Delft University of Technology

Faculty of Electrical Engineering, Mathematics and Computer Science

Delvin Karjadi

Bachelor Thesis for Applied Mathematics

Bootstrap-based bias correction for the out-of-sample Sharpe ratio

Supervisor : Dr. J. Söhl
Other committee members : Dr.ir. L.E. Meester

Delft, June 2023

Abstract

Looking for making an investment, one objective could be to find a portfolio where the Sharpe ratio for in the future, known as the out-of-sample Sharpe ratio, is maximized. Since future data is not available, the Sharpe ratio needs to be predicted using historical data, the in-sample data. This is often done using the Sharpe Ratio Information Criterion, which determines the bias for the in-sample Sharpe ratio to estimate the out-of-sample Sharpe ratio. However, this approach assumes that the covariance matrix is known.

In portfolio management, the covariance matrix is typically unknown and can only be estimated. This project will use the bootstrap method to estimate the out-of-sample Sharpe ratio using the estimated covariance matrix and analogous methods used for the Akaike Information Criterion. By eliminating the assumption of a known covariance matrix, this method becomes more applicable. Simulations will also be done with a known covariance matrix, demonstrating that the bootstrap method is an effective approach for estimating the out-of-sample Sharpe ratio. We then look at some extensions for the bootstrap method and finally we will apply the bootstrap method to stocks in the Dutch and American stock markets, showing that the in-sample Sharpe ratio is often overly optimistic compared to the out-of-sample Sharpe ratio. We reached our goal that we found an effective way to estimate the out-of-sample Sharpe ratio without the assumption that the covariance matrix is known, resulting this method becomes much more suitable for predicting the Sharpe ratio in the future.

Contents

1 Introduction **3**

2 Portfolio Optimization **4**
2.1 Sharpe Ratio 4
2.2 Sharpe Ratio Information Criterion 4
2.3 Akaike Information Criterion 6

3 The Boostrap Method **9**
3.1 Bias for the Sharpe Ratio Information Criterion 13
3.2 Bias with Unknown Covariance Matrix 18
3.3 Bias with real data 20
3.4 Parametric Bootstrapping 21
3.5 Efficient Bootstrap Simulation 24

4 Optimal out-of-sample performance on the stock market **27**
4.1 AEX index 27
4.2 Dow Jones 29

5 Conclusion **31**

6 Appendix **32**

1 Introduction

In investment analysis, the Sharpe ratio is commonly used as a measure, dividing the portfolio's return by its risk. It quantifies the additional return gained for the extra risk taken. Therefore, the objective is to find a portfolio where the future Sharpe ratio, known as the out-of-sample Sharpe ratio, is maximized. Since only historical data is available, the goal is to find a bias correction method to predict the out-of-sample Sharpe ratio based on the estimated mean of returns of the stocks $\hat{\mu}$ and the estimated covariance matrix $\hat{\Sigma}$. The covariance matrix is a matrix that represents the relationship between the different stocks and how they move together. The covariance matrix is estimated by

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{i=1}^{T}(X_i - \hat{\mu})(X_i - \hat{\mu})^T,$$

where T is the number of samples, $\hat{\mu}$ the estimated mean vector of returns and X_i represents the $i-th$ sample vector. To estimate the out-of-sample Sharpe ratio, we will first look at the Akaike Information Criterion (AIC), which focuses on log-likelihood instead of the Sharpe ratio, and the Sharpe Ratio Information Criterion (SRIC), which predicts a bias correction assuming that the true covariance matrix Σ is known, although in reality this is never known. Utilizing the bootstrap method, we will estimate a bias correction using the estimated covariance matrix. The bootstrap method works by repeatedly sampling from the available data with replacement to create new data sets of the same size as the original. By repeating this process a large number of times, a distribution of the statistic is obtained, which can be used to estimate its properties, such as its mean or standard deviation. Therefore, for every simulation, we are able to calculate the Sharpe ratio based on the bootstrap data and we can compare it with the original data.

Using the bootstrap method, we can also determine the bias for the Sharpe ratio assuming the covariance matrix is known. By simulating data based on specific mean of returns μ and covariance matrix Σ. This enables us to compare the performance of the bias estimated through the bootstrap method, the Sharpe Ratio Information Criterion (SRIC), and the actual bias.

2 Portfolio Optimization

Portfolio optimization is the process of constructing an investment portfolio using a certain criterion. In finance, there are several criteria an investor can apply to make an investment decision. The type of criterion being used depends on the interests and vision of the investor themselves. This can be maximizing the expected return on the short term or minimizing the risk for a given level of expected return on the long term. Is there a measure that maximizes returns while minimizing risk? A widely used measure for this is the Sharpe Ratio, developed by William F. Sharpe in 1966. The Sharpe ratio is a way to measure the risk-adjusted returns of your investments.

