Stockholms universitet

A comparative simulation study of two enhancements to portfolio optimization: Shrinkage and clustering

Johannes Pohjolainen

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2022:7
Matematisk statistik
Juni 2022

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2022:7**
http://www.math.su.se

A comparative simulation study of two enhancements to portfolio optimization: Shrinkage and clustering

Johannes Pohjolainen*

June 2022

Abstract

Through simulation, this thesis investigates two enhancements to minimum vari- ance portfolio estimation by comparing the methods presented in De Prado (2016) with Bodnar, Parolya, and Schmid (2018). A shrinkage methodology is presented from Bod- nar, Parolya, and Schmid (2018), while a clustering methodology - Hierarchical Risk Parity (HRP) - from De Prado (2016) is presented. The estimators were evaluated in experiments to investigate the effect of number of observations, number of assets, vari- ance, condition number and correlation. The results show that the method of Bodnar, Parolya, and Schmid (2018) performs relatively consistent while HRP generates higher errors with higher spreads. The errors of HRP show a strong relationship with asset correlation.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: theworkingbee@hotmail.com. Supervisor: Taras Bodnar.

Contents

Acknowledgements

I would like to thank my supervisor Taras Bodnar for all the interesting discussions and support during the writing of this thesis.

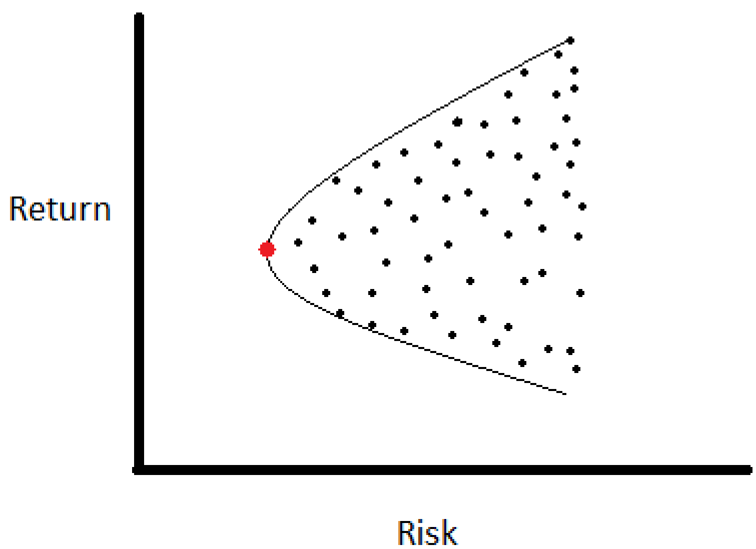
1 Introduction

Financial investments affect us all, as how our pensions and private savings are invested can change the quality of our lives. The great depression, the financial crash of 2008 and the recent market downturn during the covid-19 pandemic are all examples of how market changes can affect us. Given the impact investments can have, a lot of strategies have been developed to protect ourselves. One such strategy is to reduce risk.

1.1 Background

When investing, one needs to choose securities to include in ones portfolio and how much to allocate/invest in each. In 1952 Markowitz introduced concepts concerning the allocation step (Markowitz 1952). Depending on how much is invested in each asset, some portfolios can have the same returns although they take on different levels of risk. In finance risk is often defined as variance and Markowiz's results relate to the shape that is prevalent when plotting the returns of different portfolio allocations against their standard deviation, as displayed in Figure 1 below.

Figure 1: Relationship between expected return and risk



The red tip of the curve in Figure 1 inspired the topic of this thesis, namely estimating the portfolio with the minimum variance. Observing this curve, it is less likely that other portfolios offers better returns for a lower risk compared to the one with the minimum variance. If one as an example has decided to invest in Apple, Microsoft and Tesla, this theory advises you to divide your investment between these in the way that results in the least variability in your portfolio's worth.

To estimate the minimum variance portfolio, historical returns are often used. As an example, the returns of the three stocks above can be collected in a matrix as follows:

Table 1: Example returns for three days

	Apple	Microsoft	Tesla
Day 1	0.03	0.04	0.01
Day 2	0.06	0.02	0.03
Day 3	0.09	0.01	0.04

These returns can be seen as observations of random variables $r = (R_{Apple}, R_{Microsoft}, R_{Tesla})^t$, with the covariance matrix Σ. Using this notation, we can describe our investments as the proportion of our wealth we invest in each asset. If we as an example invest 35% in Apple, 20% in Microsoft and 45% in Tesla, we can denote our portfolio as

$$R = 0.35 \times R_{Apple} + 0.2 \times R_{Microsoft} + 0.45 \times R_{Tesla}$$

Using matrix notation, we can generalize the portfolio return as $w^t * r$, which finally gives us this following expression for portfolio variance $w^t \Sigma w$. It is this variance that is to be minimized to estimate the weights which result in the minimum variance, and the optimization problem is as follows:

$$min_w(w^t \Sigma w), \sum_i w_i = 1$$

This allows us to conclude this background section with the analytical solution for the minimum variance portfolio, i.e. the red point in Figure 1:

$$w_{traditional} = \frac{\Sigma^{-1} 1_{p\times 1}}{1_{p\times 1}^t \Sigma^{-1} 1_{p\times 1}}$$

Since the true covariance matrix Σ is not known in applications, it is often replaced with an estimated version $\hat{\Sigma}$ based on historical data for estimating the weights $\hat{w}_{traditional}$.

1.2 Problem statement and research questions

Having to use an estimate $\hat{\Sigma}$ leaves our optimization vulnerable to estimation errors and hence error maximization. Optimization problems are influenced the most by extreme values, which are the most likely to involve error (Michaud 1989). Hence, extreme values can cause our solution to display a sub-optimal portfolio. Financial covariance matrices are ill conditioned due to empirical eigenvalues being close to zero (Laloux et al. 1999), which makes this problem likely.

One further limitation is that inversions of matrices are increasingly sensitive to errors as dimensions increase, making the optimization setup vulnerable to errors too. Combining the prevalence of ill conditioned input with a sensitive calculation results in error magnifications where even small estimation errors can generate false optimal portfolios (Michaud 1989; De Prado 2016). This suggests that improvements can be made.

In a simulation study, this thesis will investigate how shrinkage and clustering for inverse variance portfolios can be used to improve the estimation of the minimum variance portfolio by comparing the methods suggested in Bodnar, Parolya, and Schmid (2018) and De Prado (2016) respectively. The following research questions will be investigated:

- Which method produces the smallest out of sample variance error generally?
- How do the market parameters correlation, variance and conditioning (condition number) affect the out of sample variance?
- How do the choices of number of assets and observations affect the out of sample variance?

To answer these questions, additionally a methodology for backtesting that aims to simulate a wide range of market conditions will be presented.

1.3 Structure of this thesis

This thesis is structured to go through previous research and the estimation methods to be compared in section 2. In section 3, this thesis' simulation study will be described and it will be performed in section 4. Section 5 will conclude our findings.

2 Theory and previous research

2.1 Literature study

At the time of writing this thesis, no paper comparing the estimators from Bodnar, Parolya, and Schmid (2018) to De Prado (2016) has been found.

