Erasmus University Rotterdam

Erasmus School of Economics

Bachelor's Thesis Econometrics and Operations Research

Efficient estimation of large scale Markowitz portfolios

Dylan Buter
415000

Supervisor Sander Barendse

Second Assessor Bart Keijsers

Abstract

This paper focuses on treating the estimation problem of the Markowitz portfolio in a large market. Several methods that have been introduced to solve this problem are investigated. The first method comes from Chen and Yuan (2016), who propose the so called subspace mean-variance portfolio, which introduces a trade-off between optimality and estimability. Another way solution for the estimation problem comes in the form of shrinkage estimation. Shrinkage estimators for the mean and variance of the asset returns have been introduced by Jorion (1986) and Ledoit and Wolf (2003) respectively. Furthermore, a combination of shrinkage estimation and the subspace mean-variance portfolio is made. Consequently, it is found through simulation that these methods are beneficial in terms of Sharpe ratio. However, these benefits were only seen in large markets when studying real return data. Furthermore, the combination of shrinkage estimation and the subspace mean-variance portfolio didn't result in a better or worse portfolio rule, which implies a robustness of the subspace method towards the quality of the in-put sample estimates.

July 2, 2017

Contents

1 Introduction

Investors can choose to invest their wealth in countless risky assets available on the financial market. Economic literature shows that investors have to diversify their wealth over all assets in order to reduce risk and optimize profits (see, e.g. Haugen (1986) or Berk and DeMarzo (2014)). This practice is commonly referred to as portfolio selection and has existed for a long time. However, portfolio selection methods were sparse until the seminal work of Markowitz (1952), before then most investors were primarily focused on finding assets with a high expected return and neglected covariance. Markowitz introduced a theoretical basis for investment strategy, which finds the optimal wealth allocation given both the mean and variance of the asset-returns, resulting in a portfolio that is mean-variance efficient. In practice, this method requires the estimation of the mean and variance from empirical returns, where the resulting sample estimates are then used as if they were the true parameters, also known as plug-in estimation. The practical issue of plug-in estimation is estimation error, which causes inadequate estimations of the theoretical mean-variance portfolio in small samples. The problem becomes worse when there is a larger amount of assets in the investment universe, as the amount of unknown parameters increases.

The Law of Large Numbers shows that sample estimates are asymptotically consistent, meaning that with enough data appropriate estimations of the mean and variance can be made. This result translates into consistent estimates of the mean-variance portfolio weights. However, estimates of the optimal portfolio are inaccurate in small samples (DeMiguel et al., 2009). When the amount of assets in the portfolio increases, the amount of unique parameters rises at a quadratic rate, as the covariance matrix and the mean vector grow larger. The result is an excessive amount of estimation error when using the standard sample moments, which causes the estimated mean-variance portfolio to perform poorly in small samples.

Several extensions on the traditional Markowitz Portfolio have since been introduced to reduce estimation error, one of which is the method of shrinkage estimation. Shrinkage estimators are alternative ways of estimating sample moments by creating a convex combination with the traditional sample estimate and a given target. Shrinkage estimation has improved estimation power for higher dimensional multivariate normal variables (Stein, 1956), when using a quadratic loss function for the estimators. Shrinkage estimators of the mean and variance for portfolio selection purposes have been proposed by Jorion (1986) and Ledoit and Wolf (2003) respectively.

Chen and Yuan (2016) propose an alternative method of handling estimation error. By restricting the investment space to a specific subspace, they find that the subspace mean-variance portfolio has better estimability in practice, as it has Sharpe ratios relatively closer to that of the theoretical optimal portfolio.

This paper is focused on comparing the small sample properties of existing portfolio selection methods that are extensions on the Markowitz portfolio with the goal to identify the best method to compute the portfolio weights. The investigated methods are the subspace mean-variance portfolio and the shrinkage estimators of Jorion (1986) and Ledoit and Wolf (2003). Furthermore, a combination of shrinkage estimation and the subspace mean-variance portfolio is made. The research gives a critical view on the effectivity of the subspace mean-variance portfolio and the shrinkage estimators, while also giving more insights into the shortcomings of the traditional sample estimation methods.

Consequently, it is found that the subspace method estimates better sample portfolios in larger markets. Furthermore, Shrinkage estimation is found to have larger estimation power than the traditional sample estimators in larger markets. And finally, the combination of shrinkage and subspace estimation didn't result in a significant improvement. However, the findings suggest that in larger markets, the traditional Markowitz portfolio can be improved by either shrinkage estimation or the subspace mean-variance portfolio.

2 Literature

Modern portfolio theory started with the seminal work of Markowitz (1952). In the proposed framework of Markowitz, an investor is able to find the optimal portfolio for a single period, given that the investor only takes the mean and variance into account. The framework of Markowitz is able to show that by combining imperfectly correlated assets an investor can create a portfolio with better expected return-risk properties. Furthermore, Markowitz shows that when a portfolio is fully diversified an investor can only get a higher expected return by taking on more risk. Since the Markowitz framework is able to capture these two important economic aspects of portfolio theory and it being easily extendable, i.e. multi-period problems or continuous time problems, the framework has become the leading portfolio model within finance.

However, Markowitz's mean-variance portfolio isn't flawless, for instance the portfolio doesn't take into account population moments higher than two. Furthermore, the Markowitz portfolio has inaccurate estimability in practice. The issue of plug-in estimates has long been noted in empirical literature (see e.g. Jobson and Korkie (1980), Michaud (1989), Best and Grauer (1991)). The general findings of these papers are that the sample portfolios are imprecise even when a large sample is used. DeMiguel et al. (2009) compares the sample mean-variance portfolio and further extensions on the portfolio to naive diversification where each asset is given equal weights. They find that around 3000 months (250 years) of data is needed for the sample mean-variance portfolio and it's extensions to consistently outperform naive diversification for a portfolio with 25 assets. The findings strongly indicate that current existing portfolio selection methods are insufficient given the realistic amount of data available in practice.

Stein (1956) introduced a method where a sample estimate is made by "shrinking" the sample moment towards a chosen target value. He finds that the shrunken estimator has an uniformly lower risk of a higher loss for the quadratic loss function than the traditional sample mean, when the dimension of the multivariate normal variable is larger than 2. In a simulation analysis, Jorion (1986) finds that his proposed Bayes-Stein estimator for the mean always outperforms the usual sample mean. Ledoit and Wolf (2003) extend the practice of shrinkage estimation towards the covariance matrix, where the sample covariance matrix is shrunk towards the single-index covariance matrix of Sharpe (1963). They find that their estimator has significantly lower out-of-sample variance than other well-known estimators when analyzing return data for portfolio selection.

In the framework of Chen and Yuan (2016), the portfolio weights are restricted to a subspace spanned by the leading eigenvectors of the covariance matrix. By restricting the investment space the subspace portfolio is less optimal, however the improved accuracy of the proposed subspace portfolio results in better sample portfolios in terms of Sharpe ratio. The research of Chen and Yuan (2016) show a trade-off between optimality and estimation accuracy similar to the bias-variance trade-off. Through a simulation study, they find that the sample subspace portfolio is relatively better than the traditional sample portfolio in estimating the theoretical Markowitz portfolio.

Another way to reduce estimation error is by assuming a model structure for the returns, such as a factor model. Factor models for return analysis originate from Sharpe (1963), who proposes a single-factor market model commonly referred to as the Capital Asset Pricing Model (CAPM). Other factor models, are for example the Fama-French model (Fama and French, 1993) or models based on principal components (see e.g. Connor and Korajczyk (1988)). Factor models have direct implications on the covariance matrix, which can substantially lower the amount of unknown parameters, resulting in lower estimation error. However, the method is prone to model misspecification. Research shows that factor models are able to significantly improve the performance of plug-in estimates if the model is correctly specified. Nonetheless, there is no clear consensus on which factor model should be preferred (Chan et al., 1999). For a detailed overview on efficient portfolio selection see Brandt (2010).

3 Data

The results of this paper are based on a simulation study and an empirical study. For the simulation, the monthly Fama-French 3 research factors are gathered from the period July 1963 to August 2007. The 3 Fama french factors are: $r_m - r_f$; SMB (Small Minus Big); and HML (High Minus Low). The factor $r_m - r_f$ is the excess market return, this factor will also be used to calculate the shrinkage estimator of Ledoit and Wolf (2003). For the empirical study, the monthly equal weight average returns on the Fama-French 25 (5×5) and 100 (10×10) formed on size and book to market are used as the asset returns. Data is collected from a 50 year period (Januari 1961 to December 2010). As the dataset is incomplete for a small number of portfolios, the missing data for a given portfolio is sampled from a normal distribution with mean and variance calculated from the data that is available for that portfolio. Furthermore, a moving window of T months is used such that the first T observations serve as the initial estimation sample to estimate the holdings at time $T + 1$. All of the datasets and further descriptions on the datasets can be found at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

4 Methodology

4.1 Markowitz Mean-Variance Portfolio

Let there be N risky assets available on the market, such that $r_t \in \mathbb{R}^N$ is the vector containing the excess returns on these assets at time $t = 1, \ldots, T$. Excess returns are defined as the returns in excess of the risk-free rate, which is the return on an investment that contains no risk. A portfolio is created by choosing $w \in \mathbb{R}^N$, which denotes a set of weights that allocates wealth over all assets. The optimal Markowitz portfolio results from the solution to the optimization problem shown in equation (1).

$$\min_{w \in \mathbb{R}^N} \{\frac{\gamma}{2} w^\intercal \Sigma w - w^\intercal \mu\} \tag{1}$$

In the equation above $\mu \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^{N \times N}$ denote the mean and variance of r respectively, furthermore γ denotes the coefficient of relative risk aversion. The solution to this problem, denoted by w_{mv}, is given in equation (2).

$$w_{\text{mv}} = \frac{1}{\gamma} \Sigma^{-1} \mu \tag{2}$$

In this representation of the Markowitz framework there are no constraints on the weights and/or a minimal target return. This paper assumes that the investor can borrow money indefinitely such that no constraint on the weight vector is needed. Furthermore, the minimal target return is resolved by the risk coefficient γ, as investors will be able to get a higher expected portfolio return by lowering their γ and by doing so accept more risk. The derivations of the Markowitz portfolio with and without constraint on the weights are both given in the appendix.

The weights that result from the Markowitz problem, incorporates both the mean and variance in the optimization problem and thus results in weights that are mean-variance efficient. However, the calculation of w_{mv} requires the knowledge of both μ and Σ, which are usually unknown in practice. Instead the parameters are estimated from historical data and the estimates $\hat{\mu}$ and $\hat{\Sigma}$ are then plugged into equation (2) which gives an the sample mean-variance portfolio, as shown in equation (3).

$$\hat{w}_{\text{mv}} = \frac{1}{\gamma} \hat{\Sigma}^{-1} \hat{\mu} \tag{3}$$

4.1.1 Subspace Mean-Variance Portfolio

Chen and Yuan (2016) propose an alternative portfolio selection method similar to the Markowitz problem, where a trade-off is made between the optimality of the mean-variance portfolio and the estimability in practice. This is done by restricting w to a linear subspace of $\mathbb{R}^N$, denoted by $\mathcal{P}$. The new optimization problem is shown in equation (4).

$$\min_{w \in \mathcal{P}} \{\frac{\gamma}{2} w^\intercal \Sigma w - w^\intercal \mu\} \tag{4}$$

The solution to this problem, $w_{mv}^{\mathcal{P}}$, is shown in equation (5)) (see Appendix for the derivation).

$$w_{mv}^{\mathcal{P}} = \frac{1}{\gamma} P_{\mathcal{P}} (P_{\mathcal{P}}^{\intercal} \Sigma P_{\mathcal{P}})^{-1} P_{\mathcal{P}}^{\intercal} \mu \tag{5}$$

Where $P_{\mathcal{P}}$ is the matrix that forms the orthonormal basis of $\mathcal{P}$. The choice of $\mathcal{P}$ is ad hoc, one can choose to construct $\mathcal{P}$ from any existing portfolio rules as the linear subspace that is spanned from those rules. Extreme choices are $\mathcal{P} = \mathbb{R}^N$ and $\mathcal{P} = \{a\mathbf{1} : a \in \mathbb{R}\}$, which denote the theoretical Markowitz framework and naive diversification respectively. In this paper $\mathcal{P}$ is chosen to be space spanned by the leading eigenvectors of the covariance matrix of the returns. Such that the orthonormal basis of $\mathcal{P}$ are the first d eigenvectors of Σ: $P_{\mathcal{P}} = [\eta_1, \ldots, \eta_d]$, plugging this into equation (5) gives the solution shown in equation (6).

$$w_{mv}^{\mathcal{P}} = \frac{1}{\gamma} \sum_{k=1}^{d} \theta_k^{-1} \eta_k \eta_k^{\intercal} \mu \tag{6}$$

In the equation above θ_k, $k = 1, \ldots, d$, are the first d eigenvalues of the covariance matrix. The choice of the dimension size d remains ad hoc and has to be estimated from data. Bai and Ng (2002) propose several methods of determining the dimension in approximate factor models. One of their methods chooses the dimension size according to equation (7).

$$\hat{d} = \underset{1 \le k \le k_{max}}{\arg\min} \left\{ \log\big(\sum_{j>k} \hat{\theta}_j\big) + \frac{k(N+T)}{NT} \log\big(\frac{NT}{N+T}\big) \right\} \tag{7}$$

Where T is the amount of observations and k_{max} is the maximum dimension size, which is chosen to be $k_{max} = 8$ in this paper. Furthermore, Bai and Ng (2002) show that $\hat{d}$ is an consistent estimator for the correct dimension of the factor model if both N and T are large.