2.1 Sharpe Ratio

The Sharpe ratio is the return of the portfolio divided by the standard deviation of the portfolio. The standard deviation is also referred to as the volatility of the portfolio, which represents the degree of movement or fluctuation of the portfolio itself. Therefore, the Sharpe Ratio is the extra risk required to get higher returns. The Sharpe Ratio in simplest form is defined as

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p}, \tag{1}$$

where R_p is the return of the portfolio, R_f the return on a risk-free asset and σ_p the standard deviation of the portfolio's excess return. To achieve the highest possible Sharpe ratio, it is important to maximize returns and minimize risk. Typically, an investment with a Sharpe ratio above 1 is considered good. It is easy to evaluate the Sharpe Ratio when there is a given data set. This is also known as the in-sample Sharpe Ratio, where the Sharpe Ratio is based on past observations, but for an investment the goal is to determine the Sharpe Ratio for the future, which is also known as the out-of-sample Sharpe Ratio.

Let n be the number of assets and k be the number of parameters such that $n = k + 1$. Let $\Theta = R^{k+1}$ be a $(k+1)$-dimensional parameter space with $\theta, \hat{\mu} \in \Theta$, where θ represents the weights of the stocks in the portfolio, and $\hat{\mu}$ represents the estimated mean returns over the stocks over T periods, with $[0, T]$ being the in-sample period. These periods could be weekly, monthly, or even yearly returns of the stocks. Denote Σ as the true known covariance matrix of the returns which is full of rank and assume that $\hat{\mu}$ is a noisy observation of the true unknown mean return μ, i.e., assume that $\hat{\mu} = \mu + \nu$ and $\nu \sim \mathcal{N}(0, \frac{1}{T}\Sigma)$. Then the in-sample Sharpe ratio for parameter θ is defined as follows

$$\rho(\theta) = \frac{\hat{\mu}^T \theta}{\sqrt{\theta^T \Sigma \theta}},$$

and the unobserved out-of-sample Sharpe ratio

$$\tau(\theta) = \frac{\mu^T \theta}{\sqrt{\theta^T \Sigma \theta}}$$

follows from removing the noise term from the mean returns.
Since μ is unknown, it is not possible to determine the out-of-sample Sharpe ratio directly, but there is an unbiased estimator for the out-of-sample Sharpe ratio in terms of $\hat{\mu}$, which is called the Sharpe Ratio Information Criterion.

2.2 Sharpe Ratio Information Criterion

For an investment the goal is to find an out-of-sample Sharpe ratio as high as possible. Let $\hat{\theta} \in \Theta$ be the vector that maximizes the in-sample Sharpe ratio and let $\theta^* \in \Theta$ be the vector maximizing the out-of-sample Sharpe ratio.

$$\hat{\theta} \in \arg\max_{\theta \in \Theta} \rho(\theta)$$

$$\theta^* \in \arg\max_{\theta \in \Theta} \tau(\theta)$$

Then there are four combinations of Sharpe ratios,

Symbol	Value	Parameter	Description
$\hat{\rho}$	$\rho(\hat{\theta})$	$\hat{\theta}$	In-sample Sharpe ratio of optimal parameter applied to in-sample data
ρ^*	$\rho(\theta^*)$	θ^*	In-sample Sharpe ratio of optimal parameter applied to out-of-sample data
$\hat{\tau}$	$\tau(\hat{\theta})$	$\hat{\theta}$	Out-of-sample Sharpe ratio of optimal parameter applied to in-sample data
τ^*	$\tau(\theta^*)$	θ^*	Out-of-sample Sharpe ratio of optimal parameter applied to out-of-sample data

Since $\hat{\tau}$ is desired to be maximized, $\hat{\tau}$ can be decomposed as follows :

$$\hat{\tau} = \hat{\rho} - (\hat{\rho} - \rho^*) - (\tau^* - \hat{\tau}) + (\tau^* - \rho^*).$$

Now the out-of-sample Sharpe ratio is decomposed into the in-sample Sharpe ratio minus three terms. These terms are the noise fit, estimation error and the noise in the in-sample data.

$$\text{Noise fit} = \hat{\rho} - \rho^*$$

$$\text{Estimation error} = \tau^* - \hat{\tau}$$

$$\text{Noise} = \tau^* - \rho^*$$

The noise fit is the difference in Sharpe ratio on the in-sample Sharpe ratio between $\hat{\theta}$ and θ^*. Since $\hat{\theta}$ is optimized on the in-sample data set and θ^* is the optimal parameter applied to the out-of-sample data, it results in a difference.
The estimation error is again the difference between $\hat{\theta}$ and θ^*, but then on the out-of-sample Sharpe ratio.
The noise is caused between the difference of the estimated mean returns $\hat{\mu}$ and the true mean returns μ, and the difference between the estimated covariance matrix $\hat{\Sigma}$ and the true covariance matrix Σ. Since there was assumed that the true covariance matrix was known (and full of rank), the noise is the difference in the estimated mean return $\hat{\mu}$ and the true mean return μ.

