R}^p$ and $\boldsymbol{a} := \mathbf{U}^\top \mathbf{z} \in \mathbb{R}^p$, then $\mathbf{z}^\top \boldsymbol{A}\mathbf{z} = \mathbf{z}^\top \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{z} = \boldsymbol{a}^\top \mathbf{\Lambda} \boldsymbol{a} = \sum_i^p \lambda_i a_i^2$ which is a second degree polynomial. If the eigenvalues are all positive, then necessarily the matrix is positive definite. If there are some eigenvalues which are zero, then the matrix is semi-positive definite.

All papers of this thesis assume that the true covariance matrix is positive definite. The assumption has a very important economical interpretation. If one (or more) eigenvalue(s) are zero, then there is a possibility to construct a portfolio which does not contain any risk. It is an arbitrage opportunity unless the elements of $\boldsymbol{\mu}$ are all zero. Assume $\lambda_p = 0$, let $\mathbf{u}_p$ be its eigenvector and set $\mathbf{w} = \mathbf{u}_p / \sum_i^p u_{ip}$. The variance of that portfolio is zero since all eigenvectors are orthonormal and its mean is $\mathbf{w}^\top \boldsymbol{\mu}$. If the true covariance matrix is not positive definite there might exist arbitrage opportunities, e.g., the possibility of making profit without taking any risk.

# 3. Statistical models and inference

To *pratically* use the portfolios described by (2.4) the two parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have to be specified. This is not really feasible if the portfolio contains many assets. There might be an opinion of what they should be, however these cannot be known precisely. Furthermore, even if there is an informed opinion of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the potential loss of using those exact parameters might be paramount. Therefore, the parameters are usually estimated through data. Let $p_{i,t}$ be the asset price of the $i$th asset at time $t$. The methods of this thesis rarely use the asset prices themselves but a transformation of the relative differences, that is, their simple and log-returns. The simple return is defined as $r_{i,t} := (p_{i,t} - p_{i,t-1})/p_{i,t-1}$ and the log-return is then defined as $y_{i,t} := \log(r_{i,t} + 1)$ and $\mathbf{y}_t = (y_{1,t}, y_{2,t}, ..., y_{p,t})$. The return of a portfolio with $p$ assets is modeled as $\mathbf{w}^\top \mathbf{y}_t$ where $\mathbf{w} = (w_1, ..., w_p)$ are the portfolio weights. This is an approximation. In reality the portfolio return should be $\sum_{i=1}^p w_i r_{i,t}$ (or even $\sum_{i=1}^p w_i p_{i,t}$). That is additive in the number of assets. However, logarithmic returns are additive in time which can be desirable. Compounding returns is simple addition and the approximation can make the statistical analysis more tractable. The difference between the two approaches is very small if the (log) returns are small, which is often true for financial assets, see Tsay (2010, p. 5).

There are many ways to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The most simple and versatile method is the Method of Moments (MM) (see e.g., Wasserman (2004, ch. 9)). Let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)$ be a sample of log-returns. Using the sample, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is replaced with the sample mean and the sample covariance matrix, i.e.,

$$\bar{\mathbf{y}} = \frac{1}{n}\sum_{i}^{n} \mathbf{y}_i, \ \mathbf{S} = \frac{1}{n}\mathbf{Y}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{Y}^\top.$$

This is always a feasible approach assuming that the first two moments actually exist. However, it introduces some issues. If the sample size $n$ is small, then the estimates are naturally imprecise. Furthermore, MPT relies on $\mathbf{S}^{-1}$ and not $\mathbf{S}$. That puts further restrictions on the sample size and the size of the portfolio, it demands that $n > p$. It is natural to ask: does an imprecise estimate of the covariance matrix provide an equally imprecise estimate of the inverse? In some simple cases the answer is no, sometimes it is worse. It is therefore very important to understand the implications of not using the true parameters but their sample counterparts. There are many approaches to this but the most natural is the bottom-up approach. If the asset returns $\mathbf{Y}$ follow some distribution, then there is most likely some statistical properties for $\bar{\mathbf{y}}$ and $\mathbf{S}$. In turn, $\mathbf{S}^{-1}$ can also have some statistical properties and in the end all the transforms given by (2.4). The models and their properties that are used in this thesis, are stated in the coming sections. Thereafter the implications of using estimates for MPT are discussed in brief.

## 3.1 Matrixvariate distributions

One of the most fundamental models for asset returns is the multivariate normal distribution. Since most of the distributions in this thesis are matrix variate, the matrixvariate normal distribution is presented. It is slightly more general but it can capture much more structure in the data.

**Definition 3.1.1** (Definition 2.2.1 A. K. Gupta and Nagar (2000))**.** The random matrix $\mathbf{Y}$ $(p \times n)$ is said to have a matrix variate normal distribution with mean matrix $\mathbf{M}$ and covariance matrix $\boldsymbol{\Sigma} \otimes \boldsymbol{\Gamma}$ where $\boldsymbol{\Sigma} > 0$ is of dimension $(p \times p)$ and $\boldsymbol{\Gamma} > 0$ is of dimension $(n \times n)$, if $\text{vec}(\mathbf{Y}^\top) \sim N_{np}(\text{vec}(\mathbf{M}^\top), \boldsymbol{\Sigma} \otimes \boldsymbol{\Gamma})$.

The multivariate normal distribution is a simple special case of it with $\boldsymbol{\Gamma} = \mathbf{I}$ and $n = 1$. A property of the matrixvariate normal distribution is that it is closed under linear transformations. That plays an important role for determining the distribution of the estimator $\bar{\mathbf{y}}$ but also to determine the portfolio return distribution. Both are linear transformations. This model is used very often and the applications are many. However, Cont (2001) described a number of stylized facts of log-returns. These stylized facts describe the characteristics of asset returns on different frequencies. It is argued that the multivariate normal distribution has too thin tails in comparison to what is usually observed in asset returns on higher frequencies. Cont argued that daily returns are usually not symmetric and often show volatility clustering. However, Cont also argued that returns on a lower frequency such as monthly or quarterly can be close to normal. The shape of the unconditional return distribution is not the same over different frequencies. A motivation for the multivariate normal model can be thought of as an investor which invests in the portfolio infrequently. That does not mean that he/she cannot observe the results of the market on a higher frequency!

In the univariate case the sample variance follows a chi-square distribution. If the returns follow a multivariate normal distribution and are independent, then $\mathbf{S}$ follows what is known as a Wishart distribution. It is essentially a generalization of the chi-square distribution. The probability density function (p.d.f.) of a Wishart disitrbution is given below.

**Definition 3.1.2** (Definition 3.2.1 A. K. Gupta and Nagar (2000))**.** A $p \times p$ random symmetric positive definite matrix $\mathbf{S}$ is said to have a Wishart distribution with parameter $n$ $(n \geq p)$ and $\boldsymbol{\Sigma} > 0$, $(p \times p)$ written as $\mathbf{S} \sim W_p(n, \boldsymbol{\Sigma})$ if its p.d.f. is given by

$$\frac{|\mathbf{S}|^{(n-p-1)/2}|\boldsymbol{\Sigma}|^{-n/2}}{2^{pn/2}\Gamma_p(n/2)} \exp\left\{-\frac{1}{2}\,\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\} \tag{3.1}$$

where $\Gamma_p(\cdot)$ is the multivariate gamma function.

In comparison to the normal distribution the Wishart distribution is used very frequently as a model for covariance matrices, although in a slightly different context. The model is very often used for realized covariance matrices, see Barndorff-Nielsen and Shephard (2004), Golosnoy, Gribisch, and Seifert (2019) or Alfelt (2021). A realized covariance matrix is an estimate of the volatility process from returns on a much higher frequency than what is used in this thesis.

If $\mathbf{Y} \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^\top, \boldsymbol{\Sigma}\otimes\mathbf{I}_n)$, then $n\mathbf{S} \sim W(n-1, \boldsymbol{\Sigma})$ by A. K. Gupta and Nagar (2000, Theorem 3.3.6). Working from the bottom-up there is now a model for the estimated parameters of the original model. As previously stated, MPT works with inverse covariance matrices and not with the covariance matrix itself. Thankfully, the Wishart distribution has an inverse counterpart.

**Definition 3.1.3** (Definition 3.4.1 A. K. Gupta and Nagar (2000))**.** A random matrix $\mathbf{V}$ is said to be distributed as an inverted Wishart distribution with $m$ degrees of freedom and parameter matrix $\boldsymbol{\Gamma}$ $(p \times p)$, denoted by $\mathbf{V} \sim IW_p(m, \boldsymbol{\Gamma})$, if its density is given by

$$\frac{2^{-(m-p-1)p/2}|\boldsymbol{\Gamma}|^{(m-p-1)/2}}{\Gamma_p((m-p-1)/2)|\mathbf{V}|^{m/2}} \exp\left\{-\frac{1}{2}\mathbf{V}^{-1}\boldsymbol{\Gamma}\right\},\ m > 2p, \mathbf{V}, \boldsymbol{\Gamma} > 0. \tag{3.2}$$

To once more connect to the univariate setting, the inverted sample variance follows an inverted chi-square distribution which is a special case of the inverted gamma distribution. It is only natural that the inverted Wishart matrix is a matrix variate generalization of the inverted gamma distribution (A. K. Gupta and Nagar (2000, page 111)). It demands quite specific constraints on the parameters of the model, namely $m > 2p$. This is a hint to an answer that was posed in the beginning of this section, does the inverse change uncertainty? The constraint

$m > 2p$ provides a hint that *something* changes with the properties of $\mathbf{S}$ when taking inverses. The constraints of the model are more strict. From Theorems 3.3.7 and 3.4.1, and Theorem 3.4.3 of A. K. Gupta and Nagar (2000) the first moments of the two are

$$\mathrm{E}\left[\mathbf{S}\right] = \frac{n-1}{n}\mathbf{\Sigma},\ \mathrm{E}\left[\mathbf{S}^{-1}\right] = \frac{n}{n-p-2}\mathbf{\Sigma}^{-1}.$$

If $n$ is sufficiently large, then the sample covariance matrix is (close to) unbiased. That is not necessarily the case for its inverse. If an investor believes in diversification, then he/she should own many assets, e.g., $p$ should be large. That in turn could make the estimator very biased! Inverses can potentially make matters worse.

