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

Erik Thorsén

January 27, 2022

Acknowledgements

First and foremost i would like to thank ...

List of papers

This thesis is based on the following papers

Papers included in this thesis

I **Sampling Distribution of Optimal Portfolios in Small and Large Dimensions** Taras Bodnar et al. (2021b). In: *Random Matrices: Theory and Applications*. DOI: `10.1142/S2010326322500083`

II **Tangency portfolio weights under a skew-normal model in small and large dimensions** Farrukh Javed, Stepan Mazur, and Erik Thorsén (2021). In: *Submitted for publication*

III **Dynamic Shrinkage Estimation of the High-Dimensional Minimum-Variance Portfolio** Taras Bodnar, Nestor Parolya, and Erik Thorsén (2021a). In: *arXiv preprint arXiv:2106.02131*

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

V **Double shrinkage of the high-dimensional Global Minimum-Variance portfolio.** Ref InsertRef and Insert Ref (2020). In: *Insert Ref* Insert Ref.5. DOI: `REF`

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

- The **DOSPortfolio** R-package which is published on CRAN. It is available here.

- **Bayesian Quantile-Based Portfolio Selection** Taras Bodnar et al. (2020). In: *Accepted with revision to Applied Mathematics and Computation*

- **Volatility Sensitive Bayesian Estimation of Portfolio VaR and CVaR** Taras Bodnar, Vilhelm Niklasson, and Erik Thorsén (2021). In: *Submitted for publication*

- **Quantile-based optimal portfolio selection** Taras Bodnar et al. (2021a). In: *Computational Management Science*

Author contribution: Thorsén E. contributed...

Contents

Chapter 1

Introduction – Decisions, uncertainty and asset allocations

Splitting a budget between your possible or existing investments is always a decision that needs to be made. How should you spend your cash? This thesis is about how to make these decisions. It is about how to make them when you have many possible choices and do not know how their payoffs will behave in the future. That turns out to be very complicated simply because we have many choices and we do not know their future payoffs and its behaviour. As an example, we start with a single investment case. Your imaginary house in Sweden has increased drastically in value[1]. How will it be valued in the future? Will it continue to increase or will there be a decrease/crash in prices? If there is an increase or a decrease, how much will it be? There are so many unknowns, which is exactly the issue. We need to describe how an asset, in this case the house, will behave in the future. The future is uncertain and we seldom know exactly what will happen. We look back and let history give insight to how the future will behave. Depending on how good you are as a historian, modeler or gambler you can usually get something out of history and the data it brings. It will never be perfect but lets assume that you are right 51% of the time.

Lets fast forward in life. Given your great guesses you have been lucky enough to now own a house, a country house and a allotment. You cant be as good of a guesser on all of those assets. Your budget needs to cover all of them. How should you spend your cash on these assets? You can invest in any of the three different assets and through that investment they might be worth more in the future. Once more, you do not know that but can perhaps make an educated guess. Given your educated guesses, should you spend all your money on one single asset? You were right 51% on the guesses regarding your house. Should you continue to place all your bets on the same outcome? If you are uncertain in your guess then your investment might be fairly risky. You just dont know it yet. There are two options, spend everything on one asset or spend a little on everything. The first option can give you the potential of a large return but also a lot of risk. The second option can potentially give you less risk. In this thesis we consider the second option and how you can do it when you dont know the future.

In reality making a good informed guess is hard. Its very seldom that you are right 51% of the time. Markets are very efficient. **Why is this important to finance and what problems can it be used to solve.**

This thesis is structured as follows. In Chapter 2 we present the framework we will work within, Markowitz (1959) Modern Portfolio Theory (MPT). Originally MPT use information which are not available. It relies on parameters which we do not know. The subsequent chapter, Chapter 3 tries to display what happens when we replace these parameters with estimates. It does so under some specific models that we consider in this thesis. Chapter 4 presents what happens in the MPT framework when we try to diversify as much as possible. We work with infinite dimensional portfolios. The last two chapters, Chapter 5 and 6 are summaries 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, here.

---

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

Chapter 2

Modern portfolio theory

Modern Portfolio Theory (MPT) was introduced by Markowitz, 1959. In his seminal work he argued that any portfolio which simply maximizes its profit will result in a naive solution. Very much like the example in Chapter 1. The future is unknown and (one of) the best model(s) we have for it is stochastic. Investing all your capital in the asset with the highest return is not sensible if you do not know the future. Such an investment will cause you to take an extreme amount of risk. Markowitz argued that any well diversified portfolio should be preferred to any non diversified portfolio. Such a portfolio can be obtained through many different procedures. He proposed the use of the first two moments for the allocation problem. If an asset has high return on average, it might make sense to invest a lot in it although not at the cost of large amounts of risk. If an asset is not risky, then it makes sense to invest in it.

We assume that the p-dimensional vector of asset returns $\mathbf{x}$ is random with mean vector $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ and covariance matrix $\mathrm{Var}(\mathbf{x}) = \boldsymbol{\Sigma}$. The matrix $\boldsymbol{\Sigma}$ is of dimensions $(p \times p)$. Although there are usually little restriction on $\boldsymbol{\mu}$ there are usually very specific restrictions on the covariance matrix $\boldsymbol{\Sigma}$. Since the covariance matrix is a subject of its own we dedicate the next section to it and disregard these restrictions for now. We will merely say that it is well behaved. The restrictions on the mean will be commented on below. Using the two moments for the asset returns the portfolio distribution $x = \mathbf{w}^\top \mathbf{x}$ has mean $\mathrm{E}(x) = \mathbf{w}^\top \boldsymbol{\mu}$ and variance $\mathrm{Var}(x) = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$. Let μ_0 be the target return that the investor would like to achieve from his/hers portfolio and $\mathbf{1}$ column vector of ones with appropriate dimensions. Markowitz (1959) considered the following optimization problem

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

This problem is a quadratic optimization problem with linear equality and inequality constraints. The objective is to minimize the variance of the portfolio. The constraint $\mathbf{w}^\top \mathbf{1} = 1$ essentially 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 et al. (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 investor expectations on the portfolio. As μ_0 grows, the return of the portfolio will grow. However, that has implications for the objective. Increasing μ_0 will change the amount of variance the portfolio can achieve. Depending on the value μ_0 we would be accepting more risk, there is a risk-return trade-off. The last constraint is rather simple though it can have quite large implications. It states that the weights can not be negative which means that we can only invest money we have. A negative value of w_i in the ith asset is called a short position. You borrow the asset from someone who owns it and then sell it, 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, we exclude it 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 what we refer 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 Gupta (2009), Bodnar, Parolya, and Schmid (2013), and Bauder et al. (2020)). We will continue in the same manner. Let $\boldsymbol{\Sigma}^{-1}$ denote the inverse matrix of $\boldsymbol{\Sigma}$, e.g. $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} = \mathbf{I}$, and

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

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

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

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 we should allocate in the self-financing portfolio. If we set μ_0 equal to $\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}/\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}$ then the portfolio is equal to the GMV portfolio. Excluding the constraint $\mathbf{w}^\top \boldsymbol{\mu} \geq \mu_0$ all together results in the same solution.