However, there has been a multitude of research and advances in portfolio optimization (Sun et al. 2019). Several new estimators have been developed, and compared both empirically and in simulations. Nevertheless as new research and estimators are produced, opinions differ on the performance of different estimators (DeMiguel, Garlappi, and Uppal 2009; Elkamhi et al. 2020; Martínez-Nieto et al. 2021; Clarke, De Silva, and Thorley 2006; Raffinot 2017; Jain and Jain 2019).

These differences in results can be interpreted as old estimators being improved. However, one cautionary aspect to consider is that there are several influential parameters when evaluating estimators as there is no consensus on datasets, performance measures or method for calculating the performance measures. Hence, selection bias can affect findings (Bailey and De Prado 2014). To exemplify with a parallel, the durability of one car tire will differ depending on what season it is tested, if the durability is measured using weight or thread and if the measure is performed using a scale or change in water volume. For estimation comparisons, this can lead to results depending not only on data, but on choice of evaluation method and its implementation too. Hence, some common choices will be outlined below.

Although markets differ in terms of several parameters – such as time period, size and location – the majority of papers above use empirical returns. A prevalent method is to choose portfolios of different sizes for markets with varying magnitudes of volatility and returns. In simulation studies, there can be a bit more variability. Here returns can be simulated using different distributions or models such as factor models. One noteworthy difference between simulations and empirical studies is that simulations often assume time independence, which is not always realistic, but sometimes warranted for comparison purposes.

To measure the performance of estimators, out of sample measures are often used as the estimators likely will be used for calculations on novel data. Out of sample variance – i.e. the variance of the portfolio outside of the sample where the weights are estimated – is a relevant measure, as the objective is to minimize variance. However, as most investors are likely to favor high returns, different versions of the Sharpe ratio are also popular. These measure returns in relation to risk.

For empirical data, out of sample measures often involve calculations on time series. Here there are different methods of dividing the time series into testing and validation sets, with a popular being the rolling window approach (Bergmeir and Benítez 2012; Tashman 2000). For simulated data, the covariance can be known, which allows the calculation of the estimation error using the minimum variance:

$$\frac{1_{p\times 1}}{1_{p\times 1}^{t}\Sigma^{-1}1_{p\times 1}}$$

.

2.2 Estimators to be compared

2.2.1 Shrinkage and the Bona-fide estimator

One method that addresses error maximization is the use of shrinkage estimators (Stein et al. 1956; Ledoit and Wolf 2004). A good explanation to how shrinkage works is that of Efron and Morris (1977). They exemplify that the mean batting ability of a baseball player can be estimated more accurately by reducing/increasing i.e. shrinking it towards the mean of all baseball players means (using a shrinkage factor). If we let m denote the mean batting ability of one player and m_m denote the mean of all players means, we can estimate the ability using:

$$ability = m_m + \alpha(m - m_m) \overset{alternatively}{=} \alpha m + (1-\alpha)m_m$$

The shrinkage factor α is tweaked differently for different problems. We can see that $\alpha = 1$ will result in the original average.

More generally we can say that shrinkage involves shrinking an estimation (m above) towards a target (m_m above).

Bodnar, Parolya, and Schmid (2018) construct a shrinkage estimator for $w_{traditional}$ using the equal weights portfolio $\hat{w}_{eq} = diag(1/\#assets)_{p\times 1}$ referred to as their "Bona-Fide estimator" (BF):

$$\hat{w}_{bf} = \hat{\alpha}\hat{w}_{traditional} + (1-\hat{\alpha})\hat{w}_{eq},\ for\ c = p/n < 1, n = \#observations$$

The formula is constructed to perform well outside of the estimation sample, using the following shrinkage factor:

$$\hat{\alpha} = \frac{(1-c)B}{c+(1-c)B}, B = (1-c)\hat{w}_{eq}^t\hat{\Sigma}\hat{w}_{eq} \times 1_{p\times 1}^t\hat{\Sigma}^{-1}1_{p\times 1} - 1$$

BF will serve as our candidate for the shrinkage approach.

2.2.2 Clustering and Hierarchical Risk Parity

Clustering is a useful tool for grouping entities into groups based on certain criteria, most commonly similarity. Generally it involves using some distance measure (e.g. distance between cities) to group data into categories (e.g municipalities)

The use of clustering can aid portfolio selection in different ways. One application is to improve reliability by being used in algorithms to filter uncertainties out of the correlation matrix (Tola et al. 2008).

Another application that we will focus on in this thesis improves methods that work better on diagonal matrices. One such method is the inverse volatility portfolio (IVP), which resembles the equal weights portfolio. The difference is that it allocates risk equally instead of capital, using $w_{ivp} = diag(\frac{1/\sigma_i^2}{\sum_i^p 1/\sigma_i^2})_{p\times 1}$. w_{ivp} are not as influenced by conditioning errors as $w_{traditional}$ due to the simpler calculation. Instead it is vulnerable when the covariance matrix is not diagonal, as nondiagonal covariance information is discarded. The algorithmic clustering technique for stabilizing IVP will be the second estimator studied in this thesis and hence outlined below.

In 2016, De Prado presented the "Hierarchical Risk Parity" (HRP) algorithm that uses hierarchical clustering to improve the conditions for IVP construction. The algorithm estimates $\widehat{w}_{hrp}$ using the following three steps:

1. Hierarchical clustering: The assets are clustered such that assets with similar correlation are close in the cluster.
2. Quazi diagonalization: Rearranges the covariance matrix so the largest values are near the diagonal.
3. Recursive bisection: Uses the column order from step 2 to recursively calculate $\widehat{w}_{ivp}$:

   A) Initialization of the algorithm: Creates a column list and starting weights.
   A1: L = $\{L_0\}$, L_0 contains the column order from step 2.
   A2: $w_i = 1, \forall i \in [1, p]$.

   B) Check if all weights are assigned: Stop if $|L_i| = 1, \forall L_i \in L$.

   C) Assign cluster weights:
   C1: Split L_i into two subsets, $L_i^{(1)} \cup L_i^{(2)} = L_i, |L_i^{(1)}| = \lfloor |L_i|/2 \rfloor$. Column order is preserved.
   C2: Calculate $\tilde{V}_i^j = \tilde{w}_i^{(j)^t} \tilde{V} \tilde{w}_i^j$, where V_i is the covariance matrix of the columns in L_i^j and w is their IVP weights.
   C3: Compute the splitfactor $\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + V_i^{(2)}}$
   C4: Rescale weights of $L_i^{(1)}$ columns by α_i
   C5: Rescale weights of $L_i^{(2)}$ columns by $1 - \alpha_i$
   C6: Loop to B)

HRP will serve as our candidate for the clustering approach.

3 Simulation design

To compare the methods from the papers, it is important to have representative data and validation methods (Choi, Lim, and Choi 2019). As there is no standard method, replicating the methods previously used could be a fair strategy. Both papers focus on generating good out of sample variance (OSV), but measure it differently and on different data. Hence, our simulation study will attempt to create a reproducible approach which covers a variety of markets.