It has been well established that returns on higher frequency the normal distribution can be a poor model. The next feature, and perhaps most common, to include is skewness of the asset returns and to assess its effect on portfolios. A $p$ dimensional Closed Skew Normal (CSN) random vector $\mathbf{z}$ has density

$$f_{\mathbf{z}}(\boldsymbol{a};\boldsymbol{\mu},\mathbf{\Sigma},\mathbf{D},\mathbf{v},\Delta) = C\phi_p(\boldsymbol{a};\boldsymbol{\mu},\mathbf{\Sigma})\Phi_q(\mathbf{D}(\boldsymbol{a}-\boldsymbol{\mu});\mathbf{v},\Delta) \tag{3.3}$$

where $C$ is a normalization constant and $\boldsymbol{\mu},\mathbf{\Sigma},\mathbf{D},\mathbf{v}$ and $\Delta$ are parameters of appropriate dimensions. Its matrixvariate counterpart is simply defined through the vec operator.

**Definition 3.1.4** (Definition 3.1 Dominguez-Molina, Gonzaez-Farias, Ramos-Quiroga, and A. K. Gupta (2007))**.** A random matrix $\mathbf{Y}$ $(p \times n)$ is said to have a matrix variate closed skew-normal distribution with parameters $\mathbf{M}$ $(p \times n)$, $\boldsymbol{A}$ $(np \times np)$, $\mathbf{B}$ $(nq \times mp)$, $\mathbf{L}$ $(q \times m)$ and $\mathbf{Q}$ $(mq \times mq)$, with $\boldsymbol{A} > 0$ and $\mathbf{Q} > 0$ if

$$\mathrm{vec}(\mathbf{Y}^\top) \sim CSN_{pm,qn}\left(\mathrm{vec}(\mathbf{M}^\top), \boldsymbol{A}, \mathbf{B}, \mathrm{vec}(\mathbf{L}^\top), \mathbf{Q}\right). \tag{3.4}$$

The closed skew-normal distribution is heavily parametrized. For each column in the matrix $\mathbf{Y}$ there is a mean vector and an additional four vectors $\boldsymbol{a}$, $\mathbf{b}$, $\mathbf{l}$ and $\mathbf{q}$ describing skewness and volatility of the asset returns. The matrixvariate distribution can capture a lot of structure. However, it also puts a heavy restriction on some of the parameters, as they need to be positive definite. The parameter $\boldsymbol{A}$ might be interpreted as a covariance matrix although that is a simplification. It is something more. It can capture variance along *both* axis of the matrix $\mathbf{Y}$, such as volatility for a specific asset but also unconditional volatility over time. There is also some type of dependence between $\boldsymbol{A}$ and how skewness is observed, by the fact that moments include *almost all of the parameters* (see, e.g., Proposition 3.2 Dominguez-Molina, Gonzaez-Farias, Ramos-Quiroga, and A. K. Gupta (2007)). Due to the stochastic representation in Proposition 2.1 Dominguez-Molina, Gonzaez-Farias, Ramos-Quiroga, and A. K. Gupta (2007), the introduction of skewness can be thought of as random shocks to the mean. [1] The overzealous parametrisation make estimating the parameters rather difficult and that is why paper 2 use the assumption that $q = m = 1$.

The last model in this thesis is the most general. It is also the model that has the least amount of interesting properties in itself. It is the following location and scale model

$$\mathbf{Y} \overset{d}{=} \boldsymbol{\mu}\mathbf{1}_n^\top + \mathbf{\Sigma}^{1/2}\mathbf{Z} \tag{3.5}$$

where $\overset{d}{=}$ stands for equality in distribution and $\mathbf{Z} = \{z_{ij}\}$, $i = 1, 2, ..., p$, $j = 1, 2, ..., n$. It is very common to assume moment conditions on $z_{ij}$, such as finite fourth moment or potentially $4 + \epsilon$ finite moment, for some $\epsilon > 0$. Although the model can capture many types of return distributions, such as skew, heavy tailed and sometimes even heteroscedasticity, there is very little to say about it.

Although estimates of the model parameters are interesting in themselves, this thesis is about portfolios. In the next section, the implications of estimation uncertainty on optimal portfolios is displayed.

[1] The proposition is a little bit misleading as there seems to be an absolute value missing and the parameter $\mathbf{v}$ appears as random but is also part of the parametrization. To arrive at the correct representation, the reader can go through the steps that begins at the end of page 1606 in the same reference.

## 3.2 Inference and sampling distributions of optimal portfolios and their characteristics

The empirical counterpart of (2.4) is obtained by replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with $\bar{\mathbf{y}}$ and $\mathbf{S}$. It is equal to

$$\hat{\mathbf{w}}_{MV} = \frac{\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}} + \frac{\mu_0 - \hat{R}_{GMV}}{\hat{V}_{GMV}}\hat{\mathbf{Q}}\bar{\mathbf{y}},\ \hat{\mathbf{Q}} = \mathbf{S}^{-1} - \frac{\mathbf{S}^{-1}\mathbf{1}\mathbf{1}^\top\mathbf{S}^{-1}}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}} \tag{3.6}$$

where

$$\hat{V}_{GMV} = \frac{1}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}},\ \hat{R}_{GMV} = \frac{\mathbf{1}^\top\mathbf{S}^{-1}\bar{\mathbf{y}}}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}} \tag{3.7}$$

and the shape parameter of the efficient frontier $\hat{s} = \bar{\mathbf{y}}^\top\hat{\mathbf{Q}}\bar{\mathbf{y}}$. This is a portfolio that can actually be invested in.

From the previous section, the distributions of $\bar{\mathbf{y}}$ and $\mathbf{S}$ are known if $\mathbf{Y}$ follows the matrix-variate normal distribution stated in Definition 3.1.1. In that scenario, the two estimators are even independent (see Theorem 3.3.6 A. K. Gupta and Nagar (2000)) which makes the analysis simpler. However, in MPT there are many complicated transforms containing both $\bar{\mathbf{y}}$ and $\mathbf{S}^{-1}$. One example is the return for the GMV portfolio

$$\hat{R}_{GMV} = \frac{\bar{\mathbf{y}}^\top\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}}. \tag{3.8}$$

Since $\bar{\mathbf{y}}$ and $\mathbf{S}$ are independent then the conditional distribution of (3.8) on $\bar{\mathbf{y}}$ is the same as the unconditional distribution. However, (3.8) still contains the sample covariance matrix in the nominator as well as the denominator. It is not safe to assume that these are independent nor is it trivial to state when they would be. To solve the problem at hand consider two $p \times p$ matrices with the following block structure

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix},\ \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \tag{3.9}$$

where $\dim(\boldsymbol{A}_{11}) = \dim(\mathbf{V}_{11}) = m \times m$, $m < p$. Let $\boldsymbol{A}_{11\cdot 2} := \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ denote the Schur complement of the matrix $\boldsymbol{A}$ and define $\mathbf{V}_{11\cdot 2}$ in the same manner. Let $\otimes$ denote the Kronocker product. The following theorem is used a lot in papers 1 and 2.