The moments of the MV portfolio are equal to

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

From equation (2.5) we can see that all values μ_0 are rescaled according to the moments of the GMV portfolio. If the taget return μ_0 is is smaller than R_{GMV}, then you are better off with the GMV portfolio in terms of return and risk. However, if you choose a value $\mu_0 > R_{GMV}$, then the portfolio return is larger than R_{GMV}. From the two expressions in (2.5) we can see that as μ_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 on the left hand side, we show its behavior for two different portfolio sizes. 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 μ_0. The points in the two plots depicts the expected returns and the standard deviations of a single-stock portfolio, i.e., when everything is invested in a specific stock. You can obtain these by investing in everything you got in a specific stock. However, diversification is always better in terms of decreasing risk. No single-stock portfolio can outperform any of the portfolios on the efficient frontier. That can not happen. The efficient frontier is the best we can do with the stocks at hand, given the objective.

The right hand-side plot of Figure 2.1, displays an extension to the mean-variance problem. It displays what happens when we include a risk-free asset in the portfolio allocation problem. To introduce this option into our allocation problem we add the risk-free rate as part of the portfolio $w_0 r_f + \mathbf{w}^\top \mathbf{x}$ and optimize over w_0 as well. We assume that the risk-free rate is deterministic and therefore (2.2) becomes

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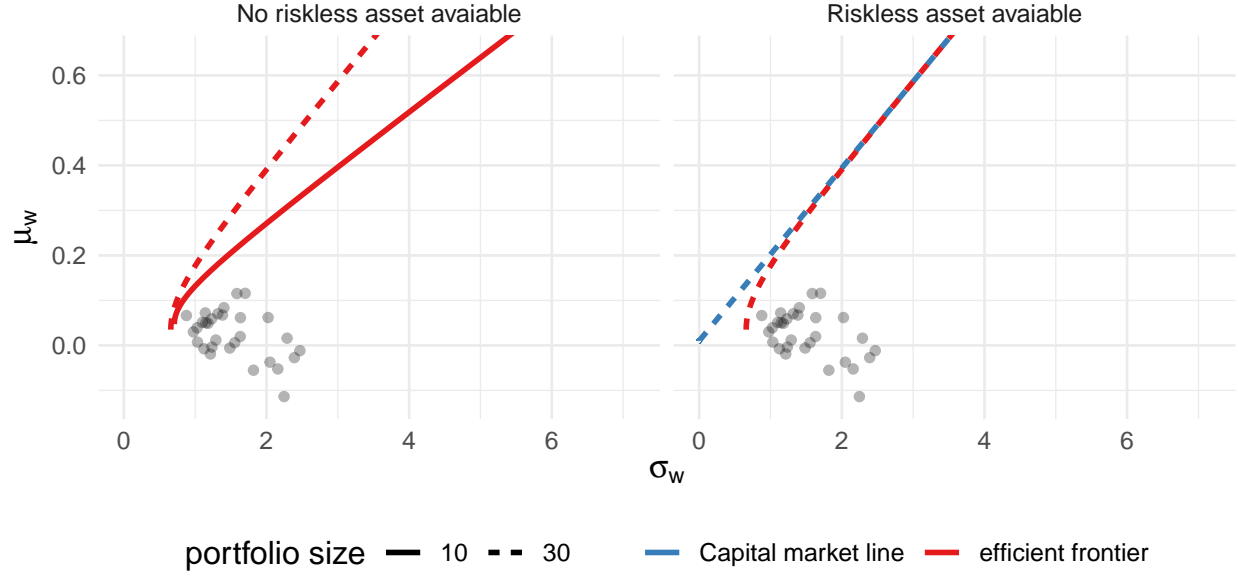\end{aligned} \tag{2.6}$$

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

However, since $w_0 + \mathbf{w}^\top \mathbf{1} = 1$ we substitute $w_0 = 1 - \mathbf{w}^\top \mathbf{1}$ and solve the unconstrained optimization problem instead. 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 we can increase the return and decrease the risk of our position in the market. This is most easily explained by the efficient frontier, displayed in Figure 2.1. For a given level of return we can sometimes get the same or less risk! The same solution can be obtained from the optimization problem with the quadratic utility, defined as

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

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 γ is a fixed constant. This is quite common in MPT, 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 equal to or less than zero. Assuming that is true for our hypothetical investor, (2.7) reduces to

---

[1]See this visualization from Riksbanken

$\mathbf{w}_{TP} = \tilde{\mu}_0 \mathbf{\Sigma}^{-1} \boldsymbol{\mu} / \boldsymbol{\mu}^\top \mathbf{\Sigma}^{-1} \boldsymbol{\mu}$. The term $\mathbf{\Sigma}^{-1} \boldsymbol{\mu}$ is also present in (2.4), although hidden with the introduction of the matrix $\mathbf{Q}$. With a little work, one can rewrite (2.4) as

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

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). To study all the portfolios on the efficient frontier we only need to study these two portfolios. The second is that the true tangency portfolio is given by (2.7) with

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

which is where the efficient frontier and the capital market line meet. Any tangency portfolio with $\mu_0 \leq R_{GMV} + \frac{\mu_0 - R_{GMV}}{V_{GMV}} s$ will be "more efficient" than the efficient frontier if there is a risk free rate. However, its not always the case that we want to optimize the amount of cash held in the risk-free asset. Given that cash is free, we should most likely borrow as much as possible to invest in the market.