Since $\hat{\mu} = \mu + \nu$ where $\nu \sim \mathcal{N}(0, \frac{1}{T}\Sigma)$, it follows that

$$\mathrm{E}[\hat{\mu}] = \mathrm{E}[\mu + \nu] = \mathrm{E}[\mu] + \mathrm{E}[\nu] = \mathrm{E}[\mu] = \mu.$$

So the expected value of the noise is 0. Furthermore, the noise fit and estimation error have the same expected value. Let $k \geq 1$, then it holds [4]

$$\mathrm{E}[\hat{\rho} - \rho^*] = \mathrm{E}[\tau^* - \hat{\tau}] = \mathrm{E}[\frac{k}{2T\hat{\rho}}].$$

Therefore, the Sharpe Ratio Information Criterion (SRIC) is an unbiased estimator of the expected out-of-sample Sharpe ratio $\mathrm{E}[\hat{\tau}]$.

$$\hat{\tau} \approx SRIC = \hat{\rho} - \frac{k}{T\hat{\rho}}$$

Therefore,

$$\mathrm{E}[\hat{\tau}] = \mathrm{E}[\hat{\rho} - \frac{k}{T\hat{\rho}}].$$

The out-of-sample Sharpe ratio is now expressed in terms of $\hat{\mu}$ instead of μ.

Given that $\hat{\mu}$ and Σ are known, the only remaining requirement is to determine the optimal weight of stocks $\hat{\theta}$ in order to calculate the out-of-sample Sharpe ratio. Since $\hat{\theta}$ is the optimal parameter of the in-sample Sharpe ratio, so $\rho(\theta)$ should be maximized over parameter space Θ.

$$\frac{d}{d\theta}\rho(\theta) = \frac{d}{d\theta}\frac{\hat{\mu}^T\theta}{\sqrt{\theta^T\Sigma\theta}}$$

Since the logarithm function is a monotonically increasing function, maximizing $\log(\rho(\theta))$ results in the same $\hat{\theta}$ as maximizing $\rho(\theta)$.

$$\log(\rho(\theta)) = \log(\frac{\hat{\mu}^T\theta}{\sqrt{\theta^T\Sigma\theta}}) = \log(\hat{\mu}^T\theta) - \log(\sqrt{\theta^T\Sigma\theta})$$

Taking the derivative with respect to θ holds

$$\frac{d}{d\theta}\log(\rho(\theta)) = \frac{1}{\hat{\mu}^T\theta}\hat{\mu} - \frac{1}{\sqrt{\theta^T\Sigma\theta}}\frac{1}{2}(\theta^T\Sigma\theta)^{-\frac{1}{2}}2\Sigma\theta = \frac{\hat{\mu}}{\hat{\mu}^T\theta} - \frac{\Sigma\theta}{\theta^T\Sigma\theta}. \tag{2}$$

$$\frac{\hat{\mu}}{\hat{\mu}^T\theta} - \frac{\Sigma\theta}{\theta^T\Sigma\theta} = 0 \iff \hat{\mu} = \Sigma\theta$$

Since Σ is full of rank, it is invertible. Therefore the optimal parameter $\hat{\theta} \in \Theta$ for the maximum in-sample Sharpe ratio is equal to $\hat{\theta} = \Sigma^{-1}\hat{\mu}$[1].

The computation of the out-of-sample Sharpe ratio is now feasible due to the availability of known values such as $\hat{\mu}$, obtained by averaging the returns of each stock, and the assumption that Σ is known. $\hat{\theta}$ can now be computed as well and since T and k are known, the calculation of the out-of-sample Sharpe ratio becomes straightforward.
However, in reality, the true Σ is unknown, so the Sharpe Ratio Information Criterion cannot be applied correctly [2]. In many cases it is assumed that Σ is equal to $\hat{\Sigma}$, since $\hat{\Sigma}$ can be computed from the data.
Let $\boldsymbol{X}_T = (X_1, X_2, ..., X_T)$ be a sample of size T, then

$$\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} X_i \qquad \hat{\Sigma} = \frac{1}{T-1}\sum_{i=1}^{T}(X_i - \hat{\mu})(X_i - \hat{\mu})^T.$$

Note that if we consider n stocks, $\hat{\mu}$ is a vector of length n and $\hat{\Sigma}$ is a $n \times n$ matrix.
Hence, it is important to evaluate the Sharpe ratio in terms of the estimated mean of returns $\hat{\mu}$ and the estimated covariance matrix $\hat{\Sigma}$, as the bias is different than of the Sharpe Ratio Information Criterion. Since this bias is difficult to determine analytically, a simulation-based method is used in the next chapter to determine a bias correction for the Sharpe ratio in terms of $\hat{\mu}$ and $\hat{\Sigma}$.