The subspace mean-variance portfolio is a variation on the Markowitz mean-variance portfolio. The knowledge of μ and Σ is still necessary for calculation, however by restricting the weights to a subspace, i.e. the space spanned by the leading eigenvectors, an investor is able to get accurate portfolio estimates.

4.2 Parameter Estimation

4.2.1 Plug-in Estimators

Traditional sample estimators for the mean μ and covariance Σ are shown in (8).

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t, \qquad \text{and} \qquad \hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} (r_t - \hat{\mu})(r_t - \hat{\mu})^{\intercal} \tag{8}$$

These estimators are widely used because they are unbiased and have satisfactory asymptotic properties. However, these estimators aren't the most efficient in small samples as they can be less accurate. Literature shows that the problem can be slightly relieved by applying "shrinkage", which should improve efficiency by using additional information on the variable at hand.

4.2.2 Shrinkage Estimators

Shrinkage estimation stems from the work of Stein (1956), who observed that for a multivariate normal variable with $N \geq 3$ the sample mean $\hat{\mu}$ isn't the most efficient estimator of μ in terms of the quadratic loss function. Stein shows that the sample mean is dominated by a convex combination of the sample mean and a common constant μ_0 as shown in equation (9) (James and Stein, 1961).

$$\hat{\mu}_s = \delta\mu_0 + (1-\delta)\hat{\mu} \tag{9}$$

Where $0 < \delta < 1$ denotes the shrinkage factor, which represents the shrinkage intensity. The benefits of shrinkage estimation is optimal by picking the shrinkage factor δ^* according to equation (10).

$$\delta^* = \frac{N+2}{N+2+T(\hat{\mu}-\mu_0\mathbf{1})^\intercal\tilde{\Sigma}^{-1}(\hat{\mu}-\mu_0)\mathbf{1}} \tag{10}$$

Where $\tilde{\Sigma} = \frac{T-1}{T-N-2}\Sigma$. In practice the optimal δ^* is unknown when Σ is unknown and is estimated from data. Brandt (2010) shows that even when Σ is unknown, the shrinkage estimator has a significant improvement in estimation. Furthermore, for any choice of μ_0 the shrinkage estimator will have a lower risk of having a high loss. However, the shrinkage estimator performs best when it's close to the true μ, in practice μ_0 is often chosen to be the grand mean. Jorion (1986) introduces a shrinkage estimator in a empirical Bayes framework. The Bayes-Stein estimator is also used in this paper. This estimator makes use of the shrinkage target shown in equation (11).

$$\hat{\mu}_0 = \frac{\mathbf{1}^\intercal\hat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}^\intercal\hat{\Sigma}^{-1}\mathbf{1}} \tag{11}$$

Shrinkage estimation can also be extended to the covariance matrix as Frost and Savarino (1986) and Ledoit and Wolf (2003) propose. Similar to the shrinkage estimator for the mean, the covariance estimator is a convex combination of the sample covariance matrix and a target matrix S_0, which is often chosen to be the identity matrix or the implied covariance matrix by an assumed factor model. The shrinkage estimator for the covariance matrix is shown in equation (12).

$$\hat{\Sigma}_s = \tau S_0 + (1-\tau)\hat{\Sigma} \tag{12}$$

Where $0 < \tau < 1$ is the shrinkage factor. Ledoit and Wolf (2003) used Sharpe (1963)'s single-index covariance matrix as the shrinkage target, which is also the shrinkage target in this paper. The single index covariance matrix follows from the Single-Index Model (SIM), which assumes that each return is generated by a constant, the market return and an idiosyncratic error, as shown in equation (13).

$$r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it} \tag{13}$$

In this model the returns are driven by the market returns r_{mt} with the respective loadings β_i and an error term ε_{it}, furthermore the market returns are uncorrelated with the error terms. The factor structure implies a covariance

Σ_s as shown in equation (14).

$$\Sigma_s = \sigma_m^2 \beta\beta' + \Sigma_\varepsilon \tag{14}$$

Where σ_m^2 is the variance of the market returns, β is the vector containing the loadings and Σ_ε is the diagonal variance matrix of the errors. The variance structure can be estimated by performing regressions of the stock returns on the market return, which results in estimates of both the loadings and the error terms, b_i and e_{it} respectively. With these estimates, Σ_s can be estimated as in equation (15).

$$\hat{\Sigma}_s = s_m^2 bb' + \hat{\Sigma}_\varepsilon \tag{15}$$

Where s_m^2 is the sample variance of the market returns, b is the vector containing the estimated loadings and $\hat{\Sigma}_\varepsilon$ is the sample covariance of the error terms. The estimate of the single index covariance matrix $\hat{\Sigma}_s$ is then used as the shrinkage target. Furthermore, the optimal shrinkage factor τ^* is shown in equation (16).

$$\tau^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(s_{ij}) - \mathrm{Cov}(f_{ij}, s_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2} \tag{16}$$

Where, s_{ij} is the (i,j)-th element of $\hat{\Sigma}$, f_{ij} is the (i,j)-th element of $\hat{\Sigma}_s$, ϕ_{ij} is the (i,j)-th element of Σ_s and σ_{ij} is the (i,j)-th element of Σ. One of the technical assumptions that should be made, is that the population covariance matrix Σ and the implied covariance matrix Σ_s are unequal, such that ϕ_{ij} and σ_{ij} are also unequal. The formulas of the estimators to calculate the optimal shrinkage factor are shown in the appendix.

The idea of shrinkage follows from the bias-variance trade-off. The shrinkage target results from assumptions on the model for the returns and thus usually contain bias as the model might be misspecified, such that shrinkage estimation can perform poorly if the structural assumptions are poor. Shrinkage estimation combines the traditional unbiased estimators that contain a lot of estimation error with biased estimators with low estimation error in the hope of attaining the optimal bias-variance trade-off. This paper uses the shrinkage estimators of Jorion (1986) and Ledoit and Wolf (2003) independently to inspect if they can be of use for portfolio selection by providing more accurate estimates of the population moments.

4.3 Sharpe Ratio

The Sharpe ratio is a commonly used portfolio measure that relates the portfolio's return to the portfolio's risk, measured by volatility (Sharpe, 1966). Given w the Sharpe ratio can be calculated using equation (17).

$$SR(w) = \frac{w^\intercal \mu}{(w^\intercal \Sigma w)^{\frac{1}{2}}} \tag{17}$$

The Sharpe ratio measures the amount of excess return per unit of standard deviation. Implying that a portfolio with a higher Sharpe ratio would be preferred over a portfolio with lower Sharpe ratio even if it has a higher return, because it's more risk efficient. The optimal feature of the mean-variance portfolio is that it's the portfolio with the highest possible Sharpe ratio given the market (see equation (18)).

$$SR(w) \leq SR(w_{\text{mv}}) = (\mu^{\intercal}\Sigma^{-1}\mu)^{\frac{1}{2}}, \quad \forall w \in \mathbb{R}^N \tag{18}$$

Using the delta method, one can show that the estimated mean-variance portfolio will reach the Sharpe ratio of the theoretical mean-variance portfolio when $T \longrightarrow \infty$ and N is fixed (see equation (19)).

$$SR(\hat{w}_{\text{mv}}) \xrightarrow{p} SR(w_{\text{mv}}) \tag{19}$$

However, this asymptotic result will most likely not occur in a finite sample due to the estimation error caused by the large set of unknown parameters and the insufficient amount of data available in practice. From this the following theorem can be constructed (Chen and Yuan, 2016).

Theorem 1. *Assume that the excess returns r_t, $t = 1, 2, \ldots, T$ are independently and normally distributed with mean μ and covariance Σ. Then the Sharpe ratio of the sample mean-variance portfolio satisfies Theorem 1.*

$$\frac{SR(\hat{w}_{mv})}{SR(w_{mv})} = O_p(\sqrt{T/N})$$

The theorem states that when the number of assets increase and the estimation window is fixed, the relative efficiency of the sample Markowitz portfolio converges to zero in probability. Under a set of general assumptions (see Appendix), Chen and Yuan (2016) state the following theorem for the subspace mean-variance portfolio.

Theorem 2. *Let d be the correct dimension size, fixed and finite. If the returns follow a Arbitrage Pricing Theory model which satisfies a set of mild assumptions, then:*

$$SR^2(\hat{w}_d) = SR^2(w_{mv}) + O_p(T^{-1/2} + N^{-1/2})$$

The theorem suggests that the Sharpe ratio of the estimated subspace portfolio will converge to the Sharpe ratio of the theoretical Markowitz portfolio if there are both lots of assets and observations.

The Sharpe ratio is used to make comparisons between the different portfolio estimation methods. Furthermore, the Sharpe ratio is used to inspect if the properties of the estimated portfolios approach that of the theoretical Markowitz mean-variance portfolio.

4.4 Simulation

The returns are simulated from a k-order common factor model, where r_t follows the model shown in equation (20).

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \beta_{21} f_{2t} + \cdots + \beta_{k1} f_{kt} + \varepsilon_{it}, \quad i = 1, \dots, N; \; t = 1, \dots, T \tag{20}$$

Where α is Jensen's alpha, f_{jt} is the j-th common factor, β_{ij} is the loading for asset i on factor j and ε_{it} is the idiosyncratic error. The factors follow the normal distribution with mean μ^f and covariance Σ_f. The error terms are normally distributed with mean zero and variance Σ_ε, which is a diagonal matrix as the error terms are independently distributed. Furthermore, the factors are also independently distributed from the error terms, such that $\mathbb{E}(\varepsilon_{jt}|f_t) = 0$. Given the distributions of the factors and the error-terms, the distribution of the returns is known. The returns are normally distributed with mean $\mu_i = \alpha_i + \beta_{i1}\mu_1^f + \ldots + \beta_{ik}\mu_k^f$ for asset i and covariance $\Sigma = B\Sigma_f B^\intercal + \Sigma_\varepsilon$, where B $(N \times K)$ contains all the factor loadings. Arbitrage Theory states that Jensen's alpha is equal to zero. Furthermore, the assumption is made that the risk premium μ^f for each factor is bigger than zero.

The simulation set-up follows that of Chen and Yuan (2016), where a Fama-French three factor model is used to simulate N assets using equation (20). The factors are generated from a normal multivariate distribution with mean and covariance calculated from historical data on the 3 Fama-French factors (July 1963 - August 2007). The factor loadings in the simulation are sampled from a uniform distribution with values ranging from 0.9 and 1.2 for $r_m - r_f$; -0.3 and 1.4 for SMB; and -0.5 and 0.9 for HML. Furthermore, the error terms are generated from a normal distribution with a zero mean and a diagonal covariance matrix, whose elements are sampled from a uniform distribution with values ranging from 0.10 and 0.30. To observe asymptotic behaviour the simulation lasts 1000 iterations and the results averaged over all iterations.

5 Results

The result section is organized as follows, first a critical analysis on the subspace mean-variance portfolio of Chen and Yuan (2016) is made. Then, an evaluation of the benefits from shrinkage estimation follows. Finally, all the different portfolio methods used in this paper are compared and a combination rule of shrinkage estimation and the subspace method is investigated.

5.1 Subspace Mean-Variance Portfolio

First, the effect of the dimension d on the subspace mean-variance portfolio is investigated. Plots of the relative Sharpe ratio for different dimension sizes can be found in figure 1, where plots are shown for $N =$ 25, 100 and $T =$ 60, 120, 240. Where the black line represents the sample subspace mean-variance portfolio and the grey line represents the naive portfolio. The relative Sharpe ratio is defined as the Sharpe ratio of a given portfolio rule divided by the Sharpe ratio of the theoretical Markowitz portfolio.