3.1 Data

As the included papers do not use the same data, a choice for backtesting data will have to be made. Since De Prado (2016) only has one dataset (see appendix), it will be easier to follow a setup similar to that paper and hence have a comparison on a common ground. However, the usage of shocks with very specific values and correlation structure might make the data too specific for general statements of the methods performance (Bailey and De Prado 2014). One way to avoid selection bias is to consider a wide variety of conditions/markets. With the assumptions of modern portfolio theory, that could be achieved by having several different covariance matrices.

There is plenty of research on the properties of financial covariance/correlation matrices and how to simulate them (Hüttner and Mai 2019; Hirschberger, Qi, and Steuer 2007; Numpacharoen and Atsawarungruangkit 2012; Hardin, Garcia, and Golan 2013; Córdoba et al. 2018; Joe 2006). Joe (2006) proposes a method that allows us to sample uniformly from the space of correlation matrices, which could allow us to cover a variety of correlation structures. Hence, this matrix and uniformly sampled variances will be used to generate covariance matrices. According to Morningstar.com (2022), the sp500 3, 5 and 10 year standard deviations are 0.1776, 0.1578 and 0.1323, hence we will sample the standard deviations from U(0,0.3) to have an expected value close to their empirical average.

In summary, the Guassian data for the experiments will be simulated using random correlation structures by generating random covariance matrices converted from random correlation matrices and standard deviations drawn from U(0,0.3) (variances will then be in [0,0.09]). Time independence will be assumed.

3.2 Simulation study description

Since we perform a simulation study, our parameters can be known. Hence, an approach similar to that of Bodnar, Parolya, and Schmid (2018) will be used for each of our experiments (see appendix). For each generated covariance matrix, the estimated weights will be used to calculate the relative error using the estimated theoretical and minimum out of sample variance ($estimated = \widehat{w}^t \Sigma \widehat{w}$, $min = \frac{1_{p\times 1}}{1^t_{p\times 1}\Sigma^{-1}1_{p\times 1}}$) as $\frac{estimated-min}{min}$.

To get some indications on the effect different markets have on the estimators, a scalar variance and correlation measure will be used. In Peña and Rodríguez (2003) we can find a measure of correlation as $|C|^{1/p}$ where C is the correlation matrix. This correlation measure – the effective correlation – displays values between 0 and 1, where 0 means high dependence and 1 means no dependence. As this is counter intuitive to the commonly used correlation coefficient, the intuitive behavior will be created by subtracting 1 from calculated effective correlations. In Puntanen (2013) both $|\Sigma|$ and $tr(\Sigma)$ are suggested as generalized variance measures. $tr(\Sigma)/p$ will be chosen as it removes the effect of correlation, which might have caused it to display similar patterns as our correlation measure.

Furthermore, condition numbers will be calculated for each matrix to allow for measuring the effect of conditioning.

We will conduct the experiments for the HRP and BF estimators, but include the traditional estimator hereafter named as MVP along with IVP to see if any improvements are made. Using this setup, three simulations will be performed:

- **Main experiment**: This experiment will try to follow the setup of De Prado (2016) by using 10 assets and an estimation sample of 260 observations. 10^6 estimations will be performed.
- **N experiment**: This experiment will investigate the effect of the number of observations, and hence estimation errors. Hence, 10^6 estimations will be performed for each of N =11, 55 and 110 observations. The number of assets will be kept at 10. Hence the c = asset/dimension ratio will be 1.1, 5.5 and 11 for each N respectively.
- **P experiment**: This experiment will investigate the effect of the number of assets, and hence conditioning errors. For performance reasons only 10^4 estimations will be performed for each of P = 50, 100 and 150 assets. The number of observations will be 1300, 2600 and 3900 respectively to have the same asset to dimension ratio as in the main experiment.

4 Simulation study

4.1 Main experiment

From Table 2 below, we can see that the general performance of IVP and HRP was substantially worse compared to MVP and BF across all measures. Furthermore we can see that HRP and BF improved median and MAD compared to the methods they were based on.

Table 2: Minimum, median and median absolute deviation of the relative errors for each estimator.

w_method	Minimum	Median	MAD
hrp	0.0294205	5.1870299	5.2573015
ivp	0.0250951	5.3782112	5.4309684
bf	0.0004716	0.0333188	0.0161636
mvp	0.0005412	0.0333551	0.0161859

4.1.1 Raw results

From Table 3 below we can observe that our measured parameters do not display strong correlations. The condition numbers however, do show tendencies to be affected by variance and more strongly by correlation.

Table 3: Spearman correlations of the mean variance, effective correlation, condition number and logarithmic condition number as mVar, eCorr and cond and log.cond respectively.

	mVar	eCor	cond	log.cond
mVar	1.0000000	0.0000060	-0.1979897	-0.1979897
eCor	0.0000060	1.0000000	-0.3987116	-0.3987116
cond	-0.1979897	-0.3987116	1.0000000	1.0000000
log.cond	-0.1979897	-0.3987116	1.0000000	1.0000000

From Figure 2 and 3 below, we can see that the generated matrices are ill conditioned. In Figure 3 we can also see that the conversion of correlation matrices to covariance matrices caused a convergence towards the dominating eigenvalue, which is close to 0.

Figure 2: Distribution of condition numbers and logarithmic condition numbers to the left and right respectively