**Theorem 3.2.1** (Theorem 3 in Bodnar and Okhrin (2008)). *Suppose $\boldsymbol{A} \sim W_k^{-1}(n, \mathbf{V})$, where $\boldsymbol{A}$ and $\mathbf{V}$ are partitioned as in* (3.9)*. Then*

*(a)* $\boldsymbol{A}_{11\cdot 2} \sim W_m^{-1}(n-k+m, \mathbf{V}_{11\cdot 2})$ *and is independent of* $\boldsymbol{A}_{22}$*;*

*(b)* $\boldsymbol{A}_{12}|\boldsymbol{A}_{22}, \boldsymbol{A}_{11\cdot 2} \sim \mathcal{N}(\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\boldsymbol{A}_{22}, \boldsymbol{A}_{11\cdot 2} \otimes \boldsymbol{A}_{22}\mathbf{V}_{22}^{-1}\boldsymbol{A}_{22})$*;*

*(c)* $\boldsymbol{A}_{22} \sim W_{p-m}^{-1}(n-2m, \mathbf{V}_{22})$*;*

*(d)* $\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$ *is independent of* $\boldsymbol{A}_{22}$*, with density given by*

$$f_{\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}}(\mathbf{X}) = \frac{|\mathbf{V}_{11\cdot 2}|^{-\frac{1}{2}(p-m)}|\mathbf{V}_{22}|^{\frac{1}{2}m}}{\pi^{\frac{(p-m)m}{2}}} \frac{\Gamma_m\left(\frac{n-m-1}{2}\right)}{\Gamma_m\left(\frac{n-p-1}{2}\right)} \times \left|\mathbf{I} + \mathbf{V}_{11\cdot 2}^{-1}\left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)\mathbf{V}_{22}\left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)^\top\right|^{-\frac{1}{2}(n-m-1)} \tag{3.10}$$

*where $\Gamma_m(\cdot)$ is the multivariate Gamma function;*

*(e)* $\boldsymbol{A}_{22}$ *is independent of* $\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$ *and* $\boldsymbol{A}_{11\cdot 2}$*;*

*(f)* $\boldsymbol{A}_{11\cdot 2}|\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} = \mathbf{X} \sim W_m^{-1}(n, \mathbf{V}_{11\cdot 2} + \left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)\mathbf{V}_{22}\left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)^\top)$

So given an inverse Wishart distribution the distribution of many, quite difficult, transformations of its sub-matrices can be derived. Let $\mathbf{H}^\top = (\mathbf{L}^\top, \tilde{\mathbf{y}}, \mathbf{1})^\top$ and note that $(\mathbf{H}\mathbf{S}^{-1}\mathbf{H}^\top)^{-1}$ follows a Wishart distribution by Theorem 3.3.13 A. K. Gupta and Nagar (2000). The inverse of $(\mathbf{H}\mathbf{S}^{-1}\mathbf{H}^\top)^{-1}$ follows an inverse Wishart distribution. Observe that

$$\mathbf{H}\mathbf{S}^{-1}\mathbf{H}^\top = \begin{pmatrix} \mathbf{L}\mathbf{S}^{-1}\mathbf{L}^\top & \mathbf{L}\mathbf{S}^{-1}\mathbf{1} & \mathbf{L}\mathbf{S}^{-1}\tilde{\mathbf{y}} \\ \tilde{\mathbf{y}}^\top\mathbf{S}^{-1}\mathbf{L}^\top & \tilde{\mathbf{y}}^\top\mathbf{S}^{-1}\tilde{\mathbf{y}} & \tilde{\mathbf{y}}^\top\mathbf{S}^{-1}\mathbf{1} \\ \mathbf{1}^\top\mathbf{S}^{-1}\mathbf{L}^\top & \mathbf{1}^\top\mathbf{S}^{-1}\tilde{\mathbf{y}} & \mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1} \end{pmatrix}.$$

In this case the product between $\boldsymbol{A}_{12}$ and $\boldsymbol{A}_{22}^{-1}$ together with the above results in

$$\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} = \begin{pmatrix} \frac{\mathbf{L}^\top\mathbf{S}^{-1}\tilde{\mathbf{y}}}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}} & \frac{\tilde{\mathbf{y}}^\top\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}} \end{pmatrix}^\top$$

which is the joint distribution of the return of the GMV portfolio, $\hat{R}_{GMV}$, and some scaled version of linear combinations of the tangency portfolio with $\gamma = \hat{V}_{GMV}^{-1}$ and $r_f = 0$. By (3.10) the joint distribution of these follows a matrixvariate t-distribution (see A. K. Gupta and Nagar (2000, Definition 4.2.1)) and the return of the GMV portfolio is independent to its variance. Similar to the matrixvariate normal distribution, the matrixvariate t-distribution is also closed under linear transformations (see A. K. Gupta and Nagar (2000, Theorem 4.3.5)). This implies that $\hat{R}_{GMV}$ follows a t-distribution, conditionally on the mean. The parameters of the distribution are presented in Lemma 7.1 in paper 1. Paper 1 uses these properties to derive the full joint distribution for all the quantities used in (3.6). The result is more general than that. It provides the joint distribution of all optimal portfolios. The joint distribution is characterized through its stochastic representation. The stochastic representation is a very verbose way of characterizing the distribution in terms of simple random variables. These random variables are simple to simulate. Any quantity of interest from the joint distribution can simply be computed through Monte Carlo approximation. For some methods, simulations are the only way to compute the quantities of interest. It can therefore be very important that simulations are fast. This is extremely simple to do with the stochastic representation.

## 3.3 Simulations, inverses and why stochastic representations are valuable

Assume that the investor cares about simulations, is interested in the GMV portfolio and that the returns follow a matrixvariate normal distribution, i.e. $\mathbf{Y} \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^\top, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$. To simulate from the sampling distribution of the variance of the GMV portfolio, there are a number of steps that needs to be performed

1. Simulate $\mathbf{Y}$ and construct $\mathbf{S}$
2. Invert $\mathbf{S}$
3. Compute $\hat{V}_{GMV}$

The second step is notoriously demanding. The default method to use in R is 'solve' which is a wrapper for certain LAPACK[2] functions. The inverse itself takes $2p^3$ flops (cpu cycles), which is not cheap (see, e.g., Higham (2002, ch 14)). If $p$ is large, then simulation of the quantity $\hat{V}_{GMV}$ will be extremely cumbersome. Another method is R's chol2inv which relies on the Cholesky decomposition. In theory, it should be faster but demands that the Cholesky decomposition is computed. The last two options that are available is to simulate $\mathbf{S}$ directly or to derive the stochastic representation of $\hat{V}_{GMV}$. Paper 1 and 2 go into great detail to derive the stochastic representation of different quantities of optimal portfolios. One of them is the sample variance

[2]For the interested reader `https://www.netlib.org/lapack/`

of the GMV portfolio. If $\mathbf{Y} \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^\top, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, then by Theorem 1 in Bodnar, Dette, Parolya, and Thorsén (2022b) $\hat{V}_{GMV} \sim V_{GMV}\xi/(n-1)$ where $\xi \sim \chi^2_{n-p}$. The inversion can be omitted all together. In Chunk 3.3.1 there is some R-code which implements a small benchmark to highlight why these types of representations can be really valuable.

Chunk 3.3.1: R-code for benchmarking different simulation approaches of the variance of the GMV portfolio.

```
# setup
p <- 150
n <- 250
Sigma <- HDShOP::RandCovMtrx(p)
Sigma_chol <- chol(Sigma)
mu <- runif(p, -0.1, 0.1)
Sigma_inv <- solve(Sigma)
V_GMV <- 1/sum(Sigma_inv)
# microbechmark
result <- microbenchmark(
  # Simulate Y directly, construct S, invert and compute GMV variance
  `Scenario 1` = {
    Y <- mu %*% t(rep(1,n)) + t(Sigma_chol)%*%matrix(rnorm(n*p), ncol = n)
    S <- var(t(Y))
    1/sum(solve(S))
  },
  # Simulate Y directly, construct S and its chol. decomp., use chol2inv and
  # compute GMV variance
  `Scenario 2` = {
    Y <- mu %*% t(rep(1,n)) + t(Sigma_chol)%*%matrix(rnorm(n*p), ncol = n)
    S <- var(t(Y))
    S_chol <- chol(S)
    1/sum(chol2inv(S_chol))
  },
  # Simulate S directly, invert and compute GMV variance
  `Scenario 3` = {
    S <- rWishart(1, df=n-1, Sigma=Sigma)[,,1]
    1/sum(solve(S))
  },
  # Simulate directly from the GMV sample variance distribution.
  `Scenario 4` = V_GMV/(n-1) * rchisq(1, df=n-p),
  times=1000
)
```

In Figure 3.1 a violinplot of the benchmark is displayed. Scenario 4 uses the stochastic representation and the time its benchmark is much smaller any of the former strategies. It is quite clear that it is the fastest. The conclusion is that inversions are very cumbersome to deal with and take a lot of time regardless if the Cholesky decomposition is used or not. It can also be a very unstable operation, especially if the matrix is close to singular.

A large portion of this thesis work with $\mathbf{S}$, $\bar{\mathbf{y}}$. That is of course a simplification and not always the case. The estimators $\mathbf{S}$ and $\bar{\mathbf{y}}$ contains a lot of uncertainty where the former contains less than the latter (see, e.g., Frankfurter, Phillips, and Seagle (1971), Merton (1980), Best and Grauer (1991)). Furthermore, there are a number of other issues at hand. If $p$ is comparable to $n$ but $n > p$, then the expectation of the inverse Wishart distribution is very biased. It is very hard to compute $\mathbf{S}^{-1}$ from $\mathbf{S}$ numerically if the matrix is close to singular. There are two startegies