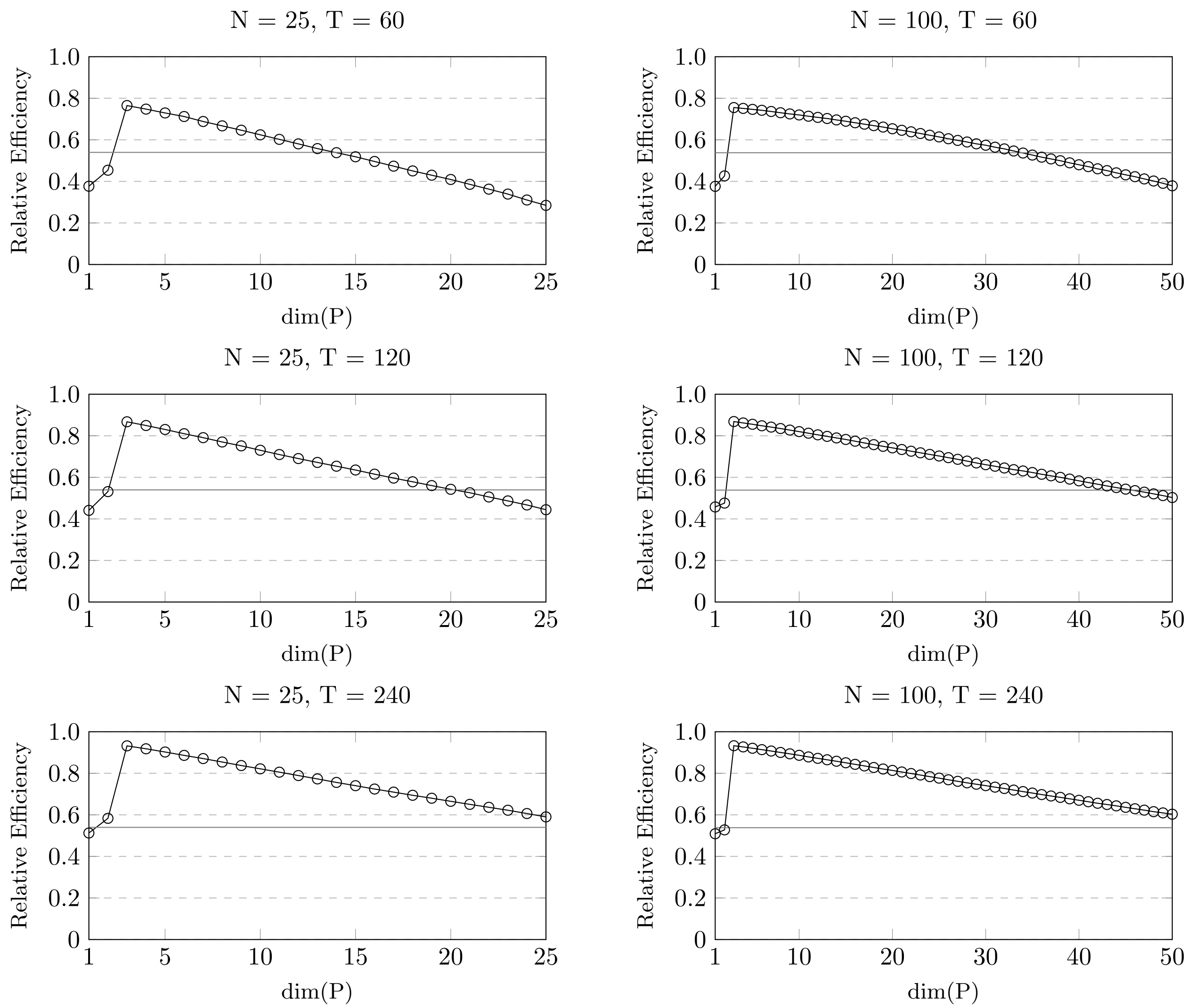


Figure 1: Plots of the relative efficiency of the estimated Subspace mean-variance portfolio at different dimension sizes d. The plot is made for all combinations $N =$ 25, 100 and $T =$ 60, 120, 240. The horizontal lines are the average relative efficiency of naive diversification.

Figure 1 shows that the subspace mean-variance portfolio performs best when the d is equal to the true amount of economic factors that generate the returns, which is three in the simulation. For lower dimensions the subspace method fails capture the full effect of all the factors. When the dimension grows larger, the subspace method starts to suffer from estimation error. However, the subspace mean-variance portfolio performs better than the naive portfolio at a wide range of other dimension sizes close to the true dimension size, especially for larger N and T where the subspace mean-variance portfolio outperforms the naive portfolio at almost all investigated dimensions. The finding support Theorem 2, which states that the performance of the subspace mean-variance portfolio improves when both N and T increase. Furthermore, the figure shows a slight concave form, indicating that the negative effects of estimation error increases as the dimension size increases, because the amount of parameters to be estimated increases at an increasing rate such that the amount of estimation error should also be increasing at an increasing rate.

To investigate the practical benefits of the subspace mean-variance portfolio, d is now estimated using equation (7). The Sharpe ratio of the different portfolio rules at different N is shown in Figure 2, where plots are shown for $T =$ 60, 120, 240, 360.

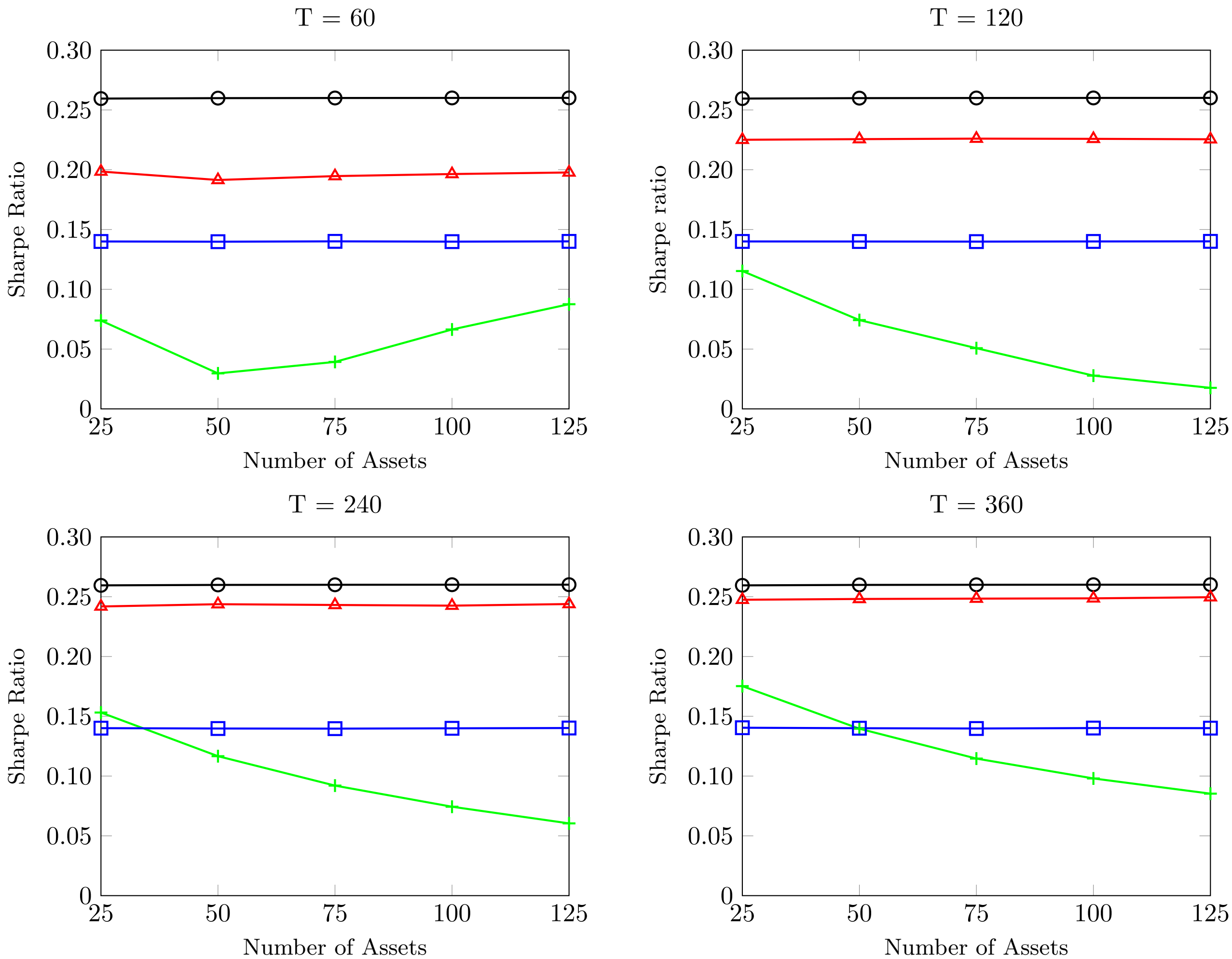


Figure 2: Plots of the Sharpe ratios at different N. The plot is made for estimation windows $T =$ 60, 120, 240, 360. Furthermore, the Sharpe ratios of the theoretical mean-variance portfolio (black, circles), the estimated subspace mean-variance portfolio (red, triangle), naive diversification (blue, squares) and the sample mean-variance portfolio (green, crosses) are shown.

Figure 2 shows several strong qualities of the subspace mean-variance portfolio. First of all, the Sharpe ratio of the subspace mean-variance portfolio remains relatively unaffected to the increase in N, while the sample mean-variance portfolio sees a rapid drop in Sharpe ratio. Furthermore, the subspace mean-variance portfolio consistently outperforms both the naive portfolio and the sample mean-variance portfolio, whereas the sample mean-variance portfolio quickly gets outperformed by the naive portfolio when N increases. Lastly, both the subspace mean-variance portfolio and the sample mean-variance portfolio show higher Sharpe ratios when T is larger, which supports both Theorem 1 and Theorem 2.

The results shown in Figure 2 differ from the findings in Chen and Yuan (2016). In their paper the Sharpe ratios are overall much higher and show much bigger jumps between different values of N. In Figure 3 the Sharpe ratio of the theoretical mean-variance portfolio is shown at $N = 1$ to 50 and $T = 120$.

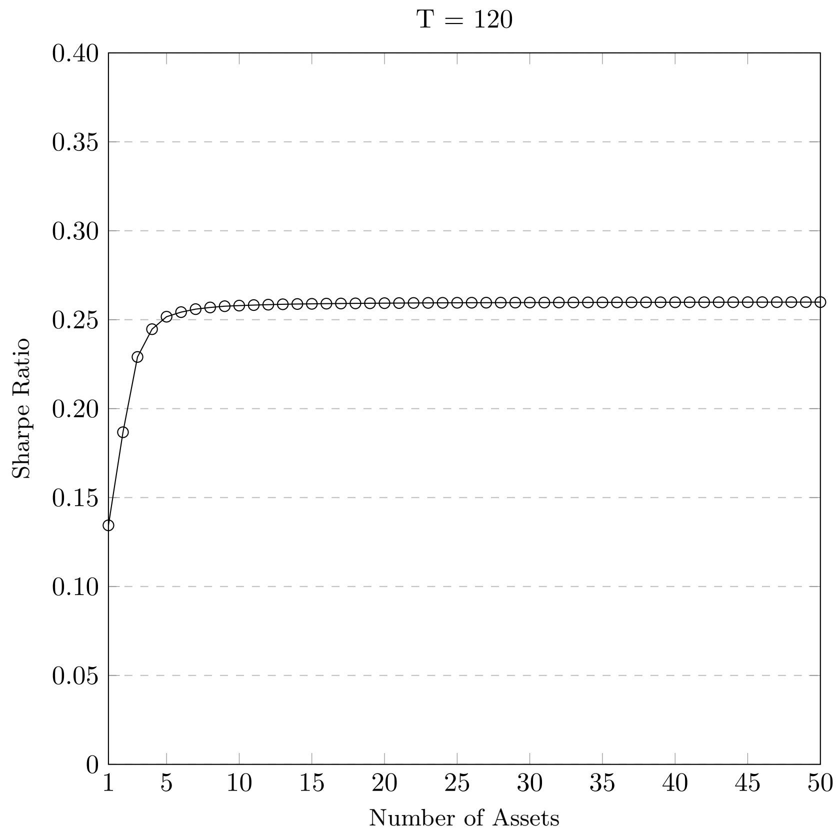


Figure 3: A plot of the Sharpe ratio of the Markowitz mean-variance portfolio as N increases.

Figure 3 shows that when N grows larger than three, the Sharpe ratio quickly starts to converge to a value around 0.26, which is also seen in Figure 2. This is because the assets are generated from a model with only three common economic factors. Portfolio theory states that the benefits of diversification aren't endless due to systemic risk (Tasca and Battiston, 2011), which implies that the Sharpe ratio should convergence after a certain amount of assets are included in the portfolio. Because only three common factors are used in the simulation, an investor doesn't need a large set of assets to reduce exposure towards the risk of the common factors, therefor the Sharpe ratio converges quickly when $N > 3$. This differs from what is shown in Chen and Yuan (2016), where there are large spikes in Sharpe ratio, such as between $N = 50$ and $N = 75$.

To further investigate the properties of the subspace mean-variance portfolio, the holdings will be compared to those of the Markowitz portfolio. In Figure 4 the holdings of the subspace mean-variance portfolio are plotted against the theoretical mean-variance portfolio and the sample mean-variance portfolio.