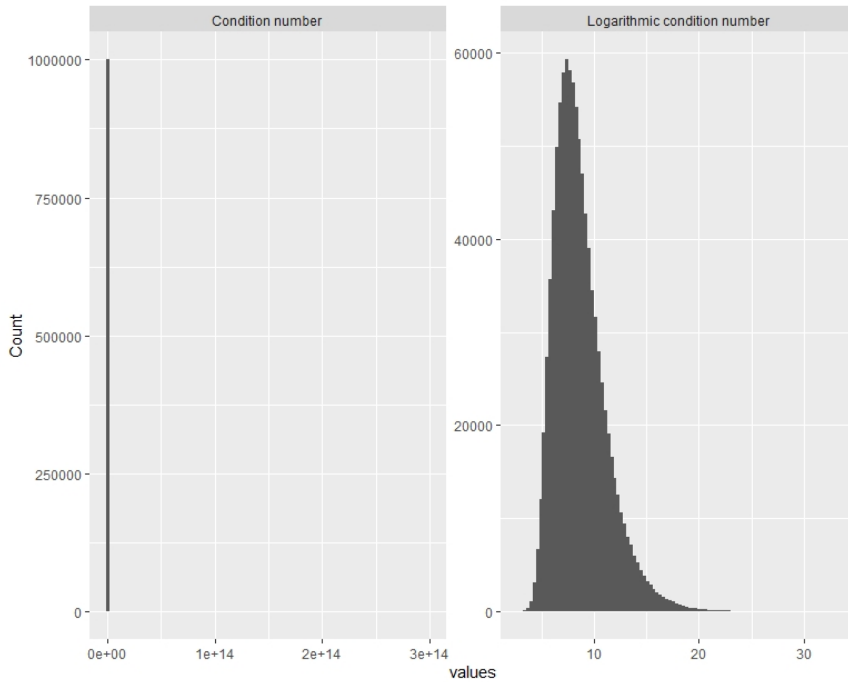
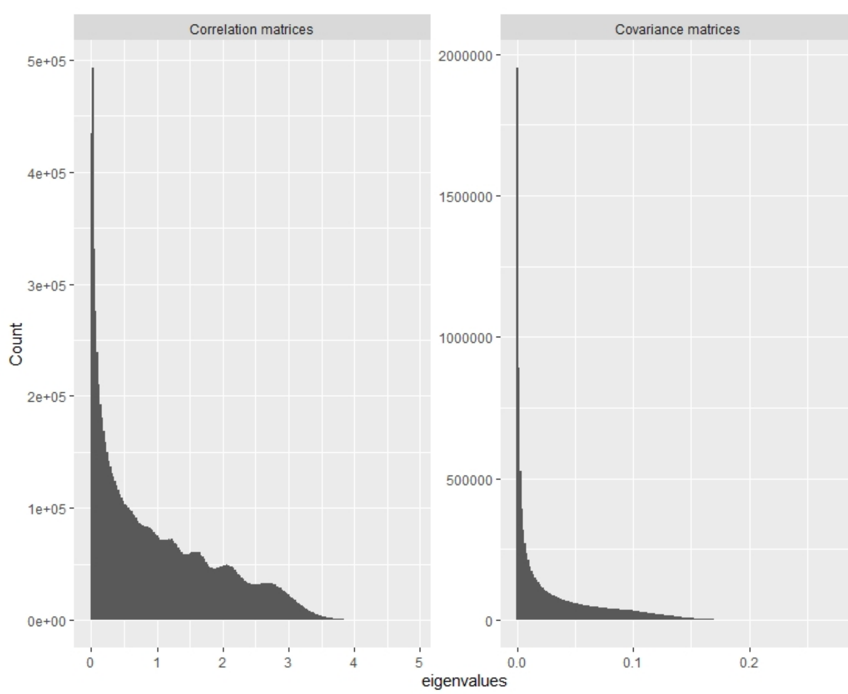
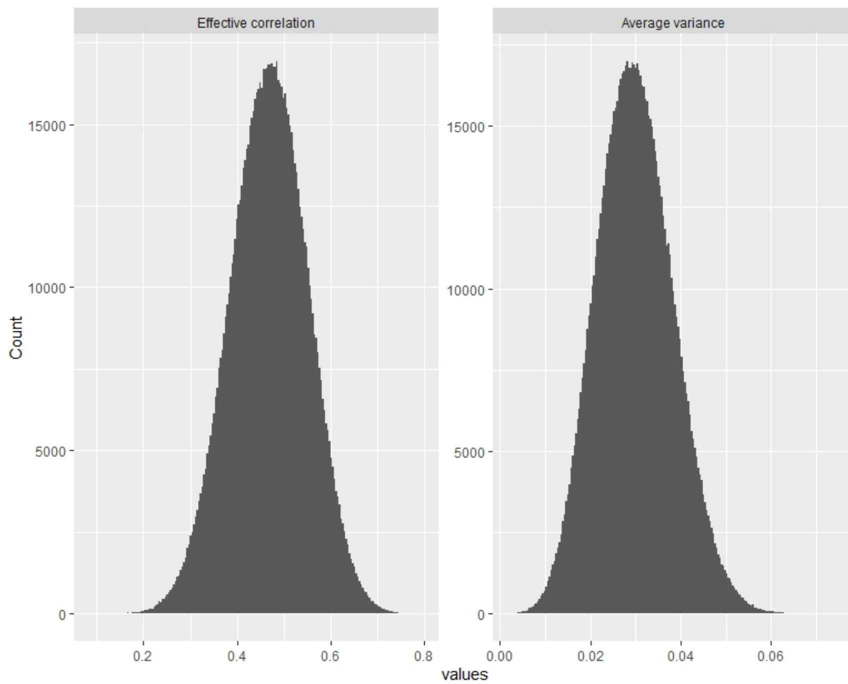


Figure 3: Distribution of eigenvalues for the correlation and covariance matrices to the left and right respectively



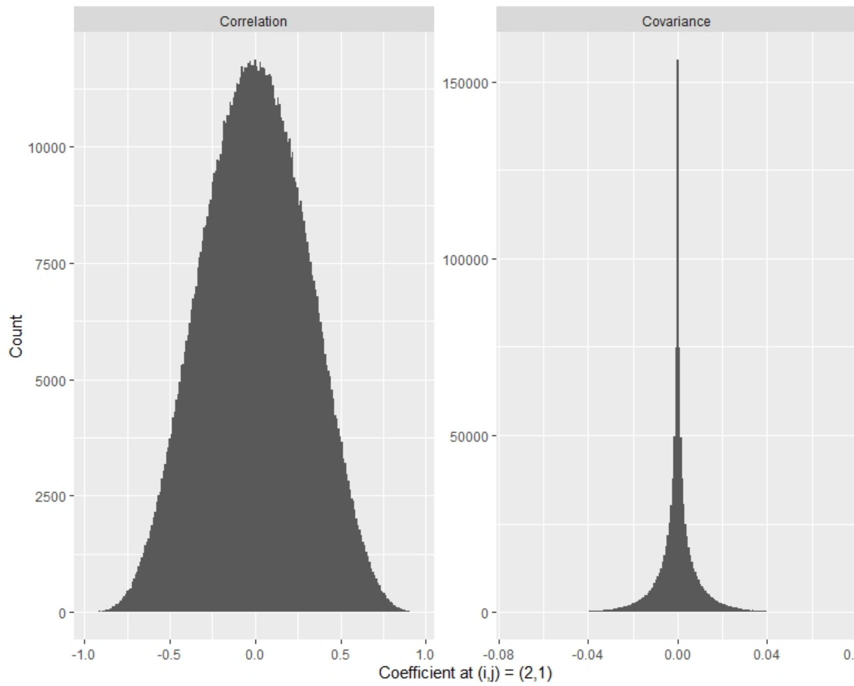
In Figure 4 below, we can see how both the effective correlation and mean variance have a bell curve distribution with means close to the expected values for the correlation and variance spans [0,1] and $[0,0.3]^2$.

Figure 4: Distributions of effective correlation and mean variances to the left and right respectively



In Figure 5 below, we can see a clear bell curve for the correlation coefficients, while the curve is much steeper for the covariances. This may be explained by the eigenvalue convergence displayed in Figure 3.

Figure 5: Distribution of correlation and covariance coefficients at (i,j) = (2, 1) to the left and right respectively



For the bell curves in figure 4 and 5, we can see differences in width which indicates an unequal mixture in favoring magnitudes. Too wide and they will favor extreme values, too narrow and the opposite occurs.

4.1.2 Correlation against error

From Figure 6 and 7 below we can see that MVP and BF seem stable, while errors seem to increase drastically with correlation for IVP and HRP. Again the performance of the original estimators are very similar to the improved ones. The increased errors for the IVP based methods are not surprising as they delete covariance information outside of the diagonal. In line with Table 3's low correlation, mean variance does not seem to affect these results.