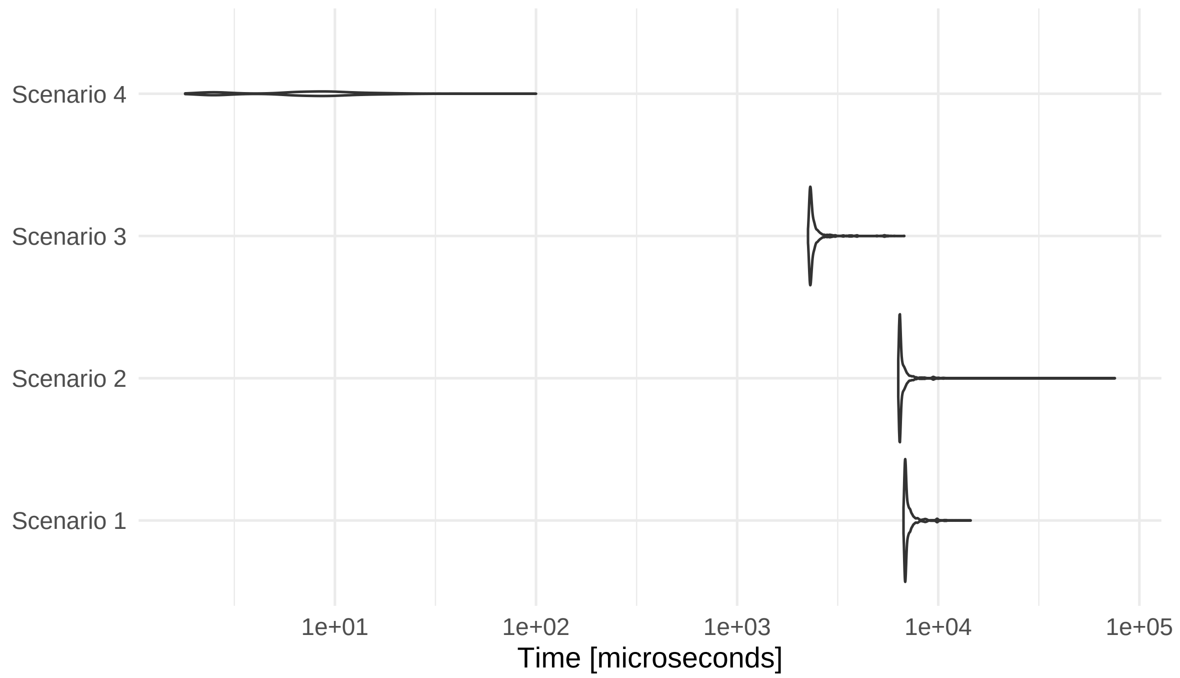


Figure 3.1: Difference in performance between the simulation methods for the estimated variance of the GMV portfolio based on 1000 simulations.

to solve the problem. The first is to derive the actual uncertainty and sample distributions of the quantities of interest, which is done in papers 1 and 2. The second is to use other estimators which introduce some bias. By introducing bias in the estimator it can reduce the variance of it. As will be seen in the next section, it is something of utmost importance if an investor believes in diversification.

# 4. Diversification, infinitely many assets and shrinkage estimators

In the previous chapter one specific way of estimating the sample covariance matrix was presented. If there is a lot of data on the assets that is in the portfolio, then the estimated weights will most likely be close to the correct portfolio. The estimated portfolio will be consistent, e.g., it estimates the correct object of interest. As previously mentioned, diversification is one of the best risk management tool there is and therefore, the asset universe is supposed to be big. However, Bodnar, Okhrin, and Parolya (2016) Proposition 2.2 states that $\hat{V}_{GMV} \to V_{GMV}/(1-c)$ whenever $p, n \to \infty$ s.t. $p/n \to c \in [0, 1)$. If $c$ is close to one, then the sample GMV portfolio variance will explode. *Estimation uncertainty dominates the diversification effect.* There are many solutions to the problem at hand (see, e.g., Ledoit and Wolf (2017) and the references therein). The solution used in this thesis is Random Matrix Theory (RMT) and the use of some types of shrinkage estimators. Both subjects are grand. A small introduction to them is presented in the following sections.

## 4.1 A short introduction to random matrix theory and the Stieltjes transform

The subject of RMT has many applications. It was originally developed in the context of quantum physics (see Mehta (2004, Chapter 1 of)). The theory and its applications have since then developed quite a lot. Many fields, such as combinatorics, computational biology, wireless communication and finance (see, e.g., Livan, Novaes, and Vivo (2018)) use these results. One of the seminal work in RMT was made by Wigner (1967). Wigner originally modeled the limiting spectral distribution of a $p \times p$ dimensional random matrices $\mathbf{X}$ with standard normal entries. The term "standard" might be a little misleading for statisticians as the matrix $\mathbf{X}$ contains independent random variables although not identically distributed. The entries on the diagonal are $N(0, 2)$ and the entries on the off-diagonal are $N(0, 1)$.

The empirical spectral distribution (ESD) of a matrix $\boldsymbol{A}$ is defined as

$$F^{\boldsymbol{A}}(x) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}(\lambda_i \le x)$$

where $\lambda_i$ are the eigenvalues from the eigenvalue decomposition, see section 2.1. The limit, in this case, is taken as $p \to \infty$ which implies that $\boldsymbol{A}$ will have infinitely many columns as well as rows. The limiting spectral distribution of $\mathbf{X}$ can be shown to converge to (see, Chapter 2 of Bai and Silverstein (2010))

$$F'(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & \text{if } |x| \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

There are many interesting facts about the empirical spectral distribution and its limiting distribution. One of the most interesting is the support of the limiting distribution. The normal distribution has unbounded support but the eigenvalues of $\mathbf{X}$ converges to a distribution with bounded support (see, Livan, Novaes, and Vivo (2018) for a good explanation to the reason why). Marchenko and Pastur (1967) extended the result of Wigner (1967) to the sample covariance matrix. Assume that $\mathbf{X}$ is a $p \times n$ matrix that contains i.i.d random variables with zero mean and variance equal to 1. The limit is now taken over the two quantities $p$ and $n$ at the same time, such that $p/n$ tends to a constant. The ratio is usually called the concentration ratio $c$. In this introduction it is assumed that $c < 1$. The density of the limiting spectral distribution of $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^\top$ was then shown to be

$$F'(x) = \begin{cases} \frac{1}{2\pi x c}\sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. The distribution has, once again, bounded support! The eigenvalues seem to attract each other. Although the sample covariance matrix appears very often in the context of MPT, it is not the object of interest. MPT cares about its inverse, as discussed in Chapter 2, which the Stieltjes transform can help with. The Stieltjes transform of a function $F : \mathbb{R} \to \mathbb{R}$ is defined as

$$m^F(z) = \int \frac{1}{x-z} dF(x) \tag{4.1}$$

where $z \in \{z \in \mathbb{C} : \mathbb{Im}(z) > 0\}$. The Stieltjes transform has many useful properties. If the Stieltjes transform is known, then the spectral distribution $F$ can be derived by the inversion formula. There is also pointwise convergence of it (see, appendix B.2 of Bai and Silverstein (2010)). To apply the results from RMT to MPT, take a sample covariance matrix $\mathbf{S}$ with ESD $F_n(x)$ and eigenvalue-matrix $\mathbf{\Lambda}$ as defined in Chapter 2, and note that

$$\frac{1}{p}\operatorname{tr}\left(\mathbf{S}^{-1}\right) = \lim_{z \to 0^+} \frac{1}{p} \operatorname{tr}\left((\mathbf{\Lambda} - z\mathbf{I})^{-1}\right) = \lim_{z \to 0^+} \int_0^\infty \frac{1}{x-z} dF_n(x) = \lim_{z \to 0^+} m^{F_n}(z). \tag{4.2}$$

To investigate the limiting properties of traces of inverse sample covariance matrices then all that is needed is the Stieltjes transform and its properties. However, to make matters slightly worse, the objects of interest in this thesis are quadratic or bilinear forms where the inverse sample covariance matrix is present. Examples are $\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}$ or $\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{b}$ for some vector $\mathbf{b}$. Although $\operatorname{tr}(\mathbf{S}^{-1})$ and $\mathbf{1}^\top\mathbf{S}^{-1}\mathbf{1}$ may look similar, their limiting objects can behave quite differently. This is due to the fact that the former does not depend on the eigenvectors while latter does. Rubio and Mestre (2011) showed the following theorem which can be used to derive limiting objects on this specific form.

**Theorem 4.1.1** (Theorem 1 of Rubio and Mestre (2011)). *(a)* $\mathbf{X}$ *is an* $p \times n$ *random matrix such that the entier of* $\sqrt{n}\mathbf{X}$ *are i.i.d complex random variables with mean 0, variance 1 and finite* $8+\epsilon$ *moment, for some* $\epsilon > 0$.

*(b)* $\boldsymbol{A}$ *and* $\mathbf{R}$ *are* $p \times n$ *hermitian nonnegative definite matrices, with the spectral norm (denoted by* $||\cdot||$*) of* $\mathbf{R}$ *being bounded uniformly in* $p$*, and* $\mathbf{T}$ *is an* $n \times n$ *diagonal matrix with real nonegative entries uniformly bounded in* $n$.

*(c)* $\mathbf{B} = \boldsymbol{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$*, where* $\mathbf{R}^{1/2}$ *is the nonnegative definite square root of* $\mathbf{R}$.

*(d)* $\mathbf{\Theta}$ *is an arbitrary nonrandom* $p \times p$ *matrix, whose trace norm (i.e.,* $\operatorname{tr}((\mathbf{\Theta}^H\mathbf{\Theta})^{1/2}) := ||\mathbf{\Theta}||_{tr}$*) is bounded uniformly in* $p$.