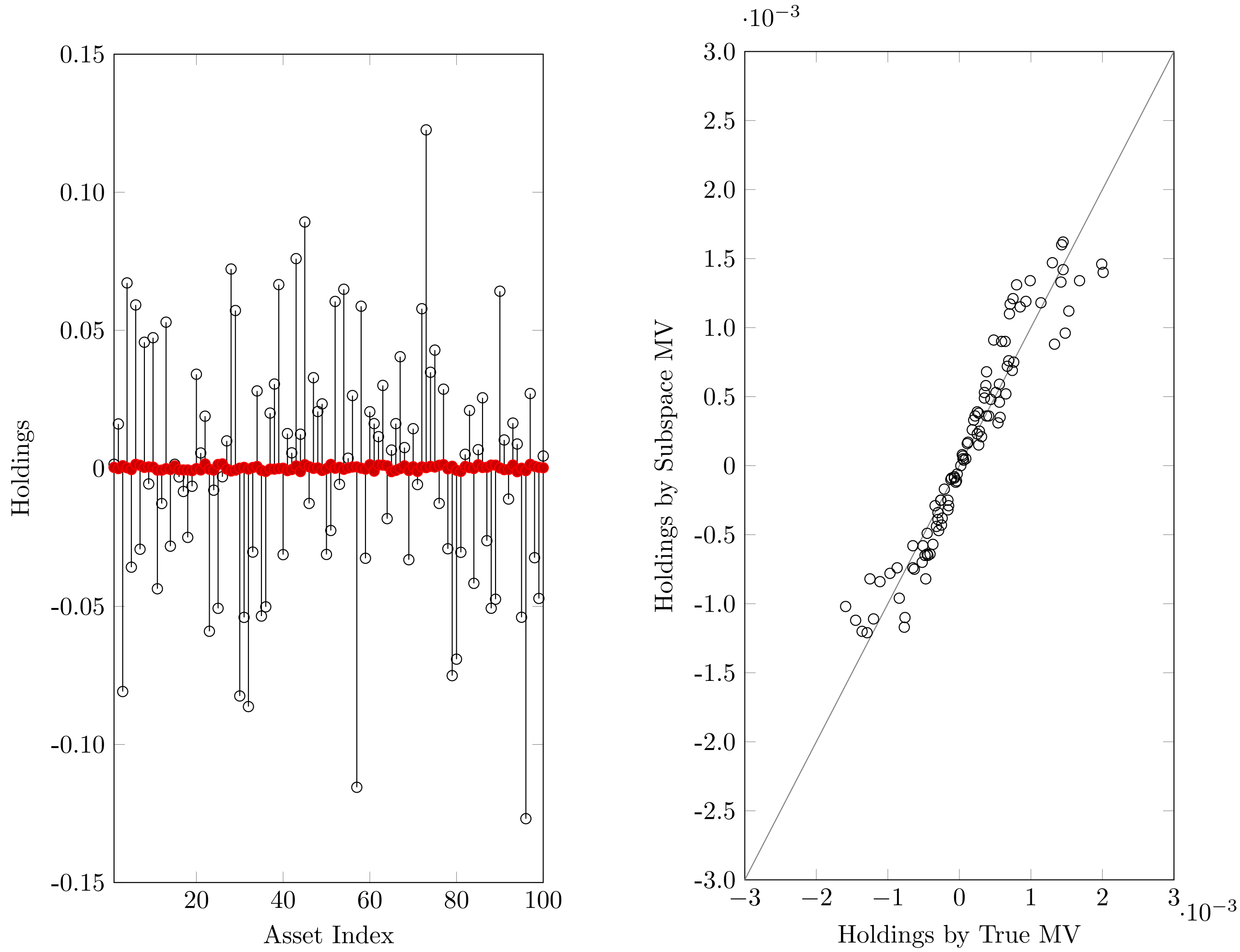


Figure 4: The plot on the left shows the holdings of the subspace mean variance portfolio (red, solid circles) compared to the holdings of the sample mean variance portfolio (black, open circles). The plot on the right shows a scatterplot of the theoretical mean-variance portfolio and the subspace mean-variance portfolio. The results are found using $T = 120$ and $N = 100$.

The left plot shows that the holdings of the subspace mean-variance portfolio are significantly smaller in scale, while also less volatile. Literature shows that portfolio rules originating from quantitative models often result in extreme and unreasonable positions for the investor (see e.g. Black and Litterman (1992)). The left plot shows that the subspace mean-variance portfolio might alleviate some of the problem since the holdings appear to be around the same values. Furthermore, the holdings of the subspace mean-variance portfolio track well with the holdings of the theoretical mean-variance portfolio, as seen in the plot on the right.

The extreme investment positions that are suggested by quantitative methods usually differ substantially between periods, which means that investors have to spend a lot of money on rebalancing their portfolios in between periods. In the simulation setting, the turnover cost C_t for changing the holdings for the next month $t+1$ can naturally be measured as in equation (21).

$$C_t = \sum_{i=1}^{N} |w_{t+1,i} - w_{t,i}| \tag{21}$$

Figure 5 shows box-plots of the turnover costs from the subspace mean-variance portfolio and the sample mean-variance portfolio for estimation windows $T = 120,\ 240$. The turnover costs are calculated over a 50-year period.

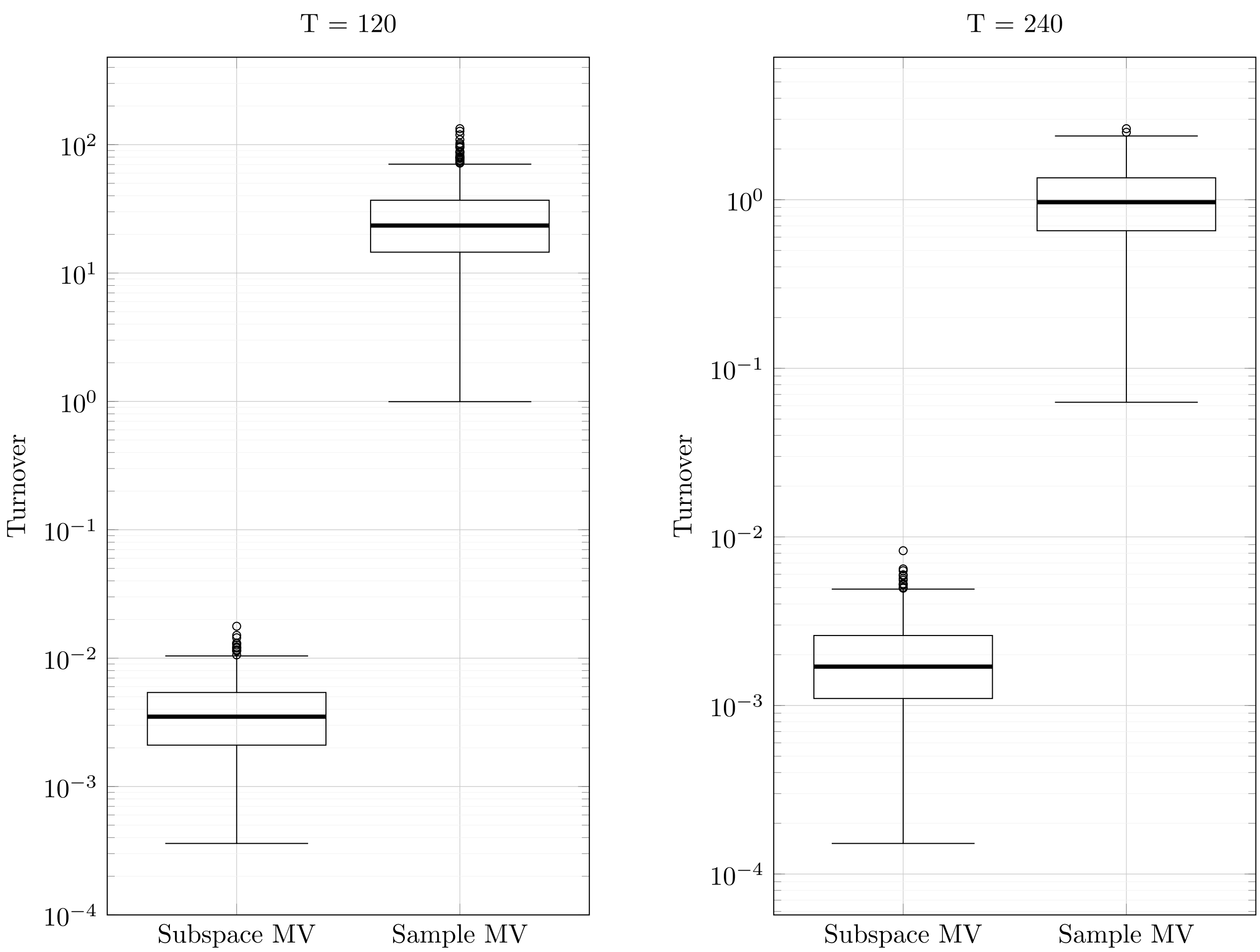


Figure 5: Box-plots are made of the turnover costs using the Subspace mean-variance portfolio and the sample mean-variance portfolio. This is done in an investment universe with $N = 100$ over a 50 year period using estimation windows $T = 120$ and 240. Furthermore, the Y-axes are in log scales.

The box plots show that the turnover costs of the subspace mean-variance portfolio is significantly lower than that of the sample mean-variance portfolio, suggesting that the investment positions of the subspace mean-variance portfolio are more stable over time.

As stated in the data section, the subspace mean-variance portfolio is also investigated using empirical returns of the Fama-French 25 (5×5) and 100 (10×10) portfolios. Figure 6 shows a plot of the Sharpe ratio achieved for different dimension sizes d, the dashed line represents the achieved Sharpe ratio where the subspace dimension is estimated according to equation (7). As the true population moments μ and Σ are unknown, the achieved Sharpe ratio is calculated as the average portfolio return divided by the sample standard deviation.

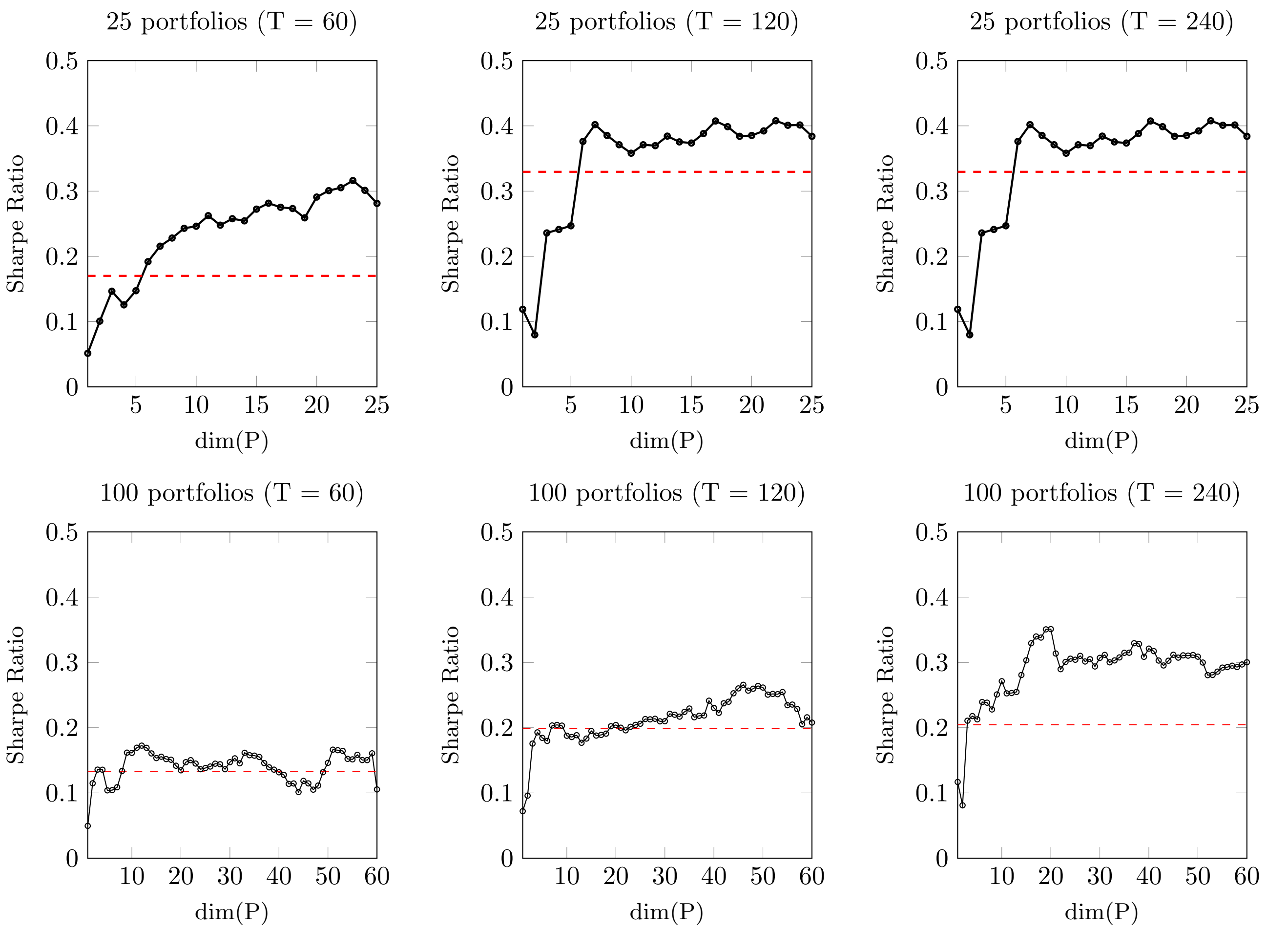


Figure 6: Plots of the Sharpe ratio at different dimension sizes for the subspace mean-variance portfolio. This is done for $T =$ 60, 120, 240 *and* $N =$ 25, 100. *The red line (dashed) denotes the Sharpe ratio of the subspace portfolio with estimated dimension sizes.*

Figure 6 shows that the subspace method works best when the dimension size is close to the true the dimension as for most plots the maximum Sharpe ratio is reached around the estimated dimension size, similar to the findings of the simulation. The estimated dimension size is usually around 7, indicating that there are 7 common factors that generate the returns. However comparing Figure 6 in to the similar Figure of Chen and Yuan (2016), the plots don't display the extreme drops at the higher dimension sizes. This indicates that the subspace mean-variance method might not be superior to the traditional sample-mean variance portfolio, since the methods should result in similar portfolios when the dimension size is close to N.