2.3 Akaike Information Criterion

Before we are going to estimate the bias for the out-of-sample Sharpe ratio in terms of $\hat{\mu}$ and $\hat{\Sigma}$, we need to look at the Akaike Information Criterion first. Therefore, we need the Kullback-Leiber information (henceforth referred to as "K-L information"). The K-L information is a measure of how one probability distribution $G(x)$ is different from a another probability distribution $F(x)$. The K-L information is given by

$$I(G;F) = E_G\left[\log(\frac{G(X)}{F(X)})\right],$$

where E_G represents the expectation with respect to the probability distribution G. Denote the density functions of $G(X)$ and $F(X)$ as $g(x)$ and $f(x)$ respectively.
If the probability distribution functions are continuous models with density function $g(x)$ and $f(x)$, the K-L information is

$$I(g;f) = \int_{-\infty}^{\infty}\log(\frac{g(x)}{f(x)})g(x)dx.$$

And if the probability distribution functions are discrete models with the probability for each x_i is $g(x_i)$ and $f(x_i)$ for $i = 1, 2, ...$. Then the K-L information is given by

$$I(g;f) = \sum_{i=1}^{\infty} g(x_i)\log(\frac{g(x_i)}{f(x_i)}).$$

The K-L information has the following properties:

(i) $I(g;f) \geq 0$
(ii) $I(g;f) = 0 \iff g(x) = f(x)$

[1] In portfolio optimization you'll need to normalize this vector such that $||\hat{\theta}|| = 1$.

[2] In large data sets, the estimated $\hat{\Sigma}$ is often used as the true Σ because the difference is not significant. The uncertainty of the estimated value of $\hat{\mu}$ is much larger than that of the estimated $\hat{\Sigma}$.

So the smaller the K-L is, the closer the model $f(x)$ is to $g(x)$. In this case we assumed that both $g(x)$ and $f(x)$ are known, however if we are given a data set $\mathbf{x}_n = (x_1, x_2, ..., x_n)$ from an unknown probability distribution $G(X)$, the K-L information cannot be calculated directly.
Since a fraction in the logarithmic function can be written in two terms, the K-L information can also be written as

$$I(g; f) = E_G\left[\log(\frac{g(X)}{f(X)})\right] = E_G[\log g(X)] - E_G[\log f(X)].$$

Note that the first-term on the right-hand side only depends on the unknown true distribution g, so it's a constant. Since we want the K-L as small as possible, we want the second term on the right-hand side as large as possible. That term is called the expected log-likelihood, the expected log-likelihood can be expressed as

$$E_G[\log f(X)] = \int \log f(x) dG(x).$$

For continuous models this can be expressed as

$$E_G[\log f(X)] = \int_{-\infty}^{\infty} g(x) \log f(x) dx,$$

and for discrete models as

$$E_G[\log f(X)] = \sum_{i=1}^{\infty} g(x_i) \log f(x_i).$$

To maximize the log likelihood, we do need the following. Let $f(x|\theta)$ be the candidate model with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)^T (\boldsymbol{\theta} \in \Theta \subset R^k)$, where k is the number of parameters. For a given data set $\mathbf{x}_n = x_1, x_2, ..., x_n$ the log-likelihood can be determined as a function of $\boldsymbol{\theta} \in \Theta$ given by

$$\ell(\boldsymbol{\theta}) = \sum_{\alpha=1}^{n} \log f(x_\alpha|\boldsymbol{\theta}).$$

Let $\hat{\boldsymbol{\theta}}$ be the parameter that satisfies

$$\ell(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}),$$

then $\hat{\boldsymbol{\theta}}$ is called the maximum likelihood estimator and $\ell(\hat{\boldsymbol{\theta}}) = \sum_{\alpha=1}^{n} \log f(x_\alpha|\hat{\boldsymbol{\theta}})$ is known as the maximum log-likelihood.

Since the K-L distance can't be computed without full knowledge of g and the parameters in each of the candidate models $f_i(x|\theta)$, Hirotogu Akaike (1973) found a way to estimate the K-L information based on the empirical maximum log-likelihood. Akaike demonstrated that estimating $E_y E_x[\log(f(x|\hat{\theta}(y)))]$ is a critical aspect in obtaining an applied K-L model selection criterion, where x and y are independent random samples from the same distribution and both expectations are taken with respect to the truth g [1]. This double expectation is the target of all model selection approaches, based on K-L information. The approximately unbiased estimator of $E_y E_x[\log(f(x|\hat{\theta}(y)))]$ for large samples and "good" models is

$$\ell(\hat{\boldsymbol{\theta}}) - k,$$

where $\ell(\hat{\boldsymbol{\theta}})$ is the maximum log-likelihood and k the number of parameters as mentioned before. This results in the Akaike Information Criterion (AIC), which goal is to maximize the fit on the out-of-sample data, given by

$$AIC = 2k - 2\ell(\hat{\boldsymbol{\theta}}). \tag{3}$$

Like the K-L information, the lower the AIC, the better the fit. The AIC deals with the trade-off between the goodness of fit of the model and the simplicity of the model, it penalizes the number of parameters in the model. Thus, rather having a simple measure of the directed distance between two models (K-L distance), we have an estimate of the expected, relative distance between the fitted candidate model and the unknown true model.
Note that we assumed that $\ell(\hat{\boldsymbol{\theta}}) - k$ is an unbiased estimator of $E_y E_x[\log(f(x|\hat{\theta}(y)))]$ for large sample sizes. However , when the sample is small, there is a substantial probability that AIC will select models that have too many parameters, i.e. that AIC will over-fit. Therefore, for small sample sizes, the Akaike Information Criterion corrected (AICc) was developed. Assume that the candidate model is univariate, linear in its parameters and has normally distributed residuals, the formula for AICc is as follows [2]