All of MPT use the inverse covariance matrix. In the next section we devote some attention to the assumption we make on the covariance matrix.

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. In this section we discuss the restrictions we place on the covariance matrix and what the precision matrix actually represent.

For a vector $\mathbf{x}$ with finite second moment, the covariance matrix is defined as $\mathbf{\Sigma} = \mathrm{E}((\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top)$. It contains the variances of each individual element of $\mathbf{x}$ 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 $\mathrm{E}((x_i - \mu_i)^2)$. 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 we involve covariances on the off-diagonal and we are therefore forced to work with a broader definition. Since we work with real matrices in this thesis, we limit the definition accordingly. From Harville (1997, ch 14.2) we say that 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 $\mathbf{z} \in \mathbb{R}^p$. We will use $\boldsymbol{A} > 0$ or $\boldsymbol{A} \geq 0$ to indicate positive or semi-positive definiteness of the matrix $\boldsymbol{A}$. In the multivariate case we need to assert that a quadratic form is (strictly) positive in comparison to the univariate setting where we can observe it through the variance. Positive- or semi-positive definiteness can be quite cumbersome to work with. We need to assert that the conditions holds for all vectors $\mathbf{z}$. One 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 λ_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, 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 $\mathbf{U} = (\mathbf{u}_1^\top, \mathbf{u}_2^\top, ..., \mathbf{u}_p^\top)^\top$. It might happen that eigenvalues are equal, such as when $\mathbf{\Sigma} = \sigma \mathbf{I}$. Using the relation between eigenvalues and their eigenvectors we can derive the eigenvalue (or spectral) decomposition of a symmetric matrix

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

Since $\mathbf{A}$ is symmetric it also holds that $\mathbf{U}^{-1} = \mathbf{U}^\top$. 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 $\mathbf{y} := \mathbf{U}^\top \mathbf{z} \in \mathbb{R}^p$, then $\mathbf{z}^\top \mathbf{A} \mathbf{z} = \mathbf{z}^\top \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{z} = \mathbf{y}^\top \mathbf{\Lambda} \mathbf{y} = \sum_i^p \lambda_i y_i^2$ 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.

In all papers of this thesis we assume that the true covariance matrix is positive definite. The assumption has quite a deep 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 with a potentially positive return. An opportunity which should not exist 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 $\mathbf{w}^\top \boldsymbol{\mu}$ which can be non-zero unless the elements of $\boldsymbol{\mu}$ are all zero. 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.

The eigenvalue decomposition is very useful. It provides a simple way to construct inverses, which is very important for MPT as seen in (2.4). We claim that $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{-1}$ is a valid inverse which is easy to verify since $\mathbf{B}\mathbf{A} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{-1}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} = \mathbf{I}$. To study the inverse of the covariance matrix we can study the eigenvectors and the inverse of its eigenvalues.

Chapter 3

Statistical models and inference

To *pratically* use the portfolios described by (2.4) we have to specify the two parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This is not really feasible if we have many assets. We might have an opinion of what they should be, however we do not know them precisely. Furthermore, even if one has an informed opinion of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the potential loss of using those exact parameters might be paramount. We usually want to rely on data to estimate the parameters of interest. In this thesis we never use the asset prices themselves but a transformation of the relative differences, that is, their simple and log-returns. Let $p_{i,t}$ be the asset price of the ith asset at time t. 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 $\sum_{i=1}^{p} w_i y_{i,t} = \mathbf{w}^\top \mathbf{y}_t$ where $\mathbf{w} = (w_1, ..., w_p)$ are the portfolio weights. Notice that this is an approximation. In reality we would want to work with $\sum_{i=1}^{p} w_i r_{i,t}$ (or even $\sum_{i=1}^{p} w_i p_{i,t}$) since it 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).

Assuming that we have a model for the log-returns there are many ways of estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The most simple and versatile method is the method of moments (MM), Wasserman, 2004, ch. 9. Let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)$ be a sample of log-returns. Using these, we replace $\boldsymbol{\mu}$ with the sample mean and $\boldsymbol{\Sigma}$ with 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 our sample size n is small, then our estimates are naturally imprecise. Furthermore, MPT relies on $\mathbf{S}^{-1}$ and not $\mathbf{S}$ which demands that $n > p$. Its natural to ask: does an imprecise estimate of the covariance matrix provide an equally imprecise estimate of the inverse? It turns out to be a very hard question to answer. In some simple cases the answer is no, sometimes its 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 we take the bottom-up approach. If we assume that the asset returns $\mathbf{y}_t$ follow some distribution then we can perhaps derive statistical properties for $\bar{\mathbf{y}}$ and $\mathbf{S}$. In turn, we need to derive the properties of $\mathbf{S}^{-1}$ and in the end all the transforms given by (2.4). In the coming sections we state the models used in this thesis and some important properties thereof. We thereafter discuss the implications for MPT.

3.1   Matrixvariate distributions

One of the most fundamental models for asset returns is the multivariate normal distribution. Since most of the distributions we work with are matrix variate, we will state the matrixvariate normal distribution. It is slightly more general but it can capture much more dynamics.

Definition 3.1.1 (Definition 2.2.1 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 $\mathbf{\Sigma} \otimes \mathbf{\Gamma}$ where $\mathbf{\Sigma} > 0$ is of dimension $(p \times p)$ and $\mathbf{\Gamma} > 0$ is of dimension $(n \times n)$, if $\text{vec}(\mathbf{Y}^\top) \sim N_{np}(\text{vec}(\mathbf{M}^\top), \mathbf{\Sigma} \otimes \mathbf{\Gamma})$.

The multivariate normal distribution is a simple special case of it with $\mathbf{\Gamma} = \mathbf{I}$ and $n = 1$.

The multivariate normal distribution can be criticized as a model for the asset returns at a higher frequency. It has to thin tails in comparison to what is usually observed in asset returns on higher frequencies. Cont (2001) describes a number of stylized facts of log-returns. These stylized facts describes the characteristics asset returns on different frequencies. He argues that daily returns are usually not symmetric, exhibit fatter tails than the normal distribution can account for and often show volatility clustering. However, he also argues 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 or rebalance their portfolio infrequently. That does not mean that they cannot observe the results of the market on a higher frequency than they invest!

In the univariate case we have that 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. We state its probability density function (p.d.f.) below.