5.2 Shrinkage Estimation

This section studies the shrinkage estimators of Jorion (1986) and Ledoit and Wolf (2003). Figure 7 shows plots of the Sharpe ratio of the sample Markowitz portfolio using both shrinkage and the traditional sample estimators for different values of N. Furthermore, the Bayes-Stein estimator can't be calculated when $T < N$ and therefor the results aren't shown for these situations.

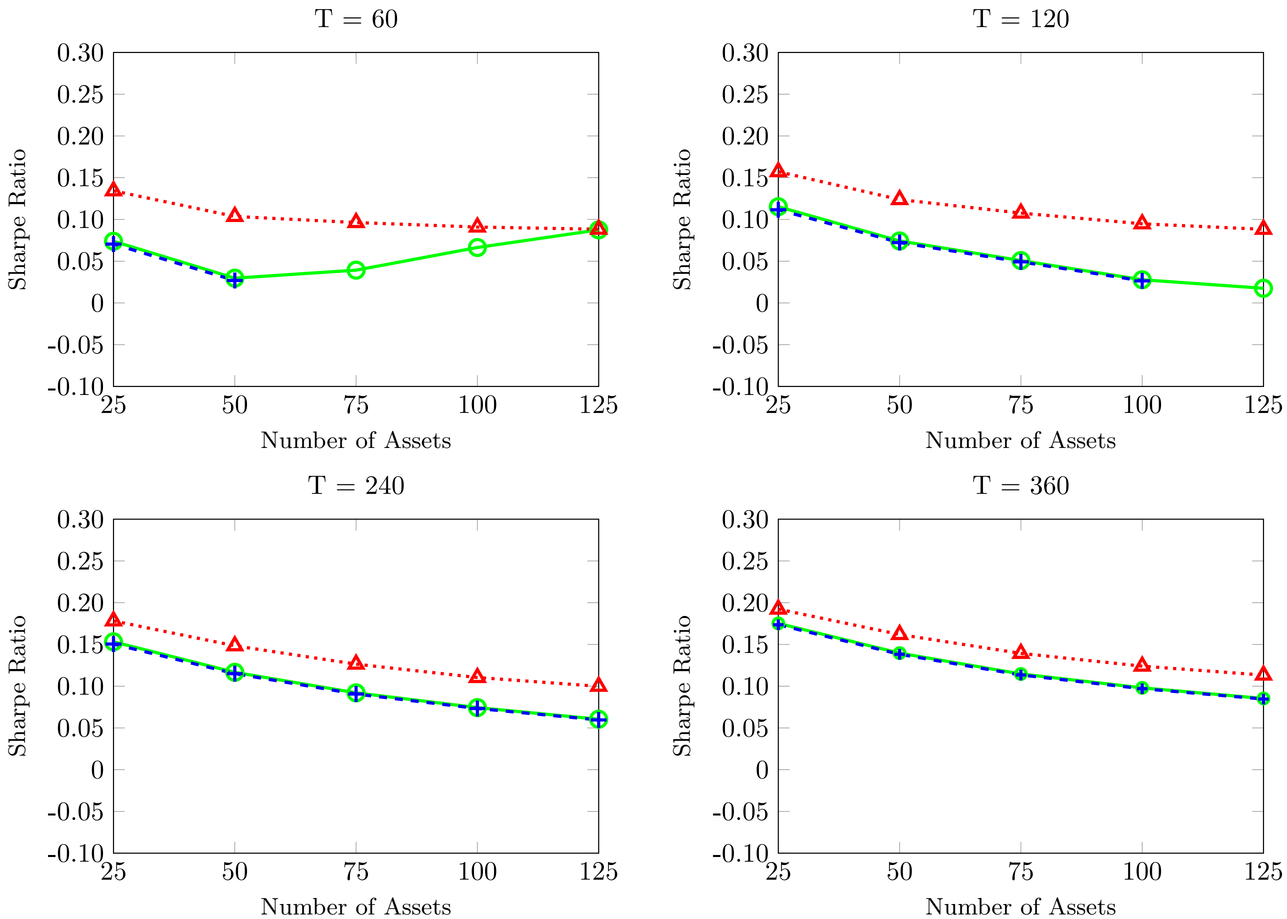


Figure 7: Plots of the Sharpe ratio for the sample Markowitz mean-variance portfolio at $N =$ 25, 50, 75, 100, 125 *using different estimator inputs. The green lines (circles) are made using the traditional sample estimators, the blue lines (crosses) are made using the mean estimator of Jorion (1986) and the red lines (triangles) are made using the covariance matrix of Ledoit and Wolf (2003).*

Figure 7 shows that of the Shrinkage estimators only the estimator of Ledoit and Wolf (2003) is able to significantly improve portfolio selection in terms of Sharpe ratio. The estimator follows from a misspecified model, as the model assumes that the returns are generated by a market return factor, but the returns are also generated by two additional factors. In the simulation setting, the shrinkage estimator is able to give an improvement in estimation power. This shows that although there is some misspecification, assuming a model structure for the returns is potentially beneficial. The Bayes-Stein estimator shows very similar estimation power to the traditional sample mean. However, this estimator requires $N < T$ to be calculated and could thus be regarded as inferior.

The holding- and turnover-plots using both shrinkage estimators can be found in the appendix. The holding plots show lower variation when calculating the sample mean-variance portfolio using the shrinkage estimator than with the traditional sample estimates. Especially for the shrunk covariance estimator, which also results in a portfolio with better tracking of the true Markowitz portfolio holdings compared to the Bayes-Stein estimator. Furthermore, the box-plots show that the shrunk covariance estimator also results in a significantly lower turnover cost, compared to both the traditional sample estimates and the Bayes-Stein estimator.

5.3 Shrinkage Estimation and Subspace Mean-Variance Portfolio combined

In this section a combination of shrinkage estimation and the subspace mean-variance portfolio is investigated. In Figure 8 the Sharpe ratio for the subspace mean-variance portfolio is plotted using both shrinkage estimators and the sample estimators at different N.

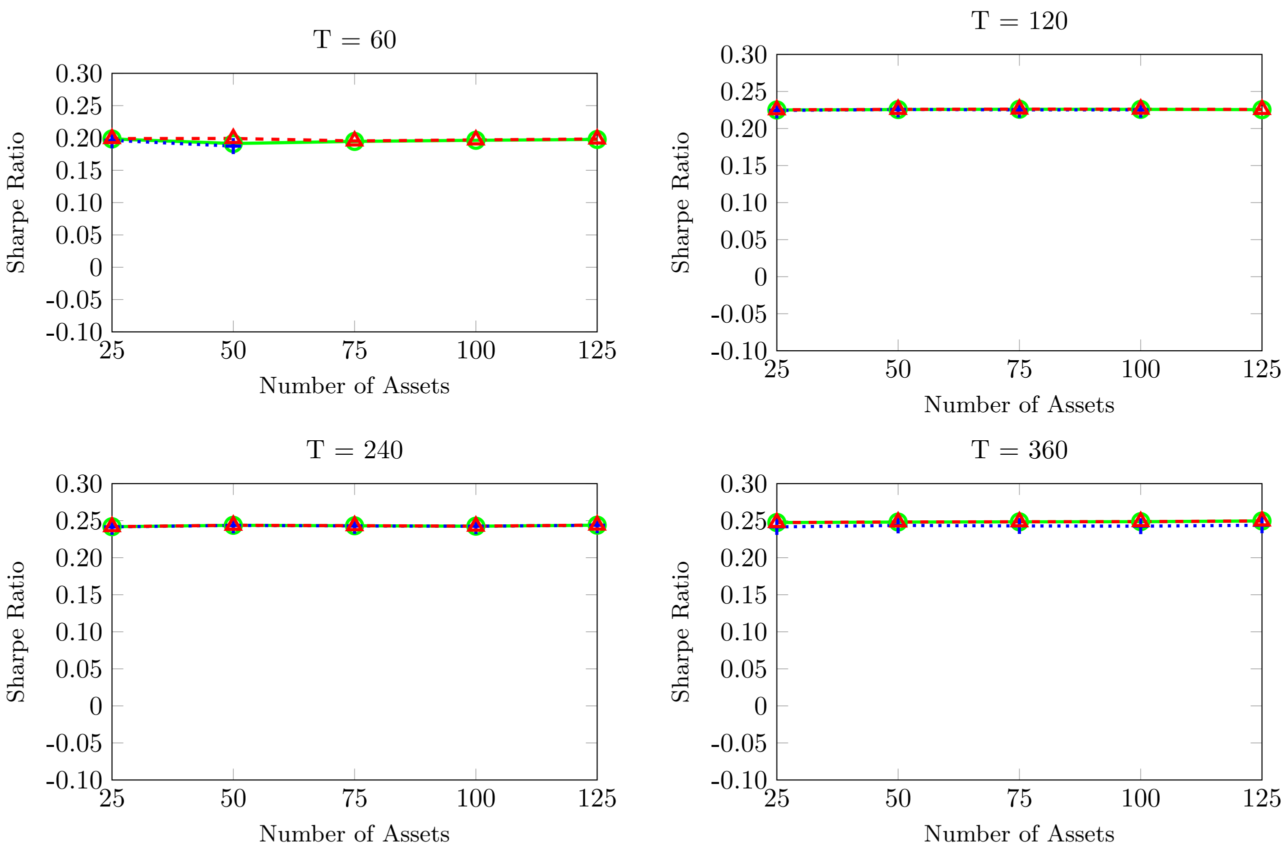


Figure 8: Plots of the Sharpe ratio for the subspace Markowitz mean-variance portfolio at $N =$ 25, 50, 75, 100, 125 *using different estimator inputs. The green lines (circles) are made using the traditional sample estimators, the blue lines (crosses) are made using the mean estimator of Jorion (1986) and the red lines (triangles) are made using the covariance matrix of Ledoit and Wolf (2003).*

Figure 7 shows that the shrinkage estimator of Ledoit and Wolf (2003) improves the sample mean-variance portfolio. However, figure 8 doesn't show a clear improvement in Sharpe ratio when using the shrinkage estimator inputs in combination with the subspace method. This result indicates that the subspace mean-variance portfolio is more robust towards the overall estimation accuracy of the estimator. The subspace mean-variance portfolio is made using the eigenvectors from the covariance, which can be considered as estimations of the common factors that generate the returns. Thus, the subspace mean-variance portfolio is only significantly improved, if the estimated covariance matrix is relatively better in capturing the common economic movement. One way to achieve this would be to assume a correctly specified factor model for the returns, whereas the model used for the shrinkage estimator is misspecified. The results of figure 8 also suggests that poor estimators will have relatively better results when combining it with the subspace mean-variance portfolio, if the estimator sufficiently captures the common movement in the returns.

Furthermore, the shrinkage estimators are investigated using the empirical returns of the Fama-French portfolios. Figure 9 shows the achieved Sharpe ratios for the subspace mean-variance portfolio calculated using the shrinkage estimators and the traditional sample estimates.