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1}.$$

where n is the sample size and k denotes the number of parameters. The AICc provides an additional penalty to the number of parameters in the model, therefore for small sample sizes the AICc is more accurate. Note that as $n \to \infty$, the extra penalty term converges to 0. So AICc $\to$ AIC as $n \to \infty$. In this case with these assumptions, the extra penalty term could be determined analytically. In particular, with other assumptions, bootstrap estimation of the formula is often feasible. The bootstrap method will be used as well in the next chapter to determine a bias correction for Sharpe ratio in terms of $\hat{\mu}$ and $\hat{\Sigma}$.

3 The Boostrap Method

Similar to the AIC and the AICc penalises the fit of the model on the out-of-sample data by the number of parameters, the SRIC penalises the out-of-sample Sharpe ratio by the number of parameters as well. However, we want to find a bias for the Sharpe ratio without the assumption of a known covariance matrix. To this end we will use the bootstrap method. The bootstrap method has the ability to provide effective solutions to problems that cannot be solved by analytic approaches based on theories or formulas. The bootstrap method is a resampling technique based on the idea to create multiple resamples by drawing observations from the original data set with replacement.
Let $\boldsymbol{X}_T = (X_1, X_2, ..., X_T)$ be a random sample of size T drawn from an unknown probability function $G(x)$. We estimate the parameter θ with respect to the probability distribution $G(x)$ by using an estimator $\hat{\theta} = \hat{\theta}(\boldsymbol{X}_T)$. Let $\boldsymbol{x}_T = (x_1, x_2, ..., x_T)$ be the observed data and let $\hat{\theta} = \hat{\theta}(\boldsymbol{x}_T)$ be the estimator of θ. The bias and variance of an estimator are essential quantities used to evaluate the error in estimation. The bias and variance of the estimator are given by

$$b(G) = E_G[\hat{\theta}] - \theta, \qquad \sigma^2(G) = E_G[(\hat{\theta} - E_G[\hat{\theta}])^2].$$

Both the bias and variance depends on the true unknown probability distribution $G(x)$. Instead of estimating these quantities analytically, we will estimate them using the bootstrap method. The bootstrap method can be executed through the following steps :

(**1**) Generate a new sample by randomly selecting observations from the original data set with allowing replacement, i.e , choose T observations from the original data set with equal probability $\frac{1}{T}$ at each point of the observations. This probability distribution function is called the empirical distribution function $\hat{G}(x)$ and we can estimate the unknown probability distribution $G(x)$ from $\hat{G}(x)$.

(**2**) Random samples of the empirical distribution function $\hat{G}(x)$ are called the bootstrap samples and are denoted as $\boldsymbol{X}_T^* = (X_1^*, X_2^*, ..., X_T^*)$. The estimator based on a bootstrap sample is denoted as $\hat{\theta}^* = \hat{\theta}(\boldsymbol{X}_n^*)$. The estimation of the bias and variance of the estimator becomes

$$b(\hat{G}) = E_{\hat{G}}[\hat{\theta}^*] - \hat{\theta}, \qquad \sigma^2(\hat{G}) = E_{\hat{G}}[(\hat{\theta}^* - E_{\hat{G}}[\hat{\theta}^*])^2].$$

where this estimation depends on the empirical distribution function $\hat{G}(x)$. The expressions of $b(\hat{G})$ and $\sigma^2(\hat{G})$ are referred to as the bootstrap estimates of $b(G)$ and $\sigma^2(G)$, respectively.

(**3**) Assume that we made B bootstrap samples, denote $\boldsymbol{X}_T^*(i) = (X_1^*(i), X_2^*(i), ..., X_T^*(i))$ as the i^{th} bootstrap sample. The bootstrap estimates $b(\hat{G})$ and $\sigma^2(\hat{G})$ are numerically approximated by using the Monte Carlo method. For finding the correction for the Sharpe ratio, we are interested in the bootstrap estimate of the bias. Let the corresponding estimator of the i-th bootstrap sample be denoted as $\hat{\theta}^*(i) = \hat{\theta}(\boldsymbol{X}_T^*(i))$. Then, the bootstrap estimate of $b(\hat{G})$ is approximated as

$$b(\hat{G}) \approx \frac{1}{B}\sum_{i=1}^{B} \hat{\theta}^*(i) - \hat{\theta}.$$

For so far we have only looked at the parameter θ itself, but the Sharpe ratio is a function of the parameter $\hat{\theta}$. Let f be the Sharpe ratio function that depends on $\hat{\theta}$, assuming the covariance matrix Σ is known. Then f is given by [3]

$$f(\boldsymbol{X}_T|\hat{\theta}) = \frac{\hat{\mu}^T\hat{\theta}}{\sqrt{\hat{\theta}^T\Sigma\hat{\theta}}} \qquad , \hat{\theta} = \Sigma^{-1}\hat{\mu}. \tag{4}$$