*Then, with probability 1, for each* $z \in \mathbb{C} - \mathbb{R}^+$*, as* $n = n(p) \to \infty$ *such that* $0 < \liminf c_p < \limsup c_p < \infty$*, with* $c_p = p/n$

$$\operatorname{tr}\left(\mathbf{\Theta}\left((\mathbf{B} - z\mathbf{I})^{-1} - (\boldsymbol{A} + x_p(e_p)\mathbf{R} - z\mathbf{I})^{-1}\right)\right) \to 0 \tag{4.3}$$

*where* $x_p(e_p)$ *is defined as*

$$x_p(e_p) = \frac{1}{n}\operatorname{tr}\left(\mathbf{T}\left(\mathbf{I}_n + c_p e_p \mathbf{T}\right)^{-1}\right) \tag{4.4}$$

*and $e_p = e_p(z)$ is the Stieltjes transform of a certain positive measure on $\mathbb{R}^+$ with total mass $\mathrm{tr}(\mathbf{R})/p$, obtained as the unique solution in $\mathbb{C}^+$ of the equation*

$$e_p = \frac{1}{p}\,\mathrm{tr}\left(\mathbf{R}\left(\boldsymbol{A} + x_p(e_p)\mathbf{R} - z\mathbf{I}_p\right)^{-1}\right). \tag{4.5}$$

This theorem is used repeatedly in papers 3, 4 and 5. However, in this thesis it is assumed that $\mathbf{X}$ has finite $4+\epsilon$ moment, while the theorem above assumes $8+\epsilon$. That can be circumvented through the supplement material of Bodnar, A. K. Gupta, and Parolya (2016).

Through Theorem 4.1.1 any limiting trace of the modified sample covariance matrix $\mathbf{B}$ can be derived by first finding $x_p(e_p)$, the transformation of the Stieltjes transform $e_p(z)$. In simple cases, an analytic solutions for the functions $x_p(e_p)$ or $e_p(z)$ can be found. This happens in papers 3 and 4 though not in paper 5. In paper 5 the solution is found using numerical methods to construct optimal portfolios.

## 4.2 Shrinkage estimators in modern portfolio theory

The estimator $\hat{V}_{GMV}$ is clearly biased, it even diverges when $c$ approaches 1. This problem is not unique. The least squares estimator is usually very volatile when there are many covariates in the regression model. An easy solution is to use $(1-c)\hat{V}_{GMV}$ as an unbiased estimator for the variance of the GMV portfolio. However, that might not be what an investor wants. He/she invests in portfolio weights, so naturally the aim should be to construct a good estimator for these. There are many solutions to this problem but the most common is to use a shrinkage estimator. By introducing a shrinkage estimator, some bias is introduced to the weights but hopefully it also reduces the variance.

Papers 3 through 5 work with the GMV portfolio and there are two natural extensions to it. The weights $\hat{\mathbf{w}}_{GMV}$ or the sample covariance matrix $\mathbf{S}$ can be shrunk towards a target. Papers 3 and 4 work with the former. Combining the GMV portfolio weights with some target portfolio $\mathbf{b}$ through a linear combination results in

$$\hat{\mathbf{w}}_{SH} = \alpha\hat{\mathbf{w}}_{GMV} + (1-\alpha)\mathbf{b}$$

which introduces the bias $\alpha(\mathrm{E}[\hat{\mathbf{w}}_{GMV}] - \mathbf{w}_{GMV})$ but decreases the variance to

$$\alpha^2\,\mathrm{E}\left[\left(\hat{\mathbf{w}}_{GMV} - \mathrm{E}[\hat{\mathbf{w}}_{GMV}]\right)\left(\hat{\mathbf{w}}_{GMV} - \mathrm{E}[\hat{\mathbf{w}}_{GMV}]\right)^\top\right]$$

It now stands to determine $\alpha$. Shrinkage intensities are most often determined by cross-validation (CV) (see, e.g., James, Witten, Hastie, and Tibshirani (2013, ch. 5)). The aim is to find the best shrinkage coefficients by dividing data into a validation and a training set. These are then used in conjunction with a loss function to determine the optimal value for the shrinkage coefficients. A natural choice of loss function for the GMV portfolio is the out-of-sample variance. The aim is to determine $\min_\alpha \hat{\mathbf{w}}_{SH}^\top(\alpha)\mathbf{\Sigma}\hat{\mathbf{w}}_{SH}(\alpha)$. The out-of-sample loss depends on $\mathbf{\Sigma}$, which is not known. The perhaps most obvious solution is to use the validation set to estimate the $\mathbf{\Sigma}$, which has its own issues. The second solution, employed by Bodnar, Parolya, and Schmid (2018), is to first solve the optimization problem analytically. The estimator is unobtainable, since it depends on $\mathbf{\Sigma}$ so it is usually referred to as an *oracle* estimator. To construct an estimator which can actually be used, the authors construct a *bona-fide* estimator. A bona-fide estimator only depends on esitmated quantities, so it is known. The aim is then to construct the bona-fide estimator such that it converges to the same limit as the oracle. It is a consistent estimator for the limiting object at hand. The two methods can differ quite substantially in their solution. The CV method is simple to implement while the other *should* be theoretically superior. In Chunk 4.2.1 there is some R-code for a small motivating example to why deriving bona-fide estimators can prove to be fruitful. It is a comparison between the estimator from Bodnar, Parolya, and Schmid (2018) and determining the shrinkage coefficient using a 5 fold CV procedure. In this example $n = 250, p = 150$ and $\mathbf{b}$ is equal to the Equally Weighted (EW) portfolio. It is a large portfolio, though $c = p/n < 1$.

Chunk 4.2.1: R-code performing a 5-fold cross-validation for determining shrinkage coefficients as well as the analytic methods from the HDShOP package.

```
# setup
set.seed(123)
p <- 150
n <- 250
K <- 5
# use EW portfolio as target
b <- rep(1,p) / p
# simulate params & dataset
Sigma <- HDShOP::RandCovMtrx(p)
mu <- runif(p, -0.1, 0.1)
# simulate from statistical model
Y <- mu %*% t(rep(1,n)) + t(chol(Sigma)) %*% matrix(rt(n*p, df=5), ncol=n)
# create test splits
folds <- split(1:n, 1:K)
grid <- expand_grid("alpha" = seq(0.01, 0.99, by=0.01), "fold" = 1:K)
# perform 5-fold CV
result <- pmap(grid, ~{
    test <- Y[,folds[[.y]]]
    train <- Y[,-folds[[.y]]]
    S <- var(t(train))
    S_inv <- solve(S)
    w <- .x * S_inv %*% rep(1, p) / sum(S_inv) + (1-.x)*b
    t(w) %*% var(t(test)) %*% w
  }) %>%
  unlist() %>%
  tibble("variance"=.) %>%
  bind_cols(grid) %>%
  group_by(alpha) %>%
  summarise(loss = mean(variance),
            sd_loss = var(variance))
min_vals <- filter(result, loss == min(loss))
# Use HDshop pkg to compute the weights
w_bodnar2018 <- HDShOP::MVShrinkPortfolio(Y, gamma=Inf, b=b, beta=0.01)
```

In Figure 4.1 the out-of-sample variance, aggregated over folds, is displayed. Cross-validation suggest that the optimal value should be 0.68. The analytic method from Bodnar, Parolya, and Schmid (2018) suggests that the optimal value is 0.7968. The "correct" value of $\alpha$, as given by the oracle estimator, is equal to 0.8151. Its limiting value (see Theorem 2.1 Bodnar, Parolya, and Schmid (2018)) is equal to 0.7933. Although none of the methods are spot on, the bona-fide estimator is closer to what the optimal value should be. Furthermore, changing the number of folds in the K-fold CV can give quite different results in the optimal value. If there is a possibility to derive a bona-fide estimator, then it seems that it is better than using a CV approach. This approach is used in paper 3, 4 and 5 as well as the linear shrinkage from (4.2).

Shrinking the weights can provide a good estimator in higher dimensions. As previously stated, another approach is to shrink the elements of $\mathbf{S}$. If the sample covariance matrix is shrunk, then the concentration ratio $c$ can be greater than one. This is the topic of the last paper of this thesis which is left for the next section where the papers are described in more detail.

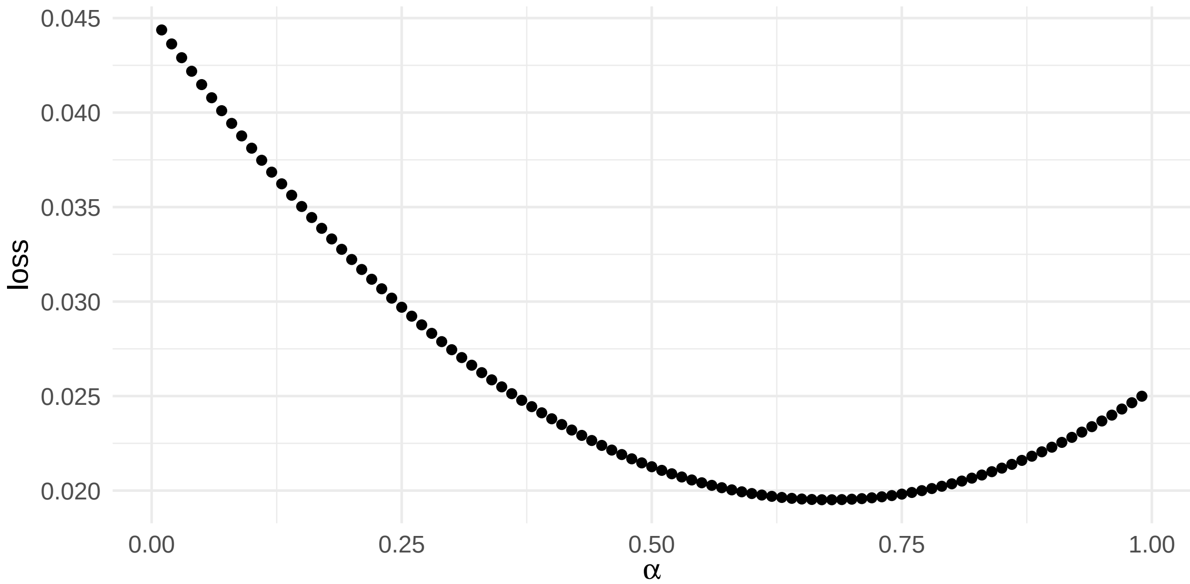


Figure 4.1: Out-of-sample variance estimates from the 5-fold cross-validation.