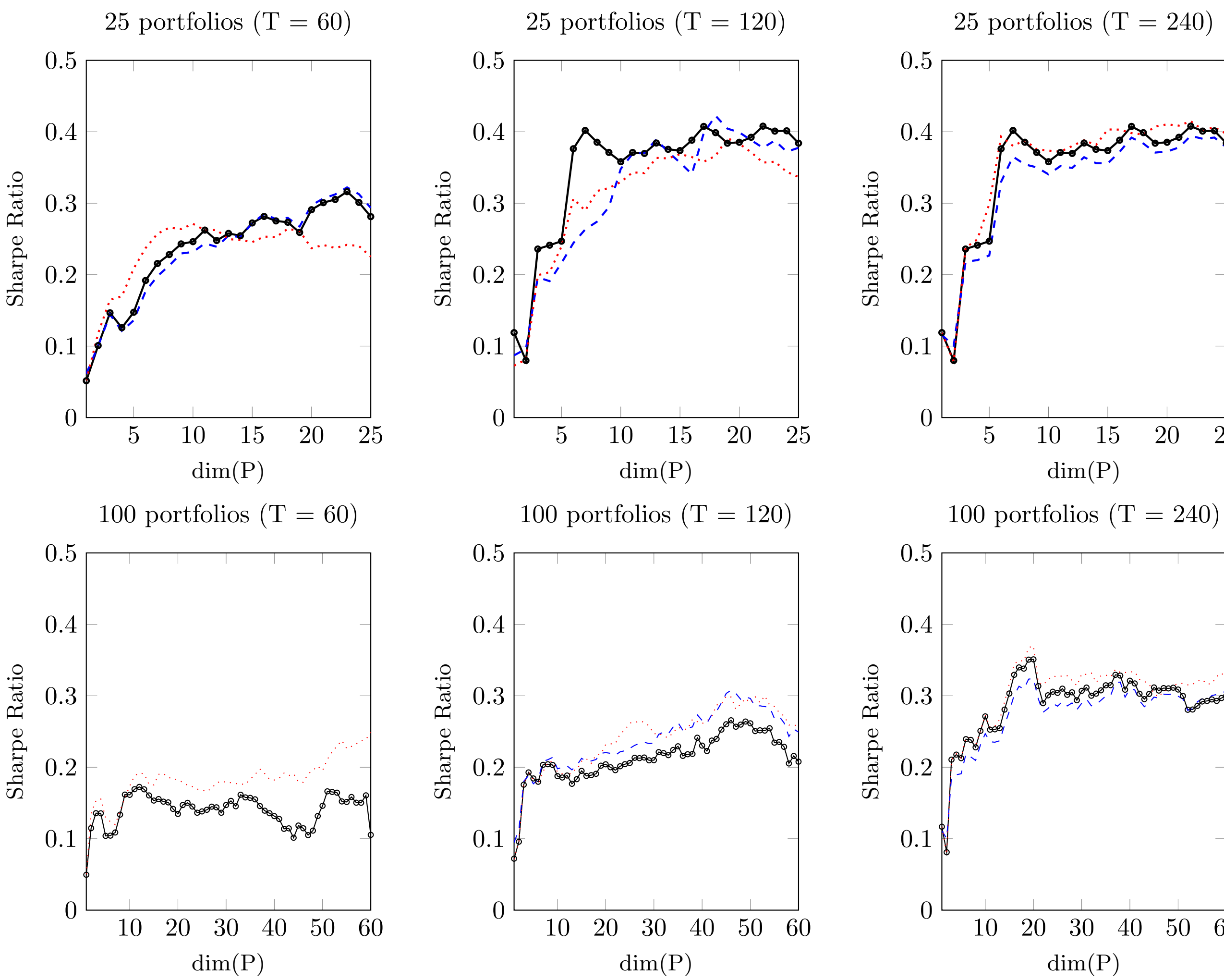


Figure 9: Plots of the Sharpe ratio at different dimension sizes for the subspace mean-variance portfolio. The black line is made using the traditional sample estimates, the blue line (dashed) uses the mean estimator of Jorion (1986) and the red line (dotted) uses the covariance estimator of Ledoit and Wolf (2003). This is done for $T =$ 60, 120, 240, $N =$ 25, 100 *and* $N < T$.

Figure 9 shows that the three different lines exhibit similar movement, such that the difference in Sharpe ratio is almost independent to the size of d. Furthermore, the shrinkage estimators don't always show improvement in estimation. However, when $N = 100$ the shrinkage estimator of Ledoit and Wolf (2003) shows higher Sharpe ratio for all T and the Bayes-Stein estimator shows improvement for $T = 120$. For $N = 25$ both estimator show similar to worse estimation power, except for $T = 60$. These findings display two important features of shrinkage estimators. First, shrinkage estimators are more robust towards larger N. Second, shrinkage estimation is perform relatively better for small T and don't benefit as much from an increased estimation window compared to the traditional sample estimates. These findings indicate that shrinkage estimation could be beneficial in settings when N is large and T is small, but could have similar to worse results otherwise. As mentioned before, the estimated dimension size is usually around 7, which means that also in the empirical setting the single index model is most likely misspecified. The subspace method in combination with shrinkage estimation could be further improved, if more advanced return models are used that are more likely to be correctly specified.

Table 1 reports the achieved Sharpe ratio of the portfolio selection methods used in this paper and three other portfolio rules that have come out of portfolio literature. These additional portfolio rules are: (KZ) the portfolio rule of Kan and Zhou (2007); (S&N) a combination rule based on the sample mean-variance portfolio and naive diversification (Tu and Zhou, 2011); and (KZ&N) a combination rule based on the KZ rule and naive diversification (Tu and Zhou, 2011). The exact implementation of the additional portfolio rules can be found in the appendix. The tables shows the achieved Sharpe ratios calculated from the empirical returns of the Fama-French portfolios.

	N = 25			N = 100		
Portfolio Rule	T = 60	T = 120	T = 240	T = 60	T = 120	T = 240
Sample MV	0.28	0.37	0.38	0.16	0.21	0.17
Naive	0.13	0.14	0.15	0.13	0.14	0.15
S&N	0.28	0.37	0.38		0.21	0.17
KZ	0.28	0.37	0.38		0.21	0.18
KZ&N	0.28	0.37	0.38		0.21	0.18
Subspace MV	0.17	0.20	0.33	0.13	0.20	0.24
SJ	0.23	0.37	0.38		0.18	0.20
SLW	0.23	0.32	0.38	0.16	0.22	0.21
Subspace MV & SJ	0.16	0.21	0.29		0.20	0.18
Subspace MV & SLW	0.22	0.24	0.31	0.15	0.20	0.17

Table 1: The table reports the achieved Sharpe ratios calculated using a selection of portfolio selection methods. A heatmap is made along the columns, where the maximum values are represented by green (light grey shade), the minimum values are represented by red (dark grey shade) and the color for the values in between is interpolated given the value. The Sharpe ratio are calculated over a period of 50 years in a market with $N = 25$, 100 and estimation windows $T = 60$, 120, 240. For certain methods the portfolio rule couldn't be calculated when $T < N$ and are left blank. Furthermore, 'SJ' stands for Shrinkage Jorion and 'SLW' for Shrinkage Ledoit and Wolf.

As seen in table 1, the subspace mean-variance portfolio isn't able to consequently beat all the other portfolio rules except for naive diversification. For $N = 25$, it even appears to be worse than other portfolio rules, however the subspace mean-variance performs better for $N = 100$ and larger T, which also where the subspace method is expected to perform best compared to the other methods. The finding supports theorem 2, which states that the subspace mean-variance portfolio is more robust towards large markets. The results in table 1 differ from what is shown in Chen and Yuan (2016), where the subspace method appeared relatively better towards the other portfolio rules, such that it should be questioned how effective the subspace mean-variance portfolio actually is.

The Sharpe ratios of the portfolio rules calculated using the shrinkage methods also are not superior to other selection methods that have appeared in literature. However, the methods do show relatively high Sharpe ratios when $N = 100$, which is also expected since shrinkage estimation is supposed to be a better estimator for high dimension multivariate normal variables. Furthermore, the combined method of shrinkage and the subspace mean-variance portfolio didn't result in a significantly improved selection method as seen from table 2.

6 Conclusion

Through simulation it is shown that both the subspace method and shrinkage estimation are beneficial towards portfolio selection in terms of Sharpe ratio. Also in the form of stable holding positions and lower turnover costs both methods could prove to be beneficial. However, the empirical study shows the limits towards the benefits of both methods in practice. It is found that for both methods only in large markets the Sharpe ratios are higher and equal to lower otherwise.

The results of Chen and Yuan (2016) are found to be difficult to replicate, such that there are some doubts towards the extent of the effectiveness of the subspace mean-variance portfolio. Of the shrinkage estimators, the estimator of Ledoit and Wolf (2003) shows the best improvement for the sample mean-variance portfolio. Since this estimator uses a misspecified model, it could be improved by applying more advanced return models. The shrinkage estimator of Jorion (1986) is found to have almost similar estimation power as the traditional sample mean. Furthermore, the combination of shrinkage estimation and the subspace mean-variance portfolio doesn't appear to be beneficial, but the subspace method could still be improved if more advanced return models are applied to calculate the subspace or the covariance estimator.

References

[1] Bai, J. and Ng, S. (2002), Determining the number of factors in approximate factor models, *Econometrica* 70(1), 191–221.

[2] Berk, J. B. and DeMarzo, P. M. (2014), *Corporate finance*, Pearson Education.

[3] Best, M. J. and Grauer, R. R. (1991), On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results, *Review of Financial Studies* 4(2), 315–342.

[4] Black, F. and Litterman, R. (1992), Global portfolio optimization, *Financial analysts journal* 48(5), 28–43.

[5] Brandt, M. (2010), *Handbook of Financial Econometrics*, Vol. 1, Elsevier, North-Holland, chapter 5: Portfolio Choice Problems, pp. 269–336. Y. Ait-Sahalia and L. P. Hansen (eds.).

[6] Chan, L. K., Karceski, J. and Lakonishok, J. (1999), On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* 12(5), 937–974.

[7] Chen, J. and Yuan, M. (2016), Efficient portfolio selection in a large market, *Journal of Financial Econometrics* 14(3), 496–524.

[8] Connor, G. and Korajczyk, R. A. (1988), Risk and return in an equilibrium apt: Application of a new test methodology, *Journal of Financial Economics* 21(2), 255–289.

[9] DeMiguel, V., Garlappi, L. and Uppal, R. (2009), Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* 22(5), 1915–1953.

[10] Fama, E. F. and French, K. R. (1993), Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33(1), 3–56.

[11] Frost, P. A. and Savarino, J. E. (1986), An empirical bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21(03), 293–305.

[12] Haugen, R. A. (1986), *Modern investment theory*, Prentice Hall.

[13] James, W. and Stein, C. (1961), Estimation with quadratic loss, *in* 'Proceedings of the fourth Berkeley symposium on mathematical statistics and probability', pp. 361–379.

[14] Jobson, J. D. and Korkie, B. (1980), Estimation for markowitz efficient portfolios, *Journal of the American Statistical Association* 75(371), 544–554.

[15] Jorion, P. (1986), Bayes-stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21(03), 279–292.

[16] Kan, R. and Zhou, G. (2007), Optimal portfolio choice with parameter uncertainty, *Journal of Financial and Quantitative Analysis* 42(03), 621–656.

[17] Ledoit, O. and Wolf, M. (2003), Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of empirical finance* 10(5), 603–621.

[18] Markowitz, H. (1952), Portfolio selection, *The journal of finance* 7(1), 77–91.

[19] Michaud, R. O. (1989), The markowitz optimization enigma: Is optimized optimal?, *ICFA Continuing Education Series* 1989(4), 43–54.

[20] Sharpe, W. F. (1963), A simplified model for portfolio analysis, *Management science* 9(2), 277–293.

[21] Sharpe, W. F. (1966), Mutual fund performance, *The Journal of business* 39(1), 119–138.

[22] Stein, C. (1956), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *in* 'Proceedings of the Third Berkeley symposium on mathematical statistics and probability', pp. 197–206.

[23] Tasca, P. and Battiston, S. (2011), Diversification and financial stability.

[24] Tu, J. and Zhou, G. (2011), Markowitz meets talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics* 99(1), 204–215.