Figure 6: Relative error against effective correlation for each estimator. Mean variance above or below the median is denoted as High and Low respectively to display its effect.

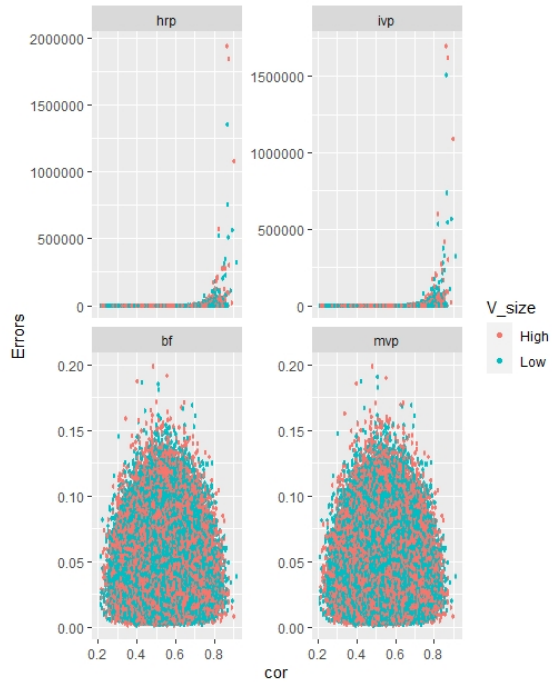
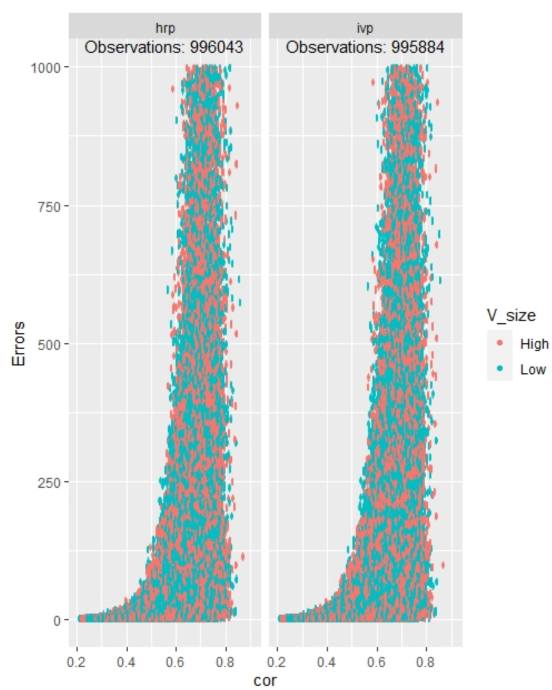


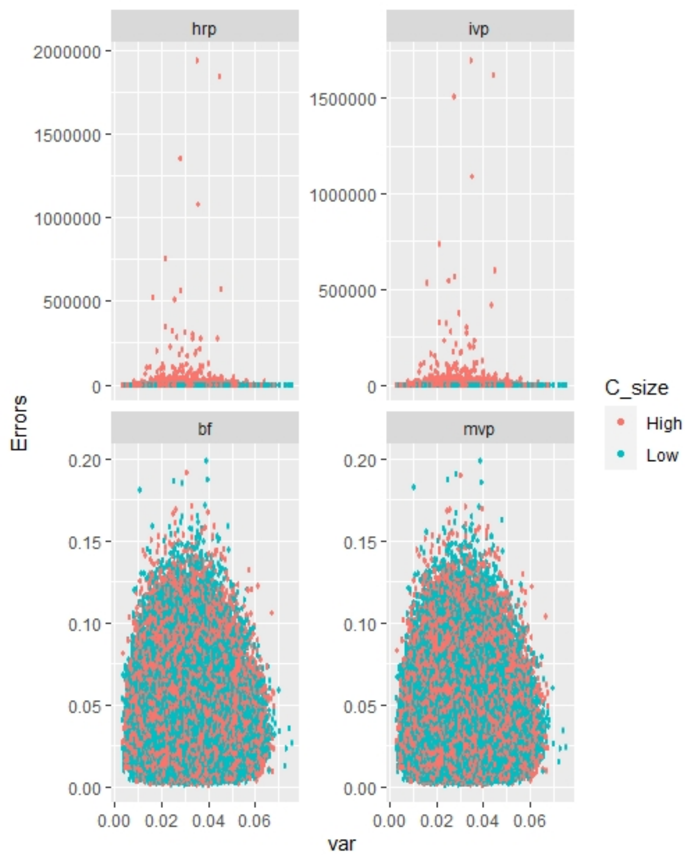
Figure 7: Plots relative error against effective correlation for IVP and HRP with smaller y-axis. Mean variance above or below the median is denoted as High and Low respectively to display its effect.



4.1.3 Variance against error

Again MVP and BF seem to show weak dependencies, while some patterns are shown for IVP and HRP in Figure 8. However, the coloring of correlation and positioning at the center of the bell curve of the mean variances from Figure 4 warrant investigating if correlation causes this pattern.

Figure 8: Plots relative error against mean variance for each estimator. Effective correlation above or below the median is denoted as High and Low respectively to display its effect.



In Figure 9 we can see that most of the pattern seems to come from high correlations, which can be supported in Figure 10 where most correlations are removed. Hence there are indications that variance does not strongly influence errors.

Figure 9: Plots relative error against effective correlation for IVP and HRP, excluding values above the median effective correlation.

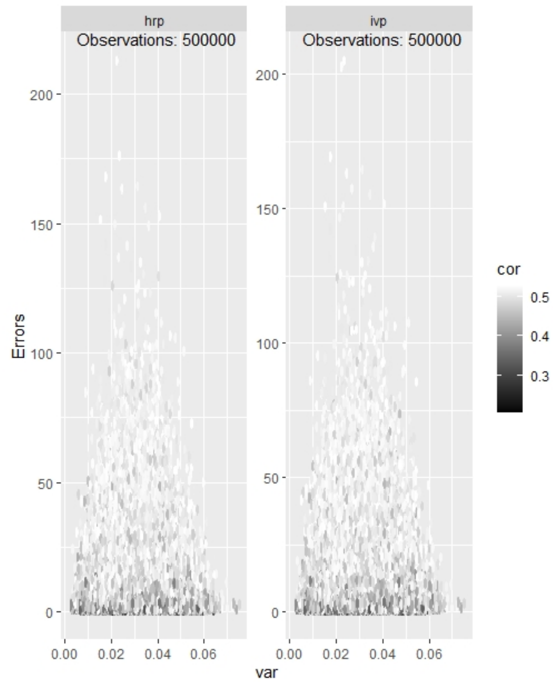
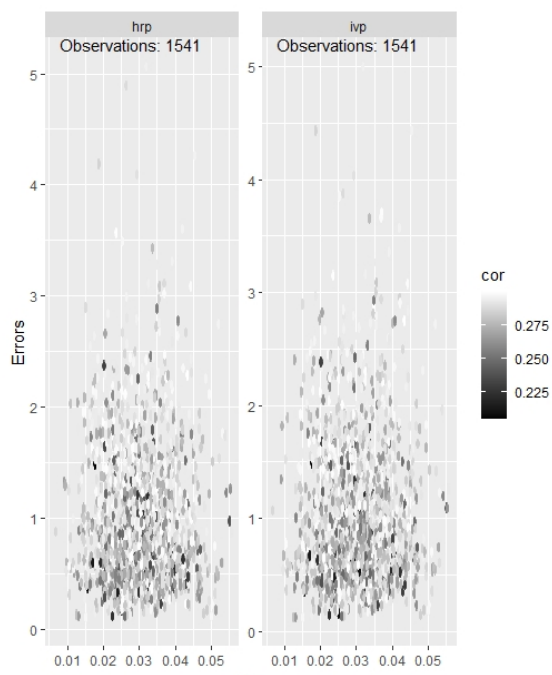