# 5. Summary of papers

The following papers are included in this thesis.

## 5.1 Paper 1 - Sampling distributions of optimal portfolio weights and characteristics in small and large dimensions

The paper investigates a fundamental question in MPT. What are the actual implications of using the sample covariance matrix $\mathbf{S}$ and the sample mean $\bar{\mathbf{y}}$ instead of the true covariance matrix $\boldsymbol{\Sigma}$ and mean vector $\boldsymbol{\mu}$? The paper presents an answer to the question when the returns follow a multivariate normal distribution. The paper contains the distribution for all optimal portfolios on the common form

$$\mathbf{L}\hat{\mathbf{w}}_{opt} = \mathbf{L}\hat{\mathbf{w}}_{GMV} + g(\hat{R}_{GMV}, \hat{V}_{GMV}, \hat{s})\frac{\mathbf{L}\hat{\mathbf{Q}}\bar{\mathbf{y}}}{\hat{s}} \tag{5.1}$$

for some matrix $\mathbf{L}$ of size $k \times p$ where $k < p$. The joint distribution of all quantities in (5.1) is derived through a stochastic representation. The stochastic representation can be used to efficiently simulate from the distribution. By simulation, quantities such as quantiles or other summary statistics are easily computed. Furthermore, the high-dimensional asymptotic joint distribution is also derived. In a simulation study, the high-dimensional asymptotic distribution is compared to simulated data. One scenario considers simulations from the stochastic representation, trying to deduce the finite-sample properties of the high dimensional distribution. The other scenarios try to investigate what happens when observations deviate from the model. As expected, the high-dimensional distribution works well under the assumptions and seems to be reasonably robust from deviations of the model. The GMV and self-financing portfolio are the most robust quantities to deviations from the model.

## 5.2 Paper 2 - Tangency portfolio weights under a skew-normal model in small and large dimensions

In this paper, the implications of skewness on the Tangency Portfolio (TP) from Chapter 2 is investigated. The portfolio is obtained from the quadratic utility function, namely

$$\min_{\mathbf{w}} \mathbf{w}^\top(\boldsymbol{\mu} - r_f\mathbf{1}) - \frac{\gamma}{2}\mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w}. \tag{5.2}$$

This paper extends paper 1 as it considers investments in a risk-free asset and use an extension of the multivariate normal model, the CSN model presented in Chapter 3. The model can include skewness in the asset returns, a trait returns usually exhibit (see e.g., Cont (2001)). Similarly to paper 1, the distribution of the sample tangency portfolio is derived and what implications the skewness has on the portfolio. In short, skewness results in a bias present in the portfolio weights. An investor will not hold the correct portfolio on average. Furthermore, the high-dimensional distribution of the sample tangency portfolio is derived. It can be seen that the skewness disappears asymptotically. The high-dimensional distribution is the same as previous research has shown (see, e.g., Karlsson, Mazur, and Muhinyuza (2021)).

## 5.3 Paper 3 - Dynamic shrinkage estimation of the high-dimensional minimum-variance portfolio

This paper solves a practical feature when investing in the GMV portfolio: how to rebalance the portfolio at fixed time points. If an investor owns a GMV portfolio and waits for a week, month or year the data will likely indicate that another GMV portfolio should be held. The change can be quite large if $n$ is sufficiently small. A natural question to ask is how to go from one portfolio to another, e.g., how to rebalance optimally when a new set of data is available.

In this paper, a dynamic rebalancing scheme for the GMV portfolio is derived. The scheme aims to decrease the out-of-sample variance between the holding portfolio, which might be a random GMV portfolio, and the GMV portfolio that is suggested to hold given this current period's data. The portfolios are on the following form

$$\hat{\mathbf{w}}_{SH;n_i} = \psi_i \hat{\mathbf{w}}_{S;n_i} + (1 - \psi_i)\hat{\mathbf{w}}_{SH;n_{i-1}}, \tag{5.3}$$

where $\hat{\mathbf{w}}_{S;n_i}$, $i = 1, 2, ..., T$, is the traditional sample GMV portfolio using the $i$th sample of size $n_i$ to estimate the GMV portfolio weights in (2.3). The initial portfolio, $\hat{\mathbf{w}}_{SH;0}$, can be a random GMV portfolio or a deterministic target portfolio $\mathbf{b}$. It is assumed that an investor have specified dates $t_i$, $i = 1, 2, ..., T$ that he/she wants to rebalance his/hers GMV portfolio. The shrinkage coefficients are then determined through the following optimization problem

$$\min_{\psi_i} \hat{\mathbf{w}}_{SH;n_i}^\top \mathbf{\Sigma} \hat{\mathbf{w}}_{SH;n_i}.$$

The problem is similar to the linear shrinkage discussed in Chapter 4. The portfolio allocation problem is an extension to the work of Bodnar, Parolya, and Schmid (2018) and use the flexible location and scale model in (3.5).

The portfolio is shown to produce great results in an extensive simulation study. It also provides a better estimator for the volatility in comparison to the traditional sample GMV as well as the GMV portfolio using Ledoit and Wolf (2020) nonlinear shrinkage estimator for the sample covariance matrix. There are also many other benefits of using the portfolio strategy. Transitions from one portfolio to the next costs money, which will diminish the return and profit that can be made. Furthermore, it is not always possible to go from one portfolio to the next in a day or even a month. The traditional GMV portfolio might suggest that an institution should first own a large long position and the next month a large short position. Depending on the size of the institution and the portfolio, the size of these positional changes might be illegal. It can be deemed market influencing or just outright impossible to sell that many assets. The rebalancing scheme provides a solution to these problems.

## 5.4 Paper 4 - Is the empirical out-of-sample variance an informative risk measure for high-dimensional portfolios?

Any empirical application using the GMV portfolio is bound to include the volatility or variance as a performance measure. However, is the empirical out-of-sample variance a consistent estimator of the variance? Furthermore, is it a good option to use or are there perhaps better options to use as performance measures? This paper investigates two different metrics of evaluation that are common to the GMV portfolio, the out-of-sample variance and the relative out-of-sample loss.

This paper also considers the location and scale model from (3.5) and three different GMV portfolios. The first portfolio is the traditional GMV portfolio from (2.3), the second is the portfolio from Bodnar, Parolya, and Schmid (2018) and the last is the linear shrinkage portfolio from Frahm and Memmel (2010). The properties of the out-of-sample variance, relative loss and their empirical counterparts are derived for the three different portfolios. It is done under different assumptions on the parameters of the model. Most notably, there is a natural ordering to the different out-of-sample losses. The empirical out-of-sample loss is smallest for Bodnar, Parolya, and Schmid (2018), second to smallest is Frahm and Memmel (2010) and the largest is the traditional sample GMV portfolio. Furthermore, the empirical out-of-sample variance for the different portfolios are presented. The assumptions that are necessary for convergence of the empirical out-of-sample variance are quite different from those used for the empirical out-of-sample loss. The assumptions necessary for convergence of the empirical out-of-sample variance are deemed stronger than the empirical out-of-sample loss. The theoretical findings are investigated in a simulation study and an empirical application which validate the ordering even when the model assumptions are violated.

## 5.5 Paper 5 - Two is better than one: Regularized shrinkage of large minimum variance portfolios.

The methods of this thesis most often use the sample covariance matrix $\mathbf{S}$. They deal with the fact that the sample covariance matrix is a noisy estimator by linear shrinkage or understanding the uncertainty it provides. This estimator can only cover $c < 1$. This paper introduces a Thikonov regularisation on the portfolio weights together with the linear shrinkage from prior papers. It results in a ridge type estimator for the sample covariance matrix together with the linear shrinkage on the portfolio weights. It enables the method to cover the case where $c > 1$ and potentially create a more robust estimator of the portfolio weights. The portfolio looks like

$$\hat{\mathbf{w}}_{SH} = \psi \frac{\left(\mathbf{S} + \eta \mathbf{I}_p\right)^{-1} \mathbf{1}_p}{\mathbf{1}_p^\top \left(\mathbf{S} + \eta \mathbf{I}_p\right)^{-1} \mathbf{1}_p} + (1-\psi)\mathbf{b}.$$

The natural loss for estimating $\psi$ and $\eta$ is the out-of-sample variance. It turns out that there is no possibility to construct a closed-form oracle estimator for the shrinkage parameters. However, there is a possibility to construct an oracle loss function, for which a bona-fide estimator can be derived. The bona-fide loss is proved to be a consistent estimator of the oracle loss.