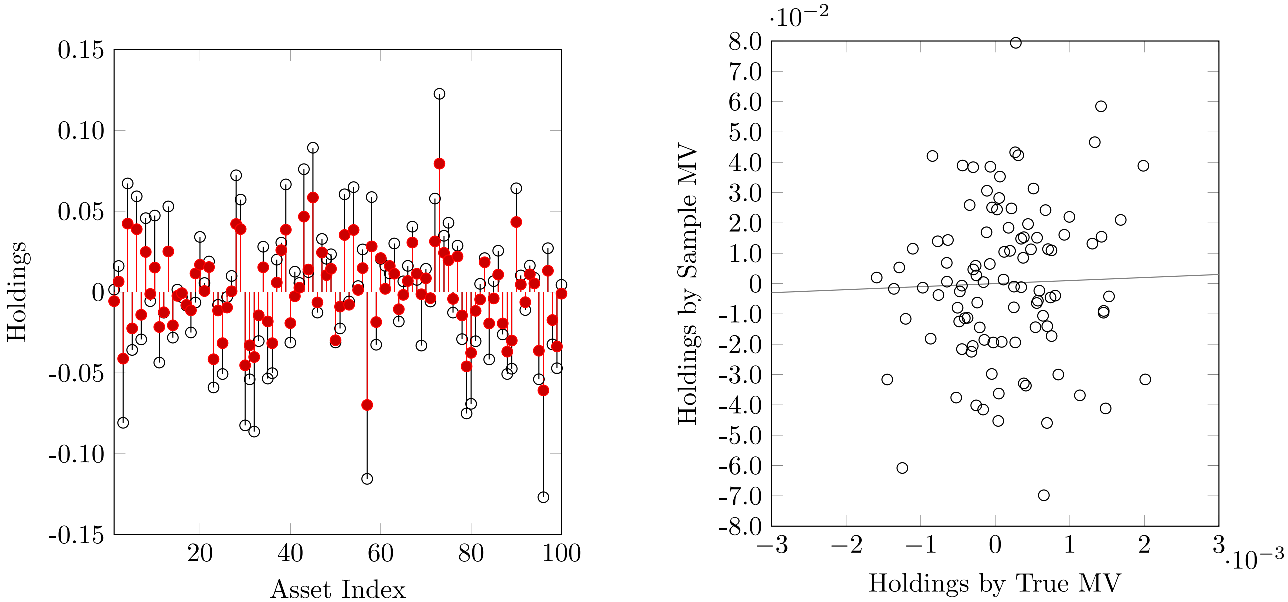
7 Appendix

A Additional Figures

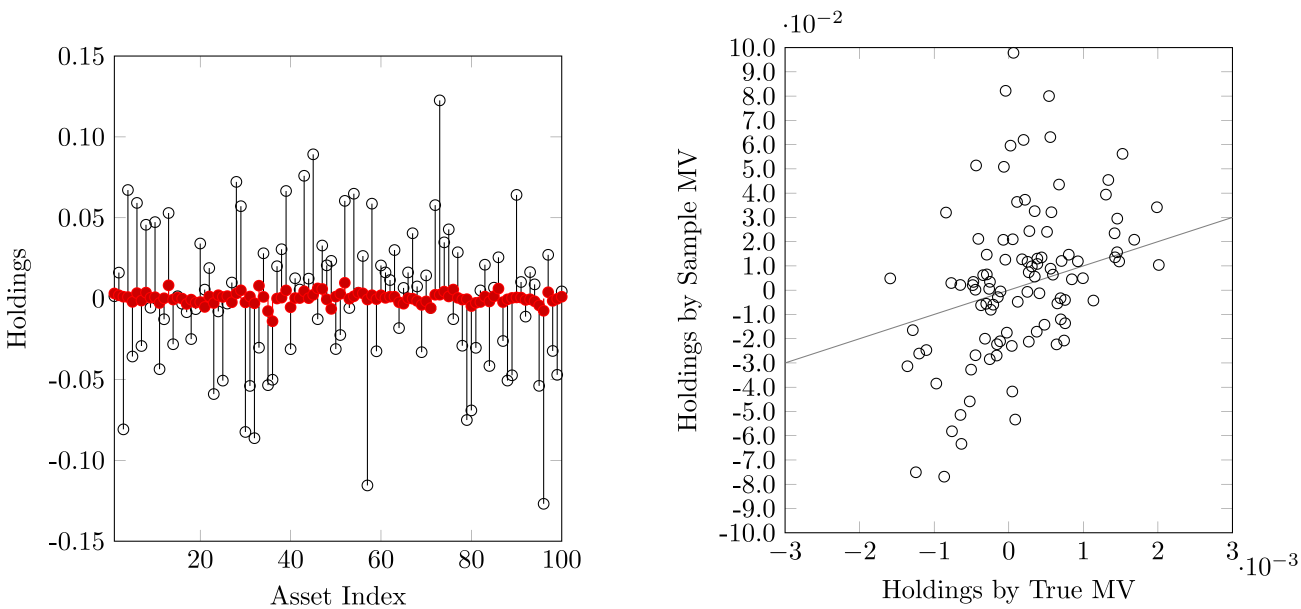
Holding plots

The holdings of the sample mean-variance portfolio calculated using shrinkage estimators of Jorion (1986) and Ledoit and Wolf (2003) are shown in the following figures.

Shrinkage estimator of Jorion (1986)



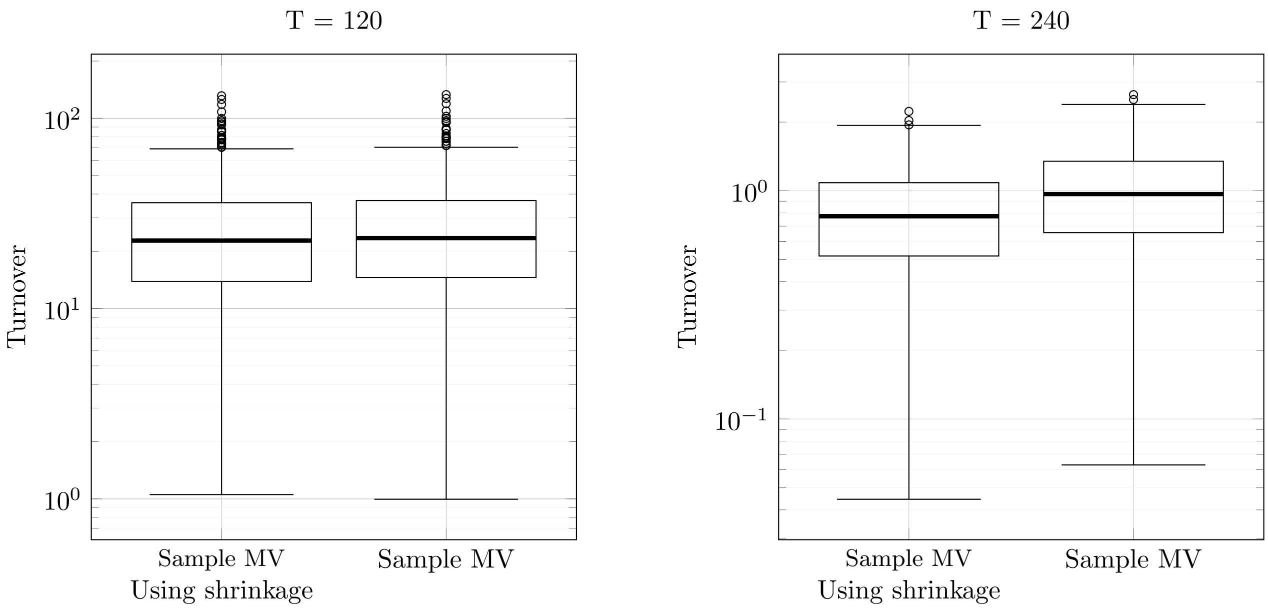
Shrinkage estimator of Ledoit and Wolf (2003)



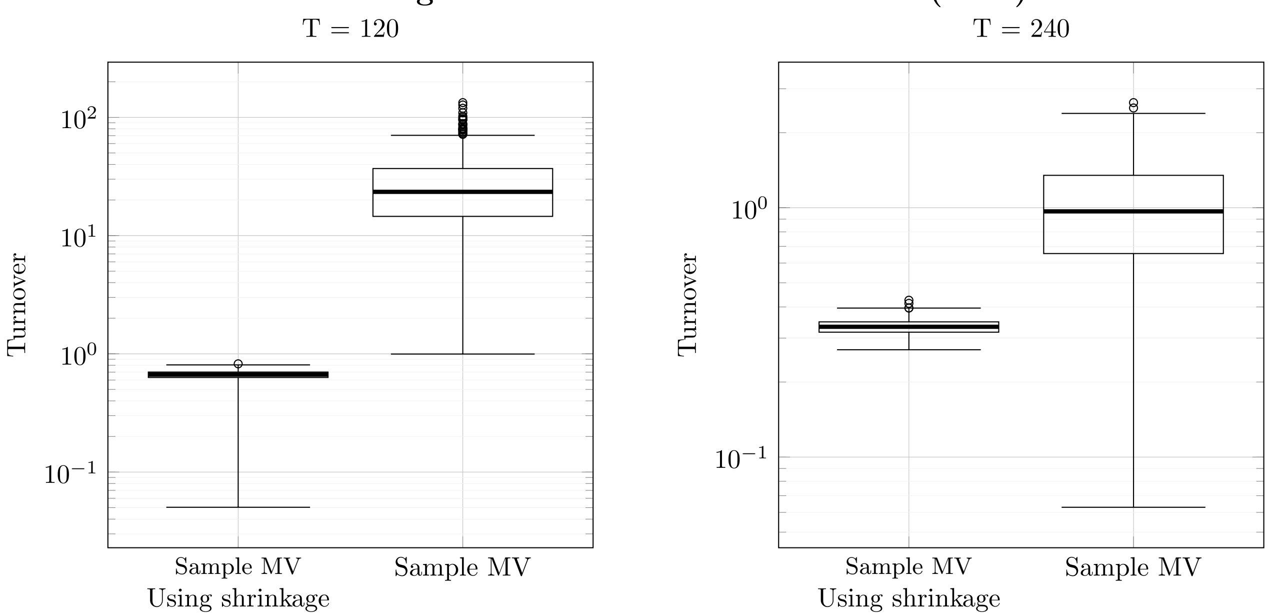
Turnover plots

The turnover costs of the sample mean-variance portfolio calculated using shrinkage estimators of Jorion (1986) and Ledoit and Wolf (2003) are shown in the following figures.

Shrinkage estimator of Jorion (1986)



Shrinkage estimator of Ledoit and Wolf (2003)



B Assumptions

As stated in the paper of Chen and Yuan (2016), the methods and theorems of this paper rely on the following set of assumptions on the factor model that generates the asset returns.

Assumption 1. *The factors have finite fourth moments such that there exists a positive constant $C_1 < \infty$ satistfying:*

$$\max_{1 \le k \le K} \mathbb{E} f_{kt}^4 \le C_1$$

Assumption 2. *There is a strictly positive definite matrix Σ_B such that:*

$$B^\intercal B/N \to \Sigma_B \qquad \text{as} \;\; N \to \infty$$

Assumption 3. *The covariance matrix of the error terms $(\varepsilon_{1t}, \dots, \varepsilon_{Nt})^\intercal$, Σ_ε, has eigenvalues which are bounded away from both zero and infinity. Moreover, the idiosyncratic risks have finite fourth moments such that there exists a positive constant $C_2 < \infty$ satisfying:*

$$\max_{1 \le j \le N} \mathbb{E}\, \varepsilon_{jt}^4 \le C_2$$

Assumption 4. *The factors $f_1, \dots, f_t$ are weakly dependent in that there exists a positive constant $C_3 < \infty$ satisfying:*

$$\max_{1 \le k_1, k_2 \le K} \mathbb{E}\bigg\{ \sum_{t=1}^{T} (f_{k_1 t} f_{k_2 t} - \Sigma^f_{k_1 k_2})^2 \bigg\} \le C_3 T$$

where $\Sigma^f_{k_1 k_2}$ is the (k_1, k_2)th entry of Σ_f.

Assumption 5. *The idiosyncratic risks $(\varepsilon_{11}, \dots, \varepsilon_{N1})^\intercal, \dots, (\varepsilon_{1t}, \dots, \varepsilon_{Nt})^\intercal$ are weakly dependent in that there exists a positive constant $C_4 < \infty$ satisfying:*

$$\max_{1 \le j_1, j_2 \le K} \mathbb{E}\bigg\{ \sum_{t=1}^{T} (\varepsilon_{j_1 t} \varepsilon_{j_2 t} - \Sigma^\varepsilon_{j_1 j_2})^2 \bigg\} \le C_4 T$$

where $\Sigma^\varepsilon_{j_1 j_2}$ is the (j_1, j_2)th entry of Σ_ε.

Assumption 6. *The factors and idiosyncratic risks are jointly weakly dependent in that ther exists a positive constant $C_5 < \infty$ satisfying:*

$$\max_{\substack{1 \le j \le N \\ 1 \le k \le K}} \mathbb{E}\bigg(\sum_{t=1}^{T} \varepsilon_{jt}^2 f_{kt}^2 \bigg) \le C_5 T$$

It has been shown that these assumptions generally hold for factor models that have been proposed in literature (Bai and Ng, 2002).