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

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

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) or Alfelt (2021). A realized covariance matrix is an estimates of the volatility process from returns on a much higher frequency than we work with in this thesis. From Theorem 3.3.6 we know that if $\mathbf{Y} \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^\top, \mathbf{\Sigma} \otimes \mathbf{I}_n)$, then $n\mathbf{S} \sim W(n-1, \mathbf{\Sigma})$, so working bottom up we can get a model for the parameters of the 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 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 $\mathbf{\Gamma}$ $(p \times p)$, denoted by $\mathbf{V} \sim IW_p(m, \mathbf{\Gamma})$, if its density is given by

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

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 (p. 111 Gupta and Nagar (2000)). It demands quite specific constraints on the parameters of the model, especially $m > 2p$. To come back to out question we posed in the beginning of this section, does the inverse change uncertainty? We can at least get a hint that *something* changes with the properties of $\mathbf{S}$ when taking inverses. From Theorems 3.3.7 and 3.4.1, and Theorem 3.4.3 of Gupta and Nagar (2000) we have that

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

If n is sufficiently large, than the sample covariance matrix is (close to) unbiased. That is not necessarily the case for its inverse. If we believe in diversification then we should own many assets, e.g., p should be

large. That in turn could make the estimator very biased! The answer is yes, inverses can potentially make matters worse. Furthermore, the noise in the sample mean can be extremely large in comparison to the noise in the sample covariance matrix. The weights from (2.4) will be much more noisy whenever $\mu_0 \neq R_{GMV}$ since these will depend on the sample mean vector. Its one of the most common motivations for considering the GMV portfolio, its very hard to estimate the mean (Golosnoy, Gribisch, and Seifert, 2019). If we want to include a target return then we should aim to incorporate this uncertainty in our analysis.

It has been well established that for higher frequency returns the normal assumption is limiting. The next, and perhaps most common feature, 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

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

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

Definition 3.1.4 (Definition 3.1 Dominguez-Molina et al. (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

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

The closed skew-normal distribution is heavily parametrized. For each column in the matrix $\mathbf{Y}$ we have 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 matrix variate distribution can capture a lot of dynamics. However, it also puts a heavy restriction on some of them, 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 et al. (2007)). It is variance in time and among assets, for our application. From the stochastic representation in Proposition 2.1 Dominguez-Molina et al. (2007) one can, after some thought and derivations [1] realize that the skewness is introduced as shocks to the mean. The mean is stochastic. The overzealous parametrisation and difficulty in estimating the parameters made us choose a special case of it for paper 2 of this thesis. We work with a special case of the distribution where $q = m = 1$.

The last model we consider 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

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

where $\overset{d}{=}$ stands for equality in distribution and $\mathbf{Z} = \{z_{ij}\}$, $i = 1, 2, ..., p$, $j = 1, 2, ..., n$. Although the model can capture many types of return distributions, such as skew heavy tailed sometimes even heteroscedasticity, there is very little to say about it. The first and foremost assumption is that it is a location and scale model. In this thesis we very often assume moment conditions on the "residuals" z_{ij} such as finite fourth moment or potentially $4 + \epsilon$ finite moment, for some $\epsilon > 0$. The difference between finite fourth moment and $4 + \epsilon$ is most often the claims of convergence we can show. With the slightly more stringent assumption we can make claims about almost sure convergence and with finite fourth we can usually make claims about convergence in probability InsertRef and Ref (2020).

---

[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 7 in the same reference (page 1606).

3.2   Inference and sampling distributions of optimal portfolios and their characteristics

Replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with $\bar{\mathbf{y}}$ and $\mathbf{S}$ in (2.4) we get its empirical counterpart

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

where

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

and the shape parameter of the efficient frontier $\hat{s} = \bar{\mathbf{y}}^\top\mathbf{Q}\bar{\mathbf{y}}$. This is a portfolio we can actually invest in. It is also the portfolio that was used for plotting in Chapter 2.

From the previous section we know the distributions of $\bar{\mathbf{y}}$ and $\mathbf{S}$ if $\mathbf{Y}$ follows distribution stated in Definition 3.1.1. However, there are many complicated transforms containing both $\bar{\mathbf{y}}$ and $\mathbf{S}^{-1}$. In our example, these two parameters are independent (see Theorem 3.3.6 Gupta and Nagar (2000)) which makes the analysis simpler. The weights for the GMV portfolio which are given by

$$\frac{\mathbf{S}^{-1}\mathbf{1}}{\mathbf{1}\mathbf{S}^{-1}\mathbf{1}}$$

contain the sample covariance matrix in the nominator as well as the demoniator. Furthermore, the sample covariance matrix is everywhere so its not safe to assume that these are independent. To solve the problem at hand consider two $p \times p$ matrices with the following block structure

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \ \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \tag{3.8}$$

where $\dim(\boldsymbol{A}_{11}) = \dim(\mathbf{V}_{11}) = m \times m$, $m < p$. Let $\boldsymbol{A}_{11\cdot2} := \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ denote the Schur complement of the matrix $\boldsymbol{A}$ and define $\mathbf{V}_{11\cdot2}$ in the same manner. Let $\otimes$ denote the Kronocker product. The following theorem is used immensely 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 (3.8). Then*

(a) $\boldsymbol{A}_{11\cdot2} \sim W_m^{-1}(n - k + m, \mathbf{V}_{11\cdot2})$ and is independent of $\boldsymbol{A}_{22}$;

(b) $\boldsymbol{A}_{12}|\boldsymbol{A}_{22}, \boldsymbol{A}_{11\cdot2} \sim \mathcal{N}(\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\boldsymbol{A}_{22}, \boldsymbol{A}_{11\cdot2} \otimes \boldsymbol{A}_{22}\mathbf{V}_{22}^{-1}\boldsymbol{A}_{22})$;

(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

$$f_{\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}}(\mathbf{X}) = \frac{|\mathbf{V}_{11\cdot2}|^{-\frac{1}{2}(p-m)}|\mathbf{V}_{22}|^{\frac{1}{2}m}}{\pi^{\frac{(p-m)m}{2}}} \frac{\Gamma_m\left(\frac{n-m-1}{2}\right)}{\Gamma_m\left(\frac{n-p-1}{2}\right)}$$
$$\times \left|\mathbf{I} + \mathbf{V}_{11\cdot2}^{-1}\left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)\mathbf{V}_{22}\left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)^\top\right|^{-\frac{1}{2}(n-m-1)} \tag{3.9}$$