Figure 10: Plots relative error against effective correlation for IVP and HRP, excluding values with correlation higher than 0.3



4.1.4 Condition number against error

In Figure 11 we see color patterns based on correlation, which indicates that it may interfere. This is supporterd in Figure 12 where filtering correlations reveals a different pattern where all methods seem to perform worse for higher condition numbers. However, again the magnitudes for HRP and IVP are notably higher.

Figure 11: Plots relative error against logarithmic condition numbers. Effective correlation above or below the median is denoted as High and Low respectively to display its effect.

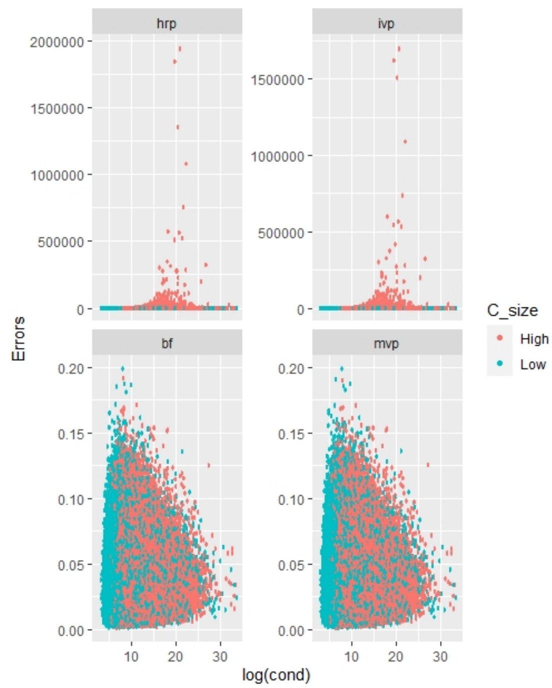
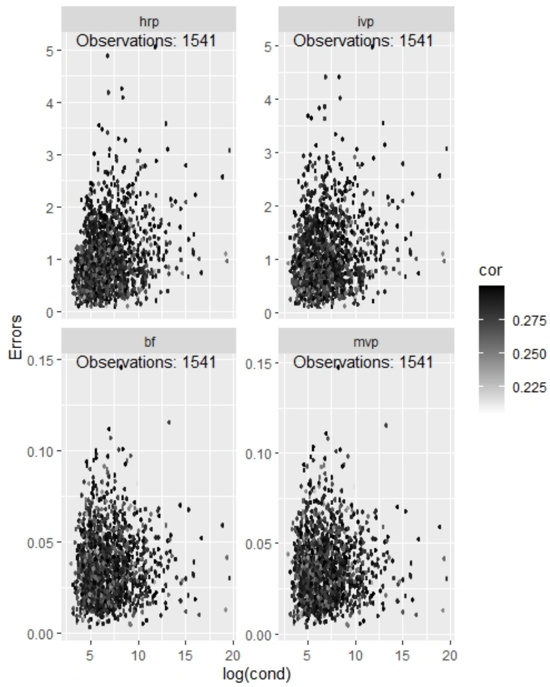


Figure 12: Plots relative error against logarithmic condition numbers after filtering correlations above 0.3



4.2 N and P experiment

From Figure 13 and 14 it can be seen that the chosen values for P and N do effect the outcomes. From Figure 13 we can see a substantially slower error and spread reduction as N increases for HRP and IVP. Figure 14 and 15 also display worse performance as dimension – and hence condition numbers – increase for HRP and IVP.

Figure 13: Violinplots of logarithmic errors with annotated raw median for different number of observations.

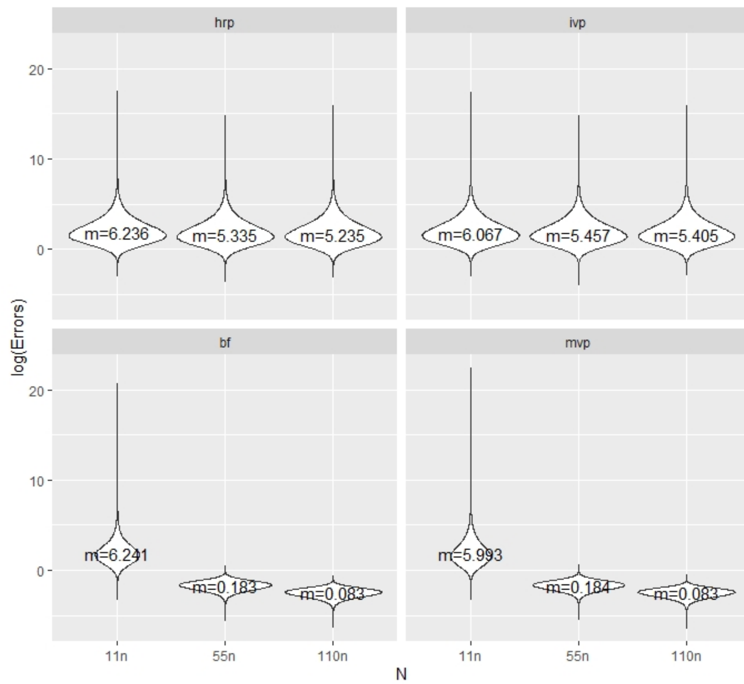


Figure 14: Violinplots of logarithmic errors with annotated raw median for varying portfolio sizes.