C Proof of Markowitz Weights

The minimization problem of Markowitz can be represented by the following equation,

$$\min_{w\in\mathbb{R}^N}\{\frac{\gamma}{2}w^\intercal\Sigma w - w^\intercal\mu\}$$

Then the differentiating this equation with respect to w and equating the differential to zero will give the solution,

$$\frac{d}{dw}(\frac{\gamma}{2}w^\intercal\Sigma w - w^\intercal\mu) = 0$$
$$\gamma\Sigma w - \mu = 0$$
$$w = \frac{1}{\gamma}\Sigma^{-1}\mu$$

Proof of Markowitz Weights with restriction

A common restriction for portfolio weights is that they add up to one, then the optimization problem is as follows,

$$\min_{w\in\mathbb{R}^N}\{\frac{\gamma}{2}w^\intercal\Sigma w - w^\intercal\mu\}$$

s.t.

$$w^\intercal\mathbf{1} = 1$$

Where $\mathbf{1}$ is a $N\times 1$ vector where each element is a 1. This problem can be solved by applying Lagrange multipliers,

$$\mathcal{L}(w,\lambda) = \frac{\gamma}{2}w^\intercal\Sigma w - w^\intercal\mu - \lambda(w'\mathbf{1} - 1)$$

Which gives,

$$\frac{d\mathcal{L}}{dw} = \gamma\Sigma w - \mu - \lambda\mathbf{1} = 0$$
$$\frac{d\mathcal{L}}{d\lambda} = w^\intercal\mathbf{1} - 1 = 0$$

Solving the upper equation for w gives,

$$w = \frac{1}{\gamma}\Sigma^{-1}\mu + \frac{\lambda}{\gamma}\Sigma^{-1}\mathbf{1}$$

This solution can be substituted into $\frac{d\mathcal{L}}{d\lambda}$ for a solution for the multiplier λ, which gives the optimal weights that add up to one.

$$w^\intercal \mathbf{1} = 1$$

$$(\frac{1}{\gamma}\Sigma^{-1}\mu + \frac{\lambda}{\gamma}\Sigma^{-1}\mathbf{1})^\intercal \mathbf{1} = 1$$

$$\mu^\intercal \Sigma^{-1}\mathbf{1} + \lambda \mathbf{1}^\intercal \Sigma^{-1}\mathbf{1} = \gamma$$

$$\lambda = \frac{\gamma - \mu^\intercal \Sigma^{-1}\mathbf{1}}{\mathbf{1}^\intercal \Sigma^{-1}\mathbf{1}}$$

Using this λ the optimal portfolio weights can be found,

$$w_{opt} = \frac{1}{\gamma}\Sigma^{-1}\mu + \frac{1}{\gamma}(\frac{\gamma - \mu^\intercal \Sigma^{-1}\mathbf{1}}{\mathbf{1}^\intercal \Sigma^{-1}\mathbf{1}})\Sigma^{-1}\mathbf{1}$$

D Proof of Subspace Mean-Variance Weights

The mapping $\mathbb{R}^d \to \mathcal{P} : x \to P_\mathcal{P} x$ is bijective. Let $x_\mathcal{P} \in \mathbb{R}^d$ be such that $w^\mathcal{P}(x) = P_\mathcal{P} x$ solves the optimization problem,

$$\min_{w \in \mathcal{P}} \{\frac{\gamma}{2} w^\intercal \Sigma w - w^\intercal \mu\}$$

Which can be rewritten to the following problem which is solved by $x_\mathcal{P}$,

$$\min_{x \in \mathbb{R}^d} \{\frac{\gamma}{2} x^\intercal P_\mathcal{P}^\intercal \Sigma P_\mathcal{P} x - x^\intercal P_\mathcal{P}^\intercal \mu\}$$

This problem is similar to the traditional Markowitz framework, however the mean and variance are scaled by $P_\mathcal{P}$ giving $P_\mathcal{P}^\intercal \Sigma P_\mathcal{P}$ and $P_\mathcal{P}^\intercal \mu$ for the mean and variance respectively. The solution of this problem is known and has the following form,

$$x_\mathcal{P} = \frac{1}{\gamma}(P_\mathcal{P}^\intercal \Sigma P_\mathcal{P})^{-1} P_\mathcal{P}^\intercal \mu$$

Plugging this equation into $w^\mathcal{P}(x)$ gives the solution,

$$w_{mv}^\mathcal{P} = \frac{1}{\gamma} P_\mathcal{P}(P_\mathcal{P}^\intercal \Sigma P_\mathcal{P})^{-1} P_\mathcal{P}^\intercal \mu$$

Deriviation of the Subspace Mean-Variance Weights using the Eigenvector Subspace

The general solution to the Subspace mean variance portfolio is given below,

$$w_{mv}^\mathcal{P} = \frac{1}{\gamma} P_\mathcal{P}(P_\mathcal{P}^\intercal \Sigma P_\mathcal{P})^{-1} P_\mathcal{P}^\intercal \mu$$

If the subspace $\mathcal{P}$ is chosen to be the subspace spanned by the first d eigenvectors of the covariance matrix, then the orthonormal base of $\mathcal{P}$ is given by $P_\mathcal{P} = [\eta_1, \ldots, \eta_d]$, with η_i being i-th eigenvector of Σ. As the covariance matrix is a symmetric matrix it is known that such an orthonormal base exists, because the eigenvectors are orthogonal and can be constructed to be unit length. Thus for the eigenvectors the following properties hold: $\eta_i^\intercal \eta_j = 0$ when $i \neq j$; $\eta_i^\intercal \eta_j = 1$ when $i = j$; and $\Sigma\eta_i = \theta_i \eta_i$ for any eigenvector. Then the general solution can be rewritten as follows,

$$P_{\mathcal{P}}^\intercal \Sigma P_{\mathcal{P}} = [\eta_1, \ldots, \eta_d]^\intercal \Sigma \, [\eta_1, \ldots, \eta_d]$$

$$= \begin{bmatrix} \eta_1^\intercal \\ \vdots \\ \eta_d^\intercal \end{bmatrix} [\theta_1 \eta_1, \ldots, \theta_d \eta_d]$$

$$= \begin{pmatrix} \theta_1 & & & \\ & \theta_2 & & \\ & & \ddots & \\ & & & \theta_d \end{pmatrix}$$

And thus,

$$(P_{\mathcal{P}}^\intercal \Sigma P_{\mathcal{P}})^{-1} = \begin{pmatrix} \theta_1^{-1} & & & \\ & \theta_2^{-1} & & \\ & & \ddots & \\ & & & \theta_{d^{-1}} \end{pmatrix}$$

This equation can be plugged back into the general solution for the eqaution of the subspace mean-variance portfolio shown below,

$$w_{mv}^{\mathcal{P}} = \frac{1}{\gamma} \sum_{k=1}^{d} \theta_k^{-1} \eta_k \eta_k^\intercal \mu$$

E Shrinkage Factor of Ledoit and Wolf (2003)

The optimal shrinkage factor of Ledoit and Wolf (2003) has the following form,

$$\tau^* = \frac{\sum_{i=1}^N \sum_{j=1}^N \mathrm{Var}(s_{ij}) - \mathrm{Cov}(f_{ij}, s_{ij})}{\sum_{i=1}^N \sum_{j=1}^N \mathrm{Var}(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2}$$

Ledoit and Wolf (2003) show that the optimal shrinkage factor satisfies the following equation,

$$\tau^* = \frac{1}{T} \frac{\pi - \rho}{\kappa} + O_p(\frac{1}{T^2})$$

Where $\pi = \sum_{i=1}^N \sum_{j=1}^N \mathrm{AsyVar}[\sqrt{T} s_{ij}]$, $\rho = \sum_{i=1}^N \sum_{j=1}^N \mathrm{AsyVar}[\sqrt{T} f_{ij}, \sqrt{T} s_{ij}]$ and $\kappa = \sum_{i=1}^N \sum_{j=1}^N (\phi_{ij} - \sigma_{ij})^2$.

Such that the optimal τ^* is of the form $\frac{1}{T}\frac{\pi-\rho}{\kappa}$. Since this equation has unobserved terms, the factor has to be estimated from data. For this purpose, the following decompositions are made:

$$\pi = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_{ij} \qquad \rho = \sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{ij} \qquad \kappa = \sum_{i=1}^{N}\sum_{j=1}^{N}\kappa_{ij}$$

The following set of equations for $\hat{\pi}_{ij}$, $\hat{\rho}_{ij}$ and $\hat{\kappa}_{ij}$ result in consistent estimators for π_{ij}, ρ_{ij} and κ_{ij} respectively.

$$\begin{aligned}
\hat{\pi}_{ij} &= \frac{1}{T}\sum_{t=1}^{T}\left[(r_{it}-\hat{\mu}_i)(r_{jt}-\hat{\mu}_j)-s_{ij}\right]^2 \\
\hat{\rho}_{ij} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\rho}_{ijt} \\
\hat{\rho}_{ijt} &= \frac{s_{jm}s_m(r_{it}-\hat{\mu}_i)+s_{im}s_m(r_{jt}-\hat{\mu}_j)-s_{im}s_{jm}(r_{mt}-\hat{\mu}_m)}{s_m^2}(r_m-\hat{\mu}_m)(r_{it}-\hat{\mu}_i)(r_{jt}-\hat{\mu}_j)-f_{ij}s_{ij} \\
\hat{\kappa}_{ij} &= \left(f_{ij}-s_{ij}\right)^2
\end{aligned}$$

Where r_{it} is the return on stock i at time t and $\hat{\mu}_i$ is the average return on stock i. Furthermore, s_m is the standard deviation of the market return, $\hat{\mu}_m$ is the average market return and s_{im} is the covariance between the market return and stock i.

F Other Portfolio Rules

- Naive Diversification:

$$w_N = \frac{1}{N}\mathbf{1}$$

- KZ: Portfolio rule of Kan and Zhou (2007):

$$\hat{w}_{KZ} = \frac{1}{\gamma}\frac{T-N-2}{c(T-1)}\left[\eta\hat{\Sigma}^{-1}\hat{\mu}+(1-\eta)\hat{\mu}_g\hat{\Sigma}^{-1}\mathbf{1}\right]$$

  where,

$$\eta = \frac{\psi^2}{\psi^2+N/T}, \qquad \psi^2 = (\hat{\mu}-\hat{\mu}_g\mathbf{1})^{\intercal}\hat{\Sigma}^{-1}(\hat{\mu}-\hat{\mu}_g\mathbf{1})$$

- S&N: Combination Rule based on sample mean-variance portfolio and the naive portfolio (Tu and Zhou, 2011) :

$$\hat{w}_{S\&N} = (1-\delta)w_N + \delta\frac{T-N-2}{T-1}\hat{w}_{mv}$$

where,

$$\delta = \frac{\pi_1}{\pi_1 + \pi_2}$$

$$\pi_1 = w_N^\intercal \hat{\Sigma} w_N - \frac{2}{\gamma} w_N^\intercal \hat{\mu} + \frac{1}{\gamma^2}\tilde{\theta}^2$$

$$\pi_2 = \frac{1}{\gamma^2}\left[\frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)} - 1\right]\tilde{\theta}^2 + \frac{(T-2)(T-N-2)}{\gamma^2(T-N-1)(T-N-4)}\frac{N}{T}$$

$$\tilde{\theta}^2 = \frac{(T-N-2)\theta^2 - N}{T} + \frac{2\theta^N(1+\theta^2)^{-(T-2)/2}}{T\int_0^{\theta^2/(1+\theta^2)} x^{N/2}(1-x)^{(T-N-2)/2}dx}$$

$$\theta^2 = \hat{\mu}^\intercal \hat{\Sigma}^{-1}\hat{\mu}$$

- KZ&N: Combination Rule based on the portfolio rule of Kan and Zhou (2007) and naive diversification (Tu and Zhou, 2011):

$$\hat{w}_{KZ\&N} = (1-\zeta)w_N + \zeta w_{KZ}$$

where,

$$\zeta = \frac{\pi_1 - \pi_{13}}{\pi_1 - 2\pi_{13} + \pi_3}$$

$$\pi_{13} = \frac{1}{\gamma^2}\tilde{\theta}^2 - \frac{1}{\gamma} w_N^\intercal \hat{\mu} + \frac{(T-N-1)(T-N-4)}{\gamma(T-2)(T-N-2)}\left[\eta w_N^\intercal \hat{\mu} + (1-\eta)\hat{\mu}_g w_N^\intercal \mathbf{1}\right]$$
$$- \frac{(T-N-1)(T-N-4)}{\gamma^2(T-2)(T-N-2)}\left[\eta \hat{\mu}^\intercal \tilde{\Sigma}^{-1}\hat{\mu} + (1-\eta)\hat{\mu}_g \hat{\mu}^\intercal \tilde{\Sigma}^{-1}\mathbf{1}\right.$$

$$\pi_3 = \frac{1}{\gamma^2}\tilde{\theta}^2 - \frac{(T-N-1)(T-N-4)}{\gamma^2(T-2)(T-N-2)}\left(\tilde{\theta}^2 - \frac{N}{T}\eta\right)$$