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

(e) $\boldsymbol{A}_{22}$ is independent of $\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$ and $\boldsymbol{A}_{11\cdot2}$;

(f) $\boldsymbol{A}_{11\cdot2}|\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} = \mathbf{X} \sim W_m^{-1}(n, \mathbf{V}_{11\cdot2} + \left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)\mathbf{V}_{22}\left(\mathbf{X} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\right)^\top)$

So given a inverse Wishart distribution we can derive the distribution of many, quite difficult, transformations of its sub-matrices. **Give an example from this setup and how to derive the distribution of different quantitites.**

Paper 1 derives the full joint distribution for all the quantities used in (3.6). The result is more general than that as we derive the distribution of all optimal portfolios but since the classical mean-variance portfolio is one of them we get that for free. The joint distribution is characterized through its stochastic representation. The stochastic representation is very verbose but it characterizes the distribution in terms of simple random variables. These random variables are simple to simulate. To understand a system we usually simulate from it and compute the quantity of interest, rinse and then repeat. For some methods simulations are the only way we can compute the quantities of interest. It can therefore be very important that simulations are fast. This is extremely simple to do when you have 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 for the train of thought that $\mathbf{Y} \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^\top, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$. To simulate from the sampling distribution of the variance of the GMV portfolio we need to

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 we compute the Cholesky decomposition. The last two options that are available is to simulate $\mathbf{S}$ directly or to derive the stochastic representation of $\hat{V}_{GMV}$ directly. Paper 1 and 2 goes into great detail to derive the stochastic representation of different quantities of optimal portfolios. One of them is the sample variance of the GMV portfolio. By Theorem 1 in Bodnar et al. (2021b) we know that if $\mathbf{Y} \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^\top, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$, then $\hat{V}_{GMV} \sim V_{GMV}\xi/(n-1)$ where $\xi \sim \chi^2_{n-p}$. We can omit inversions all together. In Chunk 3.3.1 we present 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
```

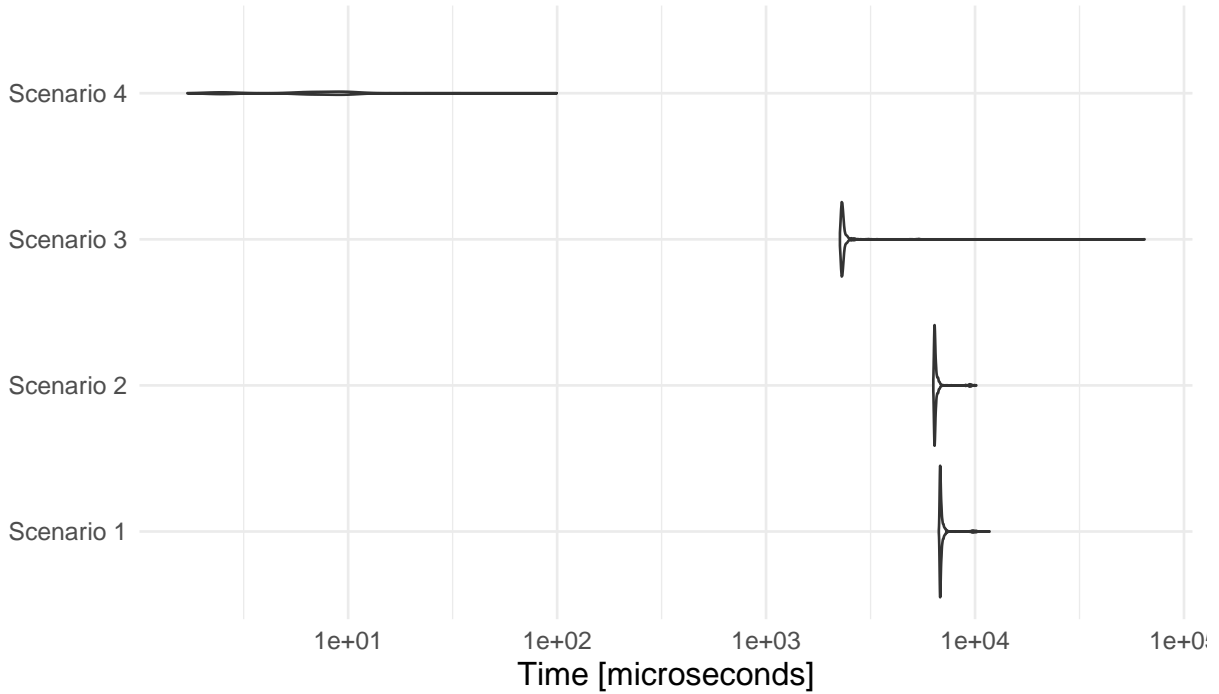[2] For the interested reader https://www.netlib.org/lapack/

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