The bias is then given by

$$b(G) = E_{G(\boldsymbol{x})}\left[\sum_{\alpha=1}^{T} f(X_\alpha|\hat{\boldsymbol{\theta}}(\boldsymbol{X}_T)) - TE_{G(z)}\left[f(Z|\hat{\boldsymbol{\theta}}(\boldsymbol{X}_T))\right]\right], \tag{5}$$

where $E_{G(\boldsymbol{x})}$ denotes the expectation with respect to the joint distribution of a random sample $\boldsymbol{X}_T$, and the second term on the right-hand side of equation 5, $E_{G(z)}\left[f(Z|\hat{\boldsymbol{\theta}}(\boldsymbol{X}_T))\right]$, represents the expectation with respect

to the distribution $G(z)$ of the future data z that is independent of the random sample $\boldsymbol{X}_T$.

Let μ and Σ be respectively the true mean returns and covariance matrix of the stocks. Let $\bar{\theta}$ be the optimal parameter based on $\hat{\mu}$ and $\hat{\Sigma}$ by $\bar{\theta} = \hat{\Sigma}^{-1}\hat{\mu}$. This is different than $\hat{\theta}$ because $\hat{\theta} = \Sigma^{-1}\hat{\mu}$ where Σ is known[3]. Then the bias for the Sharpe ratio in terms of the estimated mean returns and the estimated covariance matrix is

$$b(G) = \frac{\hat{\mu}^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}\bar{\theta}}} - \frac{\mu^T\bar{\theta}}{\sqrt{\bar{\theta}^T\Sigma\bar{\theta}}}. \tag{6}$$

Recall that both μ and Σ are unknown, so we cannot compute this bias directly. However, we can use bootstrap estimation to approximate this bias.
To approximate this bias, the true distribution $G(x)$ is replaced with an empirical distribution function $\hat{G}(x)$. The following are substituted as well :

$$G(x) \to \hat{G}(x)$$
$$X_\alpha \sim G(x) \to X^*_\alpha \sim \hat{G}(x)$$
$$Z \sim G(z) \to Z^* \sim \hat{G}(z)$$
$$E_{G(\boldsymbol{x})}, E_{G(z)} \to E_{\hat{G}(\boldsymbol{x}^*)}, E_{\hat{G}(z^*)}$$
$$\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}(\boldsymbol{X}) \to \bar{\boldsymbol{\theta}}^* = \bar{\boldsymbol{\theta}}^*(\boldsymbol{X}^*)$$

Therefore, the bootstrap bias estimate of equation 5 becomes

$$b^*(\hat{G}) = E_{\hat{G}(\boldsymbol{x}^*)}\left[\sum_{\alpha=1}^{T} f(X^*_\alpha|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)) - TE_{\hat{G}(z^*)}\left[f(Z^*|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T))\right]\right]. \tag{7}$$

Let $\boldsymbol{x}_T = (x_1, x_2, ..., x_T)$ be the given data set. Note that if the empirical distribution function is considered as the true distribution, then

$$\begin{aligned} E_{\hat{G}(z)}\left[f(Z|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T))\right] &= \int f(z|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T))d\hat{G}(z) \\ &= \frac{1}{T}\sum_{\alpha=1}^{T} f(x_\alpha|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)) \\ &\equiv \frac{1}{T} f(\boldsymbol{x}_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)). \end{aligned}$$

So for the second term on the right-hand side, we have

$$TE_{\hat{G}(z^*)}\left[f(Z^*|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T))\right] \equiv f(\boldsymbol{x}_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)). \tag{8}$$

On the other-hand, based on reusing the bootstrap sample $\boldsymbol{X}^*_T$, we have

$$\begin{aligned} E_{\hat{G}^*(z)}[f(Z|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T))] &= \int f(z|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*))d\hat{G}^*(z) \\ &= \frac{1}{T}\sum_{\alpha=1}^{T} f(X^*_\alpha|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)) \\ &\equiv \frac{1}{T} f(\boldsymbol{X}^*_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)). \end{aligned}$$

Therefore,

$$\sum_{\alpha=1}^{T} f(X^*_\alpha|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)) \equiv f(\boldsymbol{X}^*_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)). \tag{9}$$

Combining equation 8 and equation 9 into equation 7, the bootstrap bias estimate can be written as

$$b^*(\hat{G}) = E_{\hat{G}(\boldsymbol{x}^*)}\left[f(\boldsymbol{X}^*_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T)) - f(\boldsymbol{X}_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}^*_T))\right], \tag{10}$$

[3]The same method as 2, but then with $\hat{\Sigma}$ instead of Σ

where $\hat{G}$ is a known probability distribution (the empirical distribution function).