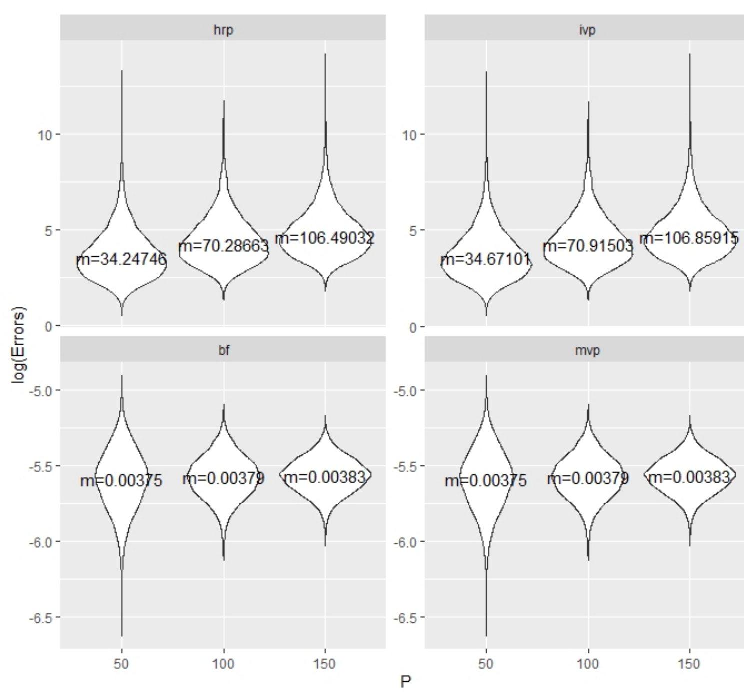
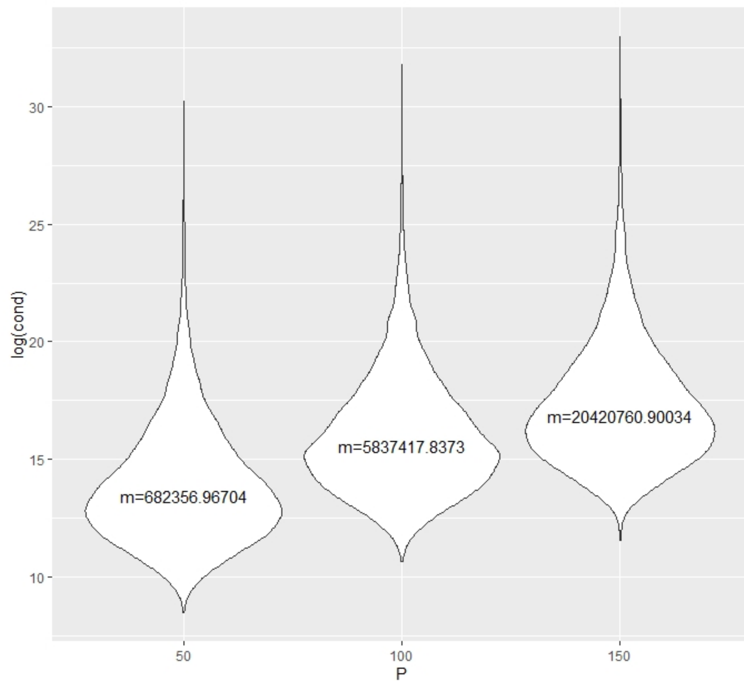


Figure 15: Violinplots of logarithmic condition numbers with annotated raw median for different portfolio sizes.



5 Conclusion and evaluation

5.1 Conclusion

To investigate the performance of HRP and BF estimators, a simulation study where variance, correlation, P and N varied was performed. The effective correlation, mean variance and condition numbers were used to measure the parameters effect on relative out of sample errors. For reference, the IVP and MVP estimators were used.

Which method produces the smallest out of sample variance error generally? From our simulations we can see that BF produces substantially lower median and MAD relative errors. Furthermore, we can see that using clustering and shrinkage produced better results than the original estimators, IVP and MVP. However, the improvements can be seen as marginal compared to the differences between analytical and IVP based estimators.

How do the market parameters correlation, variance and conditioning affect the out of sample variance? BF performed stable with respect to correlation while HRP showed a strong increase in errors. Both measures seem to not display any relationship with mean variance. Both measures also showed a tendency for increased errors when condition numbers increase, however the errors seem larger for HRP.

How do the choices of number of assets and observations affect the out of sample variance? When number of assets increase HRP displayed higher increase in errors and spread compared to BF. BF showed much faster improvement in error and spread reduction compared to HRP when number of observations increased.

5.2 Improvements and suggestion for further research

As discussed, correlation and condition-numbers affect the IVP-methods. Therefore, extended studies focusing on more specific markets could give valuable insights in recommended and ill advised areas of use. This study tried to cover a whole set of markets and hence does not have a reliable coverage for all areas, as can be seen in the centered and varying distributions for the market parameters.

In this thesis, we used time independent Guassian data from known covariance matrices to study the effect of covariance estimation errors. Here the distribution was only assumed to be able to generate a covariance matrix with a random error compared to the known one. Hence, one idea that can remove assumptions in terms of number of observations and distribution would be to directly generate matrices with random errors from a known one. As an example for intuition, the Wishart distribution could take a known covariance matrix as parameter and then generate random ones.

The generation of random covariance matrices for this thesis could be improved. The difference in shape of the covariances and correlations in Figure 3, indicate that the dominating eigenvalues close to 0 might have changed the distribution of covariances towards 0. Hence, another aspect worth investigating is how to generate random covariance matrices that are realistic for financial markets. Combining this with the use of other distributions than Gaussian data, might allow the simulations to cover more markets. The goal would be to create a common benchmark where different estimators can be evaluated. This might reduce the need for difficult backtesting, by shifting the problem to identifying the markets similar to where estimators show good performance in simulations.

Furthermore, no shocks were included in the data. Hence, another improvement would be investigating the effect of shocks in different locations and of different magnitudes. The more general problem that this can be summarized as is the need for representative data to evaluate estimators on.

One final aspect that was not investigated in this thesis was the effect of weight constraints. All estimators – except the traditional – return positive weights. Knowing the impact of such constraints will also allow for taking advantage of areas where they display favorable performance.

Bibliography

Bailey, David H, and Marcos Lopez De Prado. 2014. "The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality." *The Journal of Portfolio Management* 40 (5): 94–107.

Bergmeir, Christoph, and José M Benítez. 2012. "On the Use of Cross-Validation for Time Series Predictor Evaluation." *Information Sciences* 191: 192–213.

Bodnar, Taras, Nestor Parolya, and Wolfgang Schmid. 2018. "Estimation of the Global Minimum Variance Portfolio in High Dimensions." *European Journal of Operational Research* 266 (1): 371–90.

Choi, Young-Geun, Johan Lim, and Sujung Choi. 2019. "High-Dimensional Markowitz Portfolio Optimization Problem: Empirical Comparison of Covariance Matrix Estimators." *Journal of Statistical Computation and Simulation* 89 (7): 1278–1300.

Clarke, Roger G, Harindra De Silva, and Steven Thorley. 2006. "Minimum-Variance Portfolios in the US Equity Market." *The Journal of Portfolio Management* 33 (1): 10–24.

Córdoba, Irene, Gherardo Varando, Concha Bielza, and Pedro Larrañaga. 2018. "A Fast Metropolis-Hastings Method for Generating Random Correlation Matrices." In *International Conference on Intelligent Data Engineering and Automated Learning*, 117–24. Springer.

De Prado, Marcos Lopez. 2016. "Building Diversified Portfolios That Outperform Out of Sample." *The Journal of Portfolio Management* 42 (4): 59–69.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. "Optimal Versus Naive Diversification: How Inefficient Is the 1/n Portfolio Strategy?" *The Review of Financial Studies* 22 (5): 1915–53.

Efron, Bradley, and Carl Morris. 1977. "Stein's Paradox in Statistics." *Scientific American* 236 (5): 119–27.

Elkamhi, Redouane, Chanik Jo, Jacky SH Lee, and Marco Salerno. 2020. "The Jury Is Still Out on the Performance of Naive Diversification (1/n Rule)." *Rotman School of Management Working Paper*, no. 3638713.