```r
`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
)
```

Scenario 4 uses the stochastic representation. The execution time of scenario 4 is much smaller than 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 we use the cholesky decomposition or not. Its can also be a very unstable operation, especially if the matrix you are trying to invert is close to singular.

So far we have assumed that all we wanted to use is $\mathbf{S}$, $\bar{\mathbf{y}}$ or simply $\hat{\mathbf{w}}_{GMV}$. That is of course a simplification and not always the case. As we previously mentioned both $\mathbf{S}$ and $\bar{\mathbf{y}}$ can be noisy estimators with the former being less noisy than the latter. Furthermore, we saw that if p is comparable to n but $n > p$ then the expectation of the inverse Wishart distribution is very biased. We can cope with that through two strategies. The first is to derive the actual uncertainty and sample distributions of the quantities of interest, which we do in Paper 1 and 2. The second is to use other estimators which introduce some bias of our own. By introducing some bias in the estimator we can reduce the variance. It is something of utmost importance if we believe in diversification which we will go into detail in the next section.

Chapter 4

Diversification, infinitely many assets and shrinkage estimators

In the previous chapter we presented one specific way of estimating the sample covariance matrix. We also presented certain properties under certain conditions and/or statistical models. If we have a lot of data on the assets that we are trying to invest in then we can most often be certain that we will hold the correct portfolio. Our estimated portfolio will be consistent, e.g. it estimates the correct object of interest. Furthermore, since diversification is one, if not the best, risk management tool there is, we want our asset universe to be big. If we believe in diversification then p should be large. It should, in theory, decrease the risk (variance) of the portfolio. That is not always the case. By introducing one new asset to our portfolio of size p we need to estimate all covariances for that asset in the sample covariance matrix. They will constitute an additional $p+1$ quantities. The sample covariance matrix suffers from the curse of dimensionality. In terms of estimation uncertainty, this does not scale well. From Bodnar, Okhrin, and Parolya (2016) Proposition 2.2 we know 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 portfolios 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) or Bodnar, Bodnar, and Parolya (2021) and the references therein). We will focus on some topics in Random Matrix Theory (RMT) and the use of some type of shrinkage estimator. Both subjects are grand. Our aim is to provide a small introduction to them in the following sections.

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

The subject of Random matrix theory (RMT) has many applications. It was originally developed in the context of quantum physics (see Ch. 1 of Mehta (2004)). The theory and its applications has since then developed quite a lot. Many fields, such as combinatorics, computational biology, wireless communication and finance (see InsertRef and Ref (2020) for an overview) use these results. One of the seminal work in RMT was made by Wigner (1993). He originally modeled the limiting spectral distribution of an $p \times p$ dimensional standard Gaussian random matrices $\mathbf{X}$. 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 $N(0, 1)$. However, the more generalized definition only demands that the matrix $\mathbf{X}$ is Hermitian and its entries on the diagonal or above the diagonal are independent. We define the empirical spectral distribution (ESD) of a matrix $\boldsymbol{A}$ as

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

where λ_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))

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

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 introduction on why this is). Marchenko and Pastur (1967) extended the result of Wigner (1993) 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 stays constant. We usually call this ratio the concentration ratio c. In this introduction we assume that $c < 1$. 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, its not usually the object of interest. We are interested in its inverse, as we discussed in chapter 2. However, the Stieltjes transform can help us with that. 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 has many useful properties. If we know the Stieltjes transform, then we can also derive the spectral distribution F by its inversion formula. We also have pointwise convergence (see appendix B.2 of Bai and Silverstein (2010)). Using the results from RMT in MPT we take a sample covariance matrix $\mathbf{S}$ with ESD $F_n(x)$ and note that

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

If we are interested in the limiting properties of traces of inverse sample covariance matrices, we can investigate the properties of the Stieltjes transform. However, to make matters slightly worse, we are most often (at least in this thesis) interested in 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 $\mathbf{b}$. Altough $\operatorname{tr}(\mathbf{S}^{-1})$ 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 allows us to handle 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) $\mathbf{A}$ *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) $\mathbf{B} = \mathbf{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$*, 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* $c_p = p/n$

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

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 $\operatorname{tr}(\mathbf{R})/p$, obtained as the unique solution in $\mathbb{C}^+$ of the equation

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

This theorem is used repeatedly in papers 5.3 through 5.5. It is that powerful and flexible. However, we often assume finite $4 + \epsilon$ moment, while the theorem above assumes $8 + \epsilon$. We can circumvent that by Theorem InsertRef and Ref (2020)...

When we are able to construct sample estimators on the form of $\mathbf{B}$ it does not necessarily imply that we can find analytic solutions. **comment more on it.**

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 your 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 we want. We want to create a good estimator for the weights, since these are what we invest in! There are many solutions to this problem but the most common is using a shrinkage estimator. We will introduce bias to our weights but hopefully reduce the variance.

Looking at the GMV portfolio, there are two natural extensions. Either, we regularize the sample covariance matrix $\mathbf{S}$ or we regularize the weights $\hat{\mathbf{w}}_{GMV}$ directly. Lets start with the latter. The first extension is to combine the GMV portfolio weights with some target portfolio $\mathbf{b}$. We construct the shrunk portfolio weights $\hat{\mathbf{w}}_{SH}$ as

$$\hat{\mathbf{w}}_{SH} = \alpha \hat{\mathbf{w}}_{GMV} + (1 - \alpha)\mathbf{b} \tag{4.6}$$

which introduces the bias $(1 - \alpha)(\mathrm{E}[\hat{\mathbf{w}}_{GMV}] + \mathbf{b}) - \mathbf{w}_{GMV}$ but decreases the variance to

$$\mathrm{E}\left[(\alpha \hat{\mathbf{w}}_{GMV} - \alpha \mathrm{E}[\hat{\mathbf{w}}_{GMV}]) (\alpha \hat{\mathbf{w}}_{GMV} - \alpha \mathrm{E}[\hat{\mathbf{w}}_{GMV}])^\top \right] = \alpha^2 \mathrm{E}\left[(\hat{\mathbf{w}}_{GMV} - \mathrm{E}[\hat{\mathbf{w}}_{GMV}]) (\hat{\mathbf{w}}_{GMV} - \mathrm{E}[\hat{\mathbf{w}}_{GMV}])^\top \right] \tag{4.7}$$

It now stands to determine α. Shrinkage intensities are most often determined by cross validation (see e.g. James et al. (2013, ch. 5)). We try to find the best shrinkage coefficients by dividing data into a test and 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. Our aim is to determine $\min_\alpha \hat{\mathbf{w}}_{SH}^\top(\alpha)\boldsymbol{\Sigma}\hat{\mathbf{w}}_{SH}(\alpha)$. The loss depends on $\boldsymbol{\Sigma}$, which we do not know. The perhaps most obvious solution is to use the test set to estimate the $\boldsymbol{\Sigma}$. As we will see, that will have its own issues. The second solution, employed by Bodnar, Parolya, and Schmid (2018), is to first solve the optimization problem analytically. Since the estimator is unobtainable, its usually referred to a *oracle* estimator. To construct an estimator we can actually use, or what is called a *bona-fide* estimator, we take the limit of the oracle estimator. We then construct a consistent estimator for the limiting object at hand and use that one instead.

The two methods can differ quite substantially in their solution. One method is simple to implement while the other *should* be theoretically superior. In Chunk 4.2.1 we display 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 cross-validation procedure. In this example we use $n = 250, p = 150$ and $\mathbf{b}$ equal to the 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 implementation from the HDShOP package.
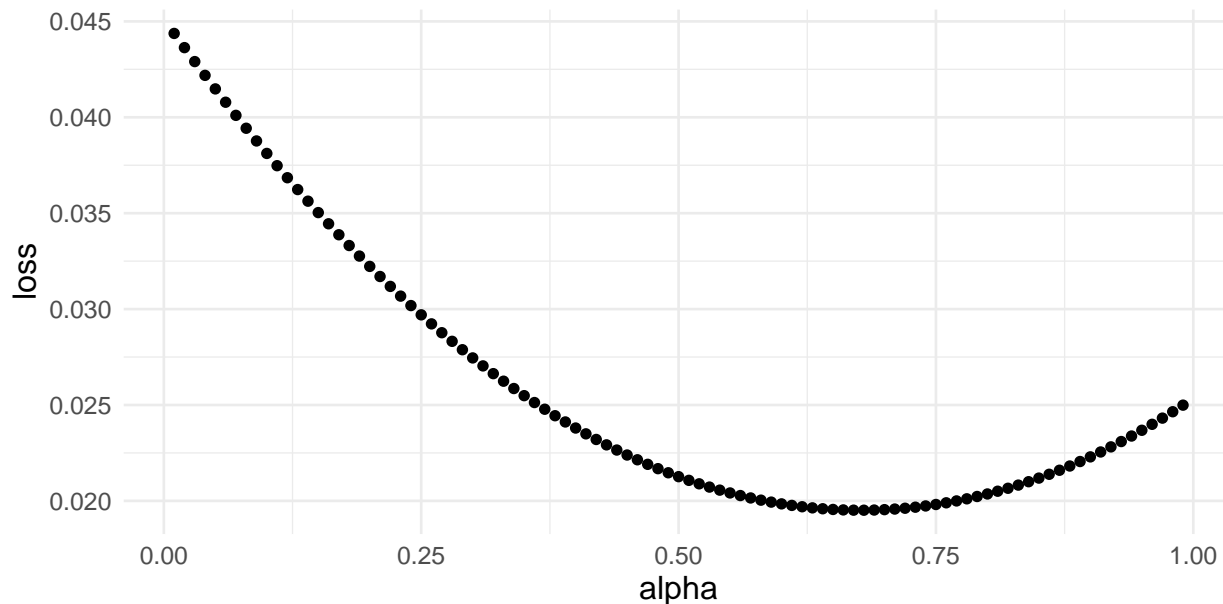
Figure 4.1: Out of sample variance estimates from the 5-fold cross validation.