Let us generate B sets of bootstrap samples, each of size T and denote the i^{th} bootstrap sample as $\boldsymbol{X}_T^*(i) = (X_1^*(i), X_2^*(i), ..., X_T^*(i))$. Denote the difference of equation 8 and equation 9 with respect to the sample $\boldsymbol{X}_T^*(i)$ as

$$D^*(i) f(\boldsymbol{X}_T^*(i)|\bar{\boldsymbol{\theta}}(\boldsymbol{X}_T^*(i))) - f(\boldsymbol{x}_T|\bar{\boldsymbol{\theta}}(\boldsymbol{X}_T^*(i))), \tag{11}$$

where $\bar{\boldsymbol{\theta}}(\boldsymbol{X}_T^*(i))$ is an estimate of $\bar{\boldsymbol{\theta}}$ obtained from the i^{th} bootstrap sample. Then the expectation in equation 10 based on B bootstrap samples can be numerically approximated as

$$b^*(\hat{G}) \approx \frac{1}{B} \sum_{i=1}^{B} D^*(i) \equiv b_B(\hat{G}), \tag{12}$$

where $b_B(\hat{G})$ is the bootstrap estimate of the bias $b(G)$ of the Sharpe ratio. So with this bootstrap estimation to approximate the bias, we are able to estimate the out-of-sample Sharpe ratio. Denote the Sharpe Ratio Information Criterion corrected as $SRIC_c$ and

$$SRIC_c \equiv \frac{\hat{\mu}^T \bar{\theta}}{\sqrt{\bar{\theta}^T \hat{\Sigma} \bar{\theta}}} - b_B(\hat{G}), \tag{13}$$

Therefore,

$$\bar{\tau} \approx \frac{\hat{\mu}^T \bar{\theta}}{\sqrt{\bar{\theta}^T \hat{\Sigma} \bar{\theta}}} - b_B(\hat{G}). \tag{14}$$

.

For unknown μ and unknown Σ, the replacements are given by

$$G(x) \to \hat{G}(x)$$

$$X_\alpha \sim G(x) \to X_\alpha^* \sim \hat{G}(x)$$

$$\mu \to \hat{\mu}$$

$$\hat{\mu} \to \hat{\mu}^*$$

$$\Sigma \to \hat{\Sigma}$$

$$\hat{\Sigma} \to \hat{\Sigma}^*$$

$$\bar{\theta} \to \bar{\theta}^*$$

This means that considering the empirical distribution $\hat{G}(x)$ as the true distribution, the estimated $\hat{\mu}$ is considered as the true mean of returns. So where in the 'real' world μ is unknown, in the empirical distribution it is known and it is equal to $\hat{\mu}$. Similarly, when it comes to the covariance matrix, the estimated covariance matrix is treated as if it were the actual covariance matrix for the empirical distribution. Therefore for the empirical distribution we have the true mean of returns as $\hat{\mu}$, the estimated mean of returns as $\hat{\mu}^*$, the true covariance matrix as $\hat{\Sigma}$ and the estimated covariance matrix as $\hat{\Sigma}^*$. For every bootstrap sample $\boldsymbol{X}^*(i)$, we can determine $\hat{\mu}^*(i)$, $\hat{\Sigma}^*(i)$ and $\bar{\theta}^*(i)$ for every $i = 1, ..., B$. Then $D^*(i)$ can be calculated by

$$D^*(i) = \frac{\hat{\mu}^*(i)^T \bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T \hat{\Sigma}^*(i) \bar{\theta}^*(i)}} - \frac{\hat{\mu}^T \bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T \hat{\Sigma} \bar{\theta}^*(i)}}.$$

Let's look at an example to provide a clearer understanding of this concept.

Example
Consider the monthly returns of 2 stocks over the past year.

Stock 1	Stock 2
0.05	-0.03
0.01	0.06
0.03	-0.04
0.05	0.01
-0.06	0.07
0.01	-0.03
0.03	0.04
-0.01	-0.02
0.04	0.04
-0.01	0.02
0.01	0.08
0.03	0.05

Now we can determine $\hat{\mu}, \hat{\Sigma}$ and $\bar{\theta}$, where $\bar{\theta} = \hat{\Sigma}^{-1}\hat{\mu}$.

$$\hat{\mu} = (0.015, 0.021)^T \qquad \hat{\Sigma} = \begin{pmatrix} 0.0009727273 & -0.0004318182 \\ -0.0004318182 & 0.0017901515 \end{pmatrix} \qquad \bar{\theta} = \hat{\Sigma}^{-1}\hat{\mu} = (0.57, 0.43)^T$$

The in-sample Sharpe ratio with the unknown covariance matrix is

$$\frac{\hat{\mu}^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}\bar{\theta}}} = \frac{0.01749}{0.02184} = 0.839.$$

Now let's generate $B = 1000$ bootstrap samples[4]. For every sample, we can determine $\hat{\mu}^*(i)$, $\hat{\Sigma}^*(i)$ and $\bar{\theta}^*(i)$.