Hardin, Johanna, Stephan Ramon Garcia, and David Golan. 2013. "A Method for Generating Realistic Correlation Matrices." *The Annals of Applied Statistics*, 1733–62.

Hirschberger, Markus, Yue Qi, and Ralph E Steuer. 2007. "Randomly Generating Portfolio-Selection Covariance Matrices with Specified Distributional Characteristics." *European Journal of Operational Research* 177 (3): 1610–25.

Hüttner, Amelie, and Jan-Frederik Mai. 2019. "Simulating Realistic Correlation Matrices for Financial Applications: Correlation Matrices with the Perron–Frobenius Property." *Journal of Statistical Computation and Simulation* 89 (2): 315–36.

Jain, Prayut, and Shashi Jain. 2019. "Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification." *Risks* 7 (3): 74.

Joe, Harry. 2006. "Generating Random Correlation Matrices Based on Partial Correlations." *Journal of Multivariate Analysis* 97 (10): 2177–89.

Laloux, Laurent, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. 1999. "Noise Dressing of Financial Correlation Matrices." *Physical Review Letters* 83 (7): 1467.

Ledoit, Olivier, and Michael Wolf. 2004. "Honey, i Shrunk the Sample Covariance Matrix." *The Journal of Portfolio Management* 30 (4): 110–19.

Markowitz, Harry. 1952. "Portfolio Selection in the Journal of Finance Vol. 7."

Martínez-Nieto, Luisa, Francisco Fernández-Navarro, Mariano Carbonero-Ruz, and Teresa Montero-Romero. 2021. "An Experimental Study on Diversification in Portfolio Optimization." *Expert Systems with Applications* 181: 115203.

Michaud, Richard O. 1989. "The Markowitz Optimization Enigma: Is 'Optimized'optimal?" *Financial Analysts Journal* 45 (1): 31–42.

Morningstar.com. 2022. "S&p 500 PR." 2022. https://www.morningstar.com/indexes/spi/spx/risk.

Numpacharoen, Kawee, and Amporn Atsawarungruangkit. 2012. "Generating Correlation Matrices Based on the Boundaries of Their Coefficients." *PLoS One* 7 (11): e48902.

Peña, Daniel, and Julio Rodríguez. 2003. "Descriptive Measures of Multivariate Scatter and Linear Dependence." *Journal of Multivariate Analysis* 85 (2): 361–74.

Puntanen, Simo. 2013. "Methods of Multivariate Analysis, by Alvin c. Rencher, William f. Christensen." *International Statistical Review* 81 (2): 328–29.

Raffinot, Thomas. 2017. "Hierarchical Clustering-Based Asset Allocation." *The Journal of Portfolio Management* 44 (2): 89–99.

Stein, Charles et al. 1956. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1:197–206. 1.

Sun, Ruili, Tiefeng Ma, Shuangzhe Liu, and Milind Sathye. 2019. "Improved Covariance Matrix Estimation for Portfolio Risk Measurement: A Review." *Journal of Risk and Financial Management* 12 (1): 48.

Tashman, Leonard J. 2000. "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review." *International Journal of Forecasting* 16 (4): 437–50.

Tola, Vincenzo, Fabrizio Lillo, Mauro Gallegati, and Rosario N Mantegna. 2008. "Cluster Analysis for Portfolio Optimization." *Journal of Economic Dynamics and Control* 32 (1): 235–58.

Appendix

Data used for backtesting in Bodnar, Parolya, and Schmid (2018) as well as De Prado (2016)

Both empirical and simulated data can be used for backtesting. In Bodnar, Parolya, and Schmid (2018), the performance of the estimator is evaluated using both a simulation and empirical study measuring out of sample variance. For the simulations, data distributed according to a normal and a t-distribution with 5 degrees of freedom is used to simulate covariance matrices to apply the methods to. For the empirical study the 417 assets between 22.04.2013 to 19.03.2014 of the of the S&P500 stock exchange were used.

De Prado (2016) used a setup to be able to repeatedly construct 520 observations of 10 assets that are given a random correlation structure and shocks using the following approach.

Creating random correlation:

1. Create the matrix x: 520 observations of five columns drawn independently from N(0,var = .01).
2. Uniformly pick between the 5 columns five times, with repetition allowed. These will be altered and used to extend x, thus introducing correlation as they are an altered copy of x.
3. The noise for altering the the result from step 3 is created by sampling as in step 1, but from N(0, var = 0.01*0.25).
4. The matrices from step 2 and 3 are added together.
5. The matrix from step 4 is used to extend the matrix from step one to 520x10 matrix.

Adding random shocks:

7. Uniformly select two observations between 260 and 520.
8. Common shock: Replace the values at the first column from step 2 and column 5 with -.5 and 2 for the first and second observation respectively.
9. Random shock: Repeat step 7, and replace the last column of step 2 with -.5 and 2 for first and second observation respectively.

Measuring out of sample variance in Bodnar, Parolya, and Schmid (2018) as well as De Prado (2016)

In the simulation of Bodnar, Parolya, and Schmid (2018) simulated data with known parameters, hence the known minimum out of sample variance $\frac{1_{p\times 1}}{1_{p\times 1}^t \Sigma^{-1} 1_{p\times 1}}$ could be used to calculate the error with the theoretical out of sample variance $\widehat{w}^t \Sigma \widehat{w}$. 1000 experiments were conducted for each of the selected ratios between the number of observations and portfolio size. For the empirical study a rolling window approach was used to estimate the out of sample variance. As the name suggests, a window within the data is used to estimate weights and the data ahead of the window is used to estimate the OSV:

1. 54 of the 417 assets are randomly chosen as a portfolio. E.g. Apple, Tesla, ... , Microsoft.
2. Observations in [1,n] (n < number of observations) in the window are used to estimate $\widehat{w}_{mvp}$. E.g. [1,260].
3. A out of sample return is generated using the observation ahead of the window. E.g. $returns_{261} \times \widehat{w}_{mvp}$ for observations [261].
4. If the window is not at the end of the time series, the window is moved one step forward [1-260]->[2-261] and we restart at step 2.
5. If the window has reached the end, the variance of the returns is calculated. The process restarts at Step 1 until 1000 variances are calculated.

De Prado (2016) uses a simulation study with 10000 iterations where each generates an out of sample return. All of them are then used to compute the out of sample variance. To calculate the out of sample returns, the timeseries is traversed using a window and 22 points ahead of it as follows:

1. Observations in [1,260] were used to estimate $\widehat{w}_{mvp}$.
2. Observations [260,260+22=282] were used to generate out of sample returns using the weights: $returns \times \widehat{w}_{mvp}$.
3. The window is shifted 22 steps forward (window + [22, 22]) and the previous steps are repeated until the 22 points ahead would exceed 520.
4. Then the dataset return is calculated as the resulting return for the out of sample returns r_i. That is done by converting the returns from proportion change to proportion results, and then multiplying them. After that one is subtracted to convert back e.g. $[\prod_{i=260}^{502}(r_i + 1)] - 1$.