```r
# setup
set.seed(123)
p <- 150
n <- 250
K <- 5
b <- rep(1,p) / p # use EW portfolio as target
# simulate params & dataset
Sigma <- HDShOP::RandCovMtrx(p)
mu <- runif(p, -0.1, 0.1)
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, pmap to map over rows
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 we illustrate the out of sample variance aggregated over folds. Cross validation suggest

that the optimal value should be 0.68. The analytic method from Bodnar, Parolya, and Schmid (2018) suggest that the optimal value is 0.7968. The "correct" value of α, 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 cross validation can give quite different results in what the optimal value should be. This linear shrinkage approach is used in Paper 5.3 through 5.5.

So far we have limited our approach to shrinking the weights. The next step is to shrink the elements of $\mathbf{S}$. If we do so then we can consider the case where $c > 1$. We leave this as an open question for the next section where we describe the papers in more detail.

Chapter 5

Summary of papers

The papers presented here are among a total of ... papers produced. These are selected based their common theme.

5.1 Paper 1 - Sampling Distributions of Optimal Portfolio Weights and Characteristics in Small and Large Dimensions

The paper investigates a fundamental question in modern portfolio theory. What are the actual implications of using the sample covariance matrix $\mathbf{S}$ and the sample mean $\bar{\mathbf{x}}$ instead of the true covariance matrix $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$? The paper does so when the returns follow a multivariate normal distribution. In it we derive the distribution for all optimal portfolios on the common form

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

for some matrix $\mathbf{L}$ of size $k \times p$ where $k < p$. To do so, we derive the stochastic representation for the joint distribution of all quantities in the equation (5.1). That enables us to efficiently simulate from the distribution, Using this representation we can easily compute quantiles of the joint distribution of the efficient frontier. Furthermore, we also derive the high-dimensional asymptotic distribution of said joint distribution. The high-dimensional asymptotic distribution is then compared to different models. One scenario considers simulations from the stochastic representation, trying to deduce the finite-sample properties. The other scenarios try to investigate what happens when we deviate from the model. As one can expect, the high-dimensional distribution works well under the assumptions though seem to be reasonably robust from deviations of the model. The hardest part to determine is what different characteristics of the return distribution have on the sample distribution of the quantities in (5.1) it remains an unanswered question.

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

In this paper we investigate a portfolio which we mention in chapter 2, the tangency portfolio. The portfolio is obtained from the quadratic utility function. It originates from the following portfolio allocation problem

$$\min_{w_0, \mathbf{w}} w_0 r_f + \mathbf{w}^\top \boldsymbol{\mu} - \frac{1}{2\gamma}\mathbf{w}^\top \mathbf{\Sigma}\mathbf{w} \tag{5.2}$$

$$\text{s.t. } w_0 + \mathbf{w}^\top \mathbf{1}_p = 1 \tag{5.3}$$

Although the portfolio is most often called the tangency portfolio it actually quantifies all portfolios on the capital market line. Increasing or decreasing the risk aversion coefficient $\gamma > 0$ decrease or increase our position in the risky assets. 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 they usually exhibit (see e.g. Cont (2001)). Similarly to Paper 1, we derive the distribution of the sample tangency portfolio. We investigate what implications the model has on the sample tangency portfolio. In short, skewness results in a bias of the portfolio weights. The investor will not hold the correct portfolio on average. Furthermore, we also investigate its high-dimensional distribution and show that it the skewness disappears asymptotically. The high-dimensional distribution is the same as the previous research have 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 deals with the fact that taking limits changes estimates. If the investor invest in a portfolio and then waits for a weak, month or year their perception of what portfolio they should hold will have changed. This is implicit when taking limits. A natural question to ask is then how to go from one portfolio to another, e.g. how to rebalance optimally when you receive a new set of data. Assuming that the investor is ok with rebalancing the portfolio at fixed time points we develop a rebalancing scheme for the GMV portfolio. This type of "dynamic" shrinkage has a lot of practical implications. If you own one portfolio it most often cost money to go from that to the next. That will take away from the return and profit you make. Furthermore, its not always possible to go from one portfolio to the next on a day. You can influence the market or make possibly make to large positional changes, which are not allowed.

   more comments...