The model is seen to perform on-par with the nonlinear shrinkage of Ledoit and Wolf (2020) in all simulations. The method also consistently beats the Ledoit and Wolf (2020) method in an extensive empirical analysis. It provides the best out-of-sample variance for five out six different configurations. The only competitor is the equally weighted portfolio. Furthermore, the method also increases/decreases the results of several other portfolio metrics.

## 5.6 Other research results

Paper 3 is accompanied by the R-package Bodnar, Parolya, and Thorsén (2021d), available on CRAN. The readers of this thesis are free, or rather encouraged(!), to install it with `install.packages("DOSPortfolio")`. The package provides a simple interface for the methods implemented in the paper. In Chunk 5.6.1 a short example on how to construct the dynamic portfolio weights using the package is displayed. The package is the first iteration of possibly many more portfolios which can be constructed in a similar fashion.

Chunk 5.6.1: R-code which showcase the use of the DOSPortfolio package.

```
library(DOSPortfolio)
df <- read_csv("../data/returns.csv")
p <- 350; n <- 400
# Sample p assets
set.seed(1234)
asset_cols <- sample(2:ncol(df), size = p)
# specify reallocation points, daily in this case
reallocation_points <- seq(n, nrow(df), by=n)
# estimate portfolio weights
dos_weights <- df %>%
  select(all_of(asset_cols), -date) %>%
  DOSPortfolio(reallocation_points = reallocation_points,
               target_portfolio = rep(1, ncol(.))/ncol(.),
               shrinkage_type = "overlapping")
```

Furthermore, the following papers were co-authored throughout the writing of this thesis Bodnar, Lindholm, Niklasson, and Thorsén, 2022b, Bodnar, Niklasson, and Thorsén, 2021b and Bodnar, Lindholm, Thorsén, and Tyrcha, 2021b. The first presents an analytic derivation of the MPT framework in the Bayesian setting. It specifically looks at how quantiles of optimal portfolios can be constructed and the effects of estimation uncertainty in these. This is especially important since the regulations in place demands that quantile-based risk measures are reported (see The Basel Committee on Banking Supervision (2019)). The second paper provides a continuation on the first. The idea is to model the belief of an investor in a prior distribution. The method then aims to construct the prior to capture what the likelihood cannot. It imposes a prior distribution which adapts to the recent observations when the market is turbulent. The algorithm is seen to work well when markets are turbulent. The third paper also considers quantile-based portfolios. The paper does so in a general framework, not necessarily the same framework as MPT where the first two moments of the return distribution are used.

# 6. Future research

There are many possible extensions and future projects to the thesis at hand. Estimation uncertainty is the primary motivation for this thesis. Bayesian statistics provide a straightforward way to integrate that. However, it demands indepth knowledge of Markov Chain Monte Carlo and also how to construct good prior distributions. Neither are easy tasks. Another approach of incorporating estimation uncertainty is robust optimization. Robust optimization, in a MPT setting, tries to incorporate the estimation uncertainty into the portfolio allocation problem itself. Are there connections to be made and especially with empirical bayes?

Paper 3 assumes that the rebalancing points are fixed. This assumption can be limiting for some investors. Can those be exchanged for stopping times, incorporated in the decision process and the portfolio allocation problem?

Many Multivariate GARCH models can be formulated as the following BEKK model (see, e.g., Engle and Kroner (1995))

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}^\top + \sum_{k=1}^{K} \boldsymbol{A}_k \boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}^\top \boldsymbol{A}_k^\top + \sum_{k=1}^{K} \mathbf{G}_k \mathbf{H}_{t-1} \mathbf{G}_k^\top, \tag{6.1}$$

where $\mathbf{H}_i$ is a sequence of conditional covariance matrices, $\boldsymbol{\epsilon}_t | \mathcal{F}_{t-1} \sim N_p(\mathbf{0}], \mathbf{H_t})$ and the matrices $\mathbf{C}, \boldsymbol{A}_i$ and $\mathbf{G}_i$ are of appropriate dimensions. These are usually very hard to fit and use for portfolio allocations. The first issue is due to the number of parameters in the model. There are a number of parametrisations but if all matrices are symmetric then there are $(K + 1/2)p(p+1)$ parameters to estimate. Building a portfolio of size 10 with $K = 1$ implies that 165 parameters need to be estimated. Furthermore, although the constraints should enforce forecasts which are positive definite it is not necessarily true that they will be numerically invertible. It can provide forecasts which are very close to singular. The first issue can be solved if one can formulate the models as Recurrent Neural Networks and use deep-learning libraries Torch or Tensorflow to fit the models. These are tailored to solve the problem of fitting very large models. Recent large Natural Language Processing models have *billions* of parameters (see, e.g., Brown et al. (2020)). The second problem can then possibly be solved by placing BEKK models in this framework. Positive definite forecasts could then be enforced by developing new layers to the network. Furthermore, it would also be easier to integrate different sources of information in the models in the context of these models.

# 7. Svensk sammanfattning

Portföljer med finansiella instrument och diversifiering går oftast hand i hand. Diversifiering är ett, om inte det bästa, verktyget för att minska risken i en portfölj. Om en av investeringarna skulle gå dåligt spelar det ingen roll för den utgör en sådan liten del av portföljen. Modern Portföljteori (MPT) är ett ramverk för att konstruera diversifierade portföljer. Ett hinder är dock att MPT använder okända parametrar. Dessa okända parametrar är de två första momentet av tillgångslagens avkastningsfördelning. När dessa ersätts av skattare introduceras estimeringsosäkerhet. Om en investerare inte förstår osäkerheten finns det en risk att den dominerar alla resultat som strategin skulle kunna påvisa. Det finns inget sätt att avgöra om strategin fungerar eller inte.

Denna avhandling innehåller fem artiklar. De innehåller resultat som kan hjälpa en investerare att hantera skattningsosäkerhet i oändliga portföljer samt, i vissa fall, beskriva portföljernas fördelning för ett ändligt stickprov. Dessa resultat gör det lättare för investerare att förstå investeringsprocessen och de empiriska resultat som kan observeras i praktiken.

Artikel 1 utforskar alla portföljer i MPT-ramverket. I artikeln härleds alla optimala portföljers empiriska fördelning, det vill säga när skattare används istället för de sanna parametrarna. Det inkluderar fördelningen för vikterna samt den avkastning, varians och andra mått som karateriserar dessa portföljer. Därefter härleds den asymptotiska fördelningen för alla mått samt vikter i stora dimensioner. Denna fördelning kan ses som ett fall då en investerare diversifierar oändligt mycket i MPT-ramverket samt har oändligt mycket data på de tillgångslag han/hon investerar i. En simuleringsstudie visar att den asymptotiska fördelningen utgör en god approximation av den fördelningen för ändliga stickprov, givet att modelantagandet håller.

Artikel 2 fortsätter vidare på artikel 1. I denna artikel undersöks portföljen från det kvadratiskt nytto-optimeringsproblem med en riskfri tillgång. En riskfri tillgång kan exempelvis vara ett räntebärande konto. Denna portfölj är också känd som den tangerande portföljen. I artikeln härleds den empiriska fördelningen för portföljvikterna fast med en generalisering av modellen från artikel 1. Fördelningen för tillgångslagen antas vara skev-normal. Resultaten visar att skevhet gör att de empiriska vikterna i portföljen är i genomsnitt fel. Detta sker i genomsnitt för ändliga stickprov men den empiriska portföljen skattar rätt objekt i stora dimensioner.

Artikel 3 ger en lösning på problemet att äga en portfölj för att sedan gå över till en annan. I artikeln utvecklas en metod som omfördelar portföljen mellan fixa tidssteg. Den gör detta genom att minimera den framtida variansen för Minsta Varians (GMV) portföljen, givet att investeraren äger en portfölj. Portföljen som ägs i denna stund kan vara deterministisk eller en skattad GMV-portfölj. En stor simuleringsstudie visar att denna metod kan uppnå goda resultat i termer av att skatta den relativa förlusten. Den relativa förlusten är en enkel utveckling av portföljvariansen. Metoden utvärderas därefter med hjälp av marknadsdata. Den lyckas bäst i att ge minsta framtida varians samt ge de minsta förändringarna i portföljvikterna bland de olika metoderna i jämförelsegruppen. Metoden har implementerats i ett R-paket som heter DOSPortfolio och finns tillgängligt på CRAN.

Artikel 4 härleder olika egenskaper för två olika prestandamått av tre olika skattare för GMV-portföljer. Prestandamåtten är den framtida variansen samt den framtida relativa förlusten för GMV-portföljerna. Det tidigare nämnda prestandamåttet används närpå alltid för att utvärdera GMV-portföljer med empirisk data. Resultaten visar att den relativa förlusten inte behöver lika strikta antaganden för att konvergera i stora dimensioner. Detta mått kan därför täcka flera modeller med detta mått till skillnad från variansen. Prestandamåtten används sedan för att bestämma en ordning på de olika portföljerna. Dessa resultat verifieras i en simuleringsstudie samt på empirisk data.

Artikel 5 utvidgar en av GMV-portföljerna från artikel 3 och 4. I den här artikeln introduseras en Thikonov-regularisering på portföljvikterna vilket resulterar i en Ridge-liknande skattare för den empiriska kovariansmatrisen. Detta kombineras sedan med den linjära shrinkage-metoden från artikel 3 och 4. Denna portfölj undersöks sedan i en omfattande simuleringsstudie och dess prestanda studeras på empirisk data. Simuleringsstudien visar att metoden är jämförbar med ett flertal metoder. Den empiriska studien visar att metoden som utvecklas ger lägre skattningar av framtida varians än ett flertal referensportföljer samt att den visar bra prestanda gällande portföljvikterna.