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. A natural question to ask is then; 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 of performance measures? In this paper we investigate another common metric of evaluation, the relative out-of-sample loss. We find that...

   more comments...

5.5   Paper 5 - Double shrinkage

In this thesis we mostly use $\mathbf{S}$ and cope with that the sample covariance matrix is a noisy estimator by linear shrinkage or understanding the "noise" it provides. That implies that we have so far only consider $c < 1$. In this paper we use ... as well as the linear shrinkage from previous papers. That enables us to cover the case where $c > 1$ and potentially create a more noise-insensitive estimator of the portfolio weights. In this paper we consider the following shrinkage estimator

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

Since we are dealing with the GMV portfolio we naturally use the out-of-sample loss as a target to determine ψ and λ. Using an estimate as this leaves all potential to recieving a closed form solution in terms of a oracle estimator. What we get is a oracle loss function, which we can then derive a bona-fide estimator for.

   The model is seen to perform on par with the non-linear shrinkage of Ledoit and Wolf (2020) in all simulations and beat their method in a empirical analysis. Furthermore, it also increases/decreases the results of several other portfolio metrics.

   more comments...

5.6   Other research results

Paper 5.3 is accompanied by a R package, available on CRAN. You 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 we display a short example on how to construct the dynamic portfolio weights. 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.

```r
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
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 also coauthored throughout the writing of this thesis Bodnar et al., 2020, Bodnar, Niklasson, and Thorsén, 2021 and Bodnar et al., 2021a. 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 you report quantile-based risk measures (see The Basel Committee on Banking Supervision (2019)). The second paper provides a continuation on the first. Our idea is to model the investors beliefs explicitly and construct a prior which captures what the likelihood cant. We impose 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. It does so in a general framework, not necessarily as the same framework as MPT where we only use the first two moments of the return distribution.

Chapter 6

Future research

There are many possible extensions and future projects to the thesis at hand. We talk very much of estimation uncertainty and how to cope with it. Bayesian statistics provide a straight forward way to integrate that. However, it demands indepth knowledge of MCMC and also 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. The literature is large. Are there connections to be made and especially with Empirical Bayes?

In Paper 5.3 we consider the allocations points fixed. That assumption can be limiting for some investors. Can we incorporate that decision process into the portfolio allocation problem?

BEKK models are usually hard to fit and use for MPT. Even when their coefficients have been estimated their forecasts are not always positive definite. 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 specifc problem of fitting very large models! Recent large Natural Languange Processing models have *billions* of parameters. By placing BEKK models in this framework one also has the possibility to develop new models. The development is solely determined by constructing new layers to the networks. It would also be easier to integrate different sources of information in the models.

Bibliography

Tobin, James (1958). "Liquidity preference as behavior towards risk". In: *The review of economic studies* 25.2, pp. 65–86.

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

Marchenko, V. A. and L. A. Pastur (1967). "Verteilung der Eigenwerte in gewissen Ensembles von Zufallsmatrizen". Russian. In: *Mat. Sb., Nov. Ser.* 72, pp. 507–536.

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.

Wigner, Eugene P (1993). "Characteristic vectors of bordered matrices with infinite dimensions i". In: *The Collected Works of Eugene Paul Wigner*. Springer, pp. 524–540.

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

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

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

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

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.

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

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

Dominguez-Molina, J Armando et al. (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.

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

Bodnar, Taras and Arjun Gupta (2009). "Construction and Inferences of the Efficient Frontier in Elliptical Models". In: *Journal of the Japan Statistical Society* 39.

Bai, Zhidong and Jack W Silverstein (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

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

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

Hult, Henrik et al. (2012). *Risk and portfolio analysis: Principles and methods*. Springer Science & Business Media.

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.

James, Gareth et al. (2013). *An introduction to statistical learning*. Vol. 112. Springer.

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

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.

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.

Livan, Giacomo, Marcel Novaes, and Pierpaolo Vivo (2018). *Introduction to random matrices: theory and practice*. Vol. 26. Springer.

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

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}.

Bauder, David et al. (2020). "Bayesian mean–variance analysis: optimal portfolio selection under parameter uncertainty". In: *Quantitative Finance*, to appear.

Bodnar, Taras et al. (2020). "Bayesian Quantile-Based Portfolio Selection". In: *Accepted with revision to Applied Mathematics and Computation*.

InsertRef, Ref and Insert Ref (2020). "Insert reference". In: *Insert Ref* Insert Ref.5, pp. 0–1. DOI: REF. URL: REF.

Ledoit, Olivier and Michael Wolf (2020). "Analytical nonlinear shrinkage of large-dimensional covariance matrices". In: *The Annals of Statistics* 48.5, pp. 3043 –3065. DOI: 10.1214/19-AOS1921. URL: https://doi.org/10.1214/19-AOS1921.

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

Bodnar, Olha, Taras Bodnar, and Nestor Parolya (2021). "Recent advances in shrinkage-based high-dimensional inference". In: *Journal of Multivariate Analysis*, p. 104826.

Bodnar, Taras, Vilhelm Niklasson, and Erik Thorsén (2021). "Volatility Sensitive Bayesian Estimation of Portfolio VaR and CVaR". In: *Submitted for publication*.

Bodnar, Taras, Nestor Parolya, and Erik Thorsén (2021a). "Dynamic Shrinkage Estimation of the High-Dimensional Minimum-Variance Portfolio". In: *arXiv preprint arXiv:2106.02131*.

— (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 et al. (2021a). "Quantile-based optimal portfolio selection". In: *Computational Management Science*, pp. 1–26.

Bodnar, Taras et al. (2021b). "Sampling distributions of optimal portfolio weights and characteristics in small and large dimensions". In: *Random Matrices: Theory and Applications*, p. 2250008. DOI: 10.1142/S2010326322500083.

Javed, Farrukh, Stepan Mazur, and Erik Thorsén (2021). "Tangency portfolio weights under a skew-normal model in small and large dimensions". In: *Submitted for publication*.

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