# Bibliography


Alfelt, Gustav (2021). “Modeling the covariance matrix of financial asset returns”. PhD thesis. Department of Mathematics, Stockholm University.

Bai, Zhidong and Jack W. Silverstein (2010). *Spectral analysis of large dimensional random matrices*. Springer.

Barndorff-Nielsen, Ole E. and Neil Shephard (2004). “Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics”. In: *Econometrica* 72.3, pp. 885–925.

Best, Michael J. and Robert R. Grauer (1991). “On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results”. In: *The Review of Financial Studies* 4.2, pp. 315–342.

Bodnar, Taras, Holger Dette, Nestor Parolya, and Erik Thorsén (2022b). “Sampling distributions of optimal portfolio weights and characteristics in small and large dimensions”. In: *Random Matrices: Theory and Applications*.

Bodnar, Taras and Arjun Gupta (2009). “Construction and Inferences of the efficient frontier in elliptical models”. In: *Journal of the Japan Statistical Society* 39.2, pp. 193–207.

Bodnar, Taras, Arjun K. Gupta, and Nestor Parolya (2016). “Direct shrinkage estimation of large dimensional precision matrix”. In: *Journal of Multivariate Analysis* 146, pp. 223–236.

Bodnar, Taras, Mathias Lindholm, Vilhelm Niklasson, and Erik Thorsén (2022b). “Bayesian quantile-based portfolio selection”. In: *Under revision in Applied Mathematics and Computation*.

Bodnar, Taras, Mathias Lindholm, Erik Thorsén, and Joanna Tyrcha (2021b). “Quantile-based optimal portfolio selection”. In: *Computational Management Science* 18.3, pp. 1–26.

Bodnar, Taras, Vilhelm Niklasson, and Erik Thorsén (2021b). “Volatility sensitive Bayesian estimation of portfolio VaR and CVaR”. In: *Research report in Mathematical Statistics, 2021:6*.

Bodnar, Taras and Yarema Okhrin (2008). “Properties of the singular, inverse and generalized inverse partitioned Wishart distributions”. In: *Journal of Multivariate Analysis* 99, pp. 2389–2405.

Bodnar, Taras, Yarema Okhrin, and Nestor Parolya (2016). “Optimal shrinkage-based portfolio selection in high dimensions”. In: *arXiv preprint arXiv:1611.01958*.

Bodnar, Taras, Yarema Okhrin, and Nestor Parolya (2021). “Optimal shrinkage-based portfolio selection in high dimensions”. In: *Journal of Business & Economic Statistics*, pp. 1–17.

Bodnar, Taras, Nestor Parolya, and Wolfgang Schmid (2013). “On the equivalence of quadratic optimization problems commonly used in portfolio theory”. In: *European Journal of Operational Research* 229.3, pp. 637–644.

Bodnar, Taras, Nestor Parolya, and Wolfgang Schmid (2018). "Estimation of the global minimum variance portfolio in high dimensions". In: *European Journal of Operational Research* 266.1, pp. 371–390.

Bodnar, Taras, Nestor Parolya, and Erik Thorsén (2021d). *DOSPortfolio: dynamic optimal shrinkage of high-dimensional portfolios*. R package version 0.1. URL: `https://CRAN.R-project.org/package=DOSPortfolio`.

Brown, Tom B. et al. (2020). *Language models are few-shot learners*. arXiv: `2005.14165 [cs.CL]`.

Cont, Rama (2001). "Empirical properties of asset returns: stylized facts and statistical issues". In: *Quantitative Finance* 1.2, pp. 223–236.

Ding, Yi, Yingying Li, and Xinghua Zheng (2021). "High dimensional minimum variance portfolio estimation under statistical factor models". In: *Journal of Econometrics* 222.1, pp. 502–515.

Dominguez-Molina, J. Armando, Graciela Gonzaez-Farias, Rogelio Ramos-Quiroga, and Arjun K. Gupta (2007). "A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis". In: *Communications in Statistics—Theory and Methods* 36.9, pp. 1691–1703.

El Karoui, Noureddine (2010). "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation". In: *The Annals of Statistics* 38.6, pp. 3487–3566.

Engle, Robert F. and Kenneth F. Kroner (1995). "Multivariate simultaneous generalized ARCH". In: *Econometric Theory* 11.1, pp. 122–150.

Frahm, Gabriel and Christoph Memmel (2010). "Dominating estimators for minimum-variance portfolios". In: *Journal of Econometrics* 159.2, pp. 289–302.

Frankfurter, George M., Herbert E. Phillips, and John P. Seagle (1971). "Portfolio selection: the effects of uncertain means, variances, and covariances". In: *Journal of Financial and Quantitative Analysis* 6.5, pp. 1251–1262.

Golosnoy, Vasyl, Bastian Gribisch, and Miriam Isabel Seifert (2019). "Exponential smoothing of realized portfolio weights". In: *Journal of Empirical Finance* 53, pp. 222–237.

Gupta, Arjun K. and Daya K. Nagar (2000). *Matrix variate distributions*. Chapman and Hall/CRC.

Harville, David A. (1997). *Matrix algebra from a statistician's perspective*. Vol. 1. Springer.

Higham, Nicholas J. (2002). *Accuracy and stability of numerical algorithms*. SIAM.

Hult, Henrik, Filip Lindskog, Ola Hammarlid, and Carl Johan Rehn (2012). *Risk and portfolio analysis: principles and methods*. Springer Science & Business Media.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2013). *An introduction to statistical learning*. Vol. 112. Springer.

Karlsson, Sune, Stepan Mazur, and Stanislas Muhinyuza (2021). "Statistical inference for the tangency portfolio in high dimension". In: *Statistics* 55.3, pp. 532–560.

Kempf, Alexander and Christoph Memmel (2006). "Estimating the global minimum variance portfolio". In: *Schmalenbach Business Review* 58.4, pp. 332–348.

Ledoit, Olivier and Michael Wolf (2017). "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks". In: *The Review of Financial Studies* 30.12, pp. 4349–4388.

Ledoit, Olivier and Michael Wolf (2020). "Analytical nonlinear shrinkage of large-dimensional covariance matrices". In: *The Annals of Statistics* 48.5, pp. 3043–3065.

Li, Jian, P. Stoica, and Zhisong Wang (2004). "Doubly constrained robust capon beamformer". In: *IEEE Transactions on Signal Processing* 52.9, pp. 2407–2423.

Livan, Giacomo, Marcel Novaes, and Pierpaolo Vivo (2018). *Introduction to random matrices: theory and practice*. Springer.

Maillet, Bertrand, Sessi Tokpavi, and Benoit Vaucher (2015). “Global minimum variance portfolio optimisation under some model risk: a robust regression-based approach”. In: *European Journal of Operational Research* 244.1, pp. 289–299.

Marchenko, Vladimir Alexandrovich and Leonid Andreevich Pastur (1967). “Distribution of eigenvalues for some sets of random matrices”. In: *Matematicheskii Sbornik* 114.4, pp. 507–536.

Markowitz, Harry (1959). *Portfolio selection: efficient diversification of investments.* John Wiley & Sons, Inc and Chapman & Hall, Ltd.

Mehta, Madan Lal (2004). *Random matrices*. Elsevier.

Merton, Robert C. (1972). “An analytic derivation of the efficient portfolio frontier”. In: *The Journal of Financial and Quantitive Analysis* 7.4, pp. 1851–1872.

Merton, Robert C. (1980). “On estimating the expected return on the market: An exploratory investigation”. In: *Journal of Financial Economics* 8.4, pp. 323–361.

Okhrin, Yarema and Wolfgang Schmid (2006). “Distributional properties of portfolio weights”. In: *Journal of Econometrics* 134.1, pp. 235–256.

Ross, Stephen A. (1976). “The arbitrage theory of capital asset pricing”. In: *Journal of Economic Theory*, pp. 341–360.

Rubio, Francisco and Xavier Mestre (2011). “Spectral convergence for a general class of random matrices”. In: *Statistics & Probability Letters* 81.5, pp. 592–602.

Stein, Charles (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, Volume 1: Contributions to the Theory of Statistics.* Vol. 1. 1. University of California Press, pp. 197–206.

The Basel Committee on Banking Supervision (2019). *Minimum capital requirements for market risk.* Online: accessed 12-April-2021. URL: `https://www.bis.org/bcbs/publ/d457.pdf`.

Thorsén, Erik (2019). “Assessment of the uncertainty in small and large dimensional portfolio allocation”. PhD thesis. Department of Mathematics, Stockholm University.

Tobin, James (1958). “Liquidity preference as behavior towards risk”. In: *The Review of Economic Studies* 25.2, pp. 65–86.

Tsay, Ruey S. (2010). *Analysis of financial time series.* John Wiley & Sons.

Wasserman, Larry (2004). *All of statistics: a concise course in statistical inference.* Vol. 26. Springer.

Wigner, Eugene P. (1967). “Random matrices in physics”. In: *SIAM review* 9.1, pp. 1–23.

# Part II

# Papers