For the first bootstrap sample $\boldsymbol{X}^*(1)$ we get

$\boldsymbol{X}^*(1) =$

Stock 1	Stock 2
0.01	0.08
0.04	0.04
0.04	0.04
-0.01	0.02
-0.01	0.02
0.03	-0.04
0.01	0.08
0.03	0.04
0.04	0.04
0.05	0.01
0.04	0.04
0.04	0.04

with

$$\hat{\mu}^*(1) = (0.026, 0.034)^T \qquad \hat{\Sigma}^*(1) = \begin{pmatrix} 0.0004265152 & -0.00008106061 \\ -0.00008106061 & 0.0009901515 \end{pmatrix} \qquad \bar{\theta}^*(1) = (0.63, 0.37)^T$$

Then

$$D^*(1) = \frac{\hat{\mu}^*(1)^T\bar{\theta}^*(1)}{\sqrt{\bar{\theta}^*(1)^T\hat{\Sigma}^*(1)\bar{\theta}^*(1)}} - \frac{\hat{\mu}^T\bar{\theta}^*(1)}{\sqrt{\bar{\theta}^*(1)^T\hat{\Sigma}\bar{\theta}^*(1)}} = 1.77 - 0.83 = 0.94.$$

Repeating this 999 times gives us

$$\frac{1}{1000}\sum_{i=1}^{1000} D^*(i) = 0.341.$$

[4]If the number B of bootstrap samples becomes infinitely large, errors in the approximation by Monte Carlo simulation can be ignored. In this paper we will use B = 1000.

Therefore, the bootstrap estimate of the bias $b(G)$ is approximately 0.341. The out-of-sample Sharpe ratio of the 2 stocks is estimated as

$$\bar{\tau} \approx \frac{\hat{\mu}^T \bar{\theta}}{\sqrt{\bar{\theta}^T \hat{\Sigma} \bar{\theta}}} - b_B(\hat{G}) = 0.839 - 0.341 = 0.498.$$

3.1 Bias for the Sharpe Ratio Information Criterion

We can estimate the bias of the Sharpe ratio using an unknown covariance matrix as well as with a known covariance matrix. In the case of an known covariance matrix, the bias can be estimated analytically as well, so we are able to compare both results. However, when the covariance matrix is known, we need to simulate data based on that covariance matrix. To accomplish this, we consider an n dimensional multivariate normal distribution and require the true mean, denoted as μ. Therefore, we are able to determine the real bias as well, as μ and Σ are known. By comparing the actual bias, the analytically estimated bias, and the bias estimated through bootstrapping, we can evaluate and compare the results.

Let $\boldsymbol{X} \sim N(\mu, \Sigma)$ be an n dimensional multivariate normal distribution where $\mu \in R^n$ and $\Sigma \in R^{n \times n}$ is a symmetric, positive definite $n \times n$ matrix, so

$$\Sigma^T = \Sigma \qquad \text{and} \qquad x^T \Sigma x > 0 \quad \forall x \in R^n \backslash \{\mathbf{0}\}.$$

Now we able to simulate the returns of n components using the *mvrnorm* package in R with a given T. Remember that we assumed that Σ is known, so the replacements for the bootstrap simulation is as follows :

$$G(x) \to \hat{G}(x)$$

$$X_\alpha \sim G(x) \to X_\alpha^* \sim \hat{G}(x)$$

$$\mu \to \hat{\mu}$$

$$\hat{\mu} \to \hat{\mu}^*$$

$$\Sigma \to \Sigma$$

$$\hat{\theta} \to \hat{\theta}^*$$

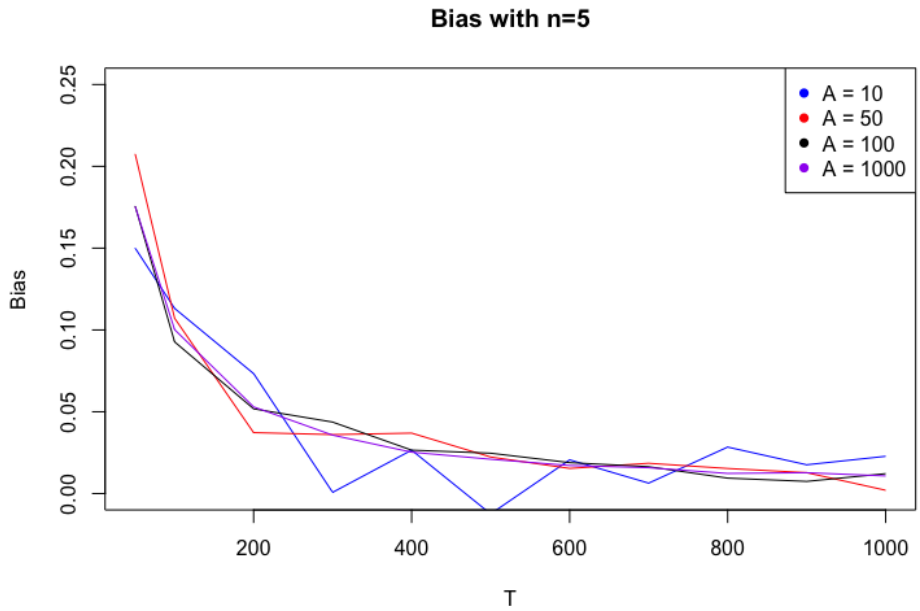
Let $\hat{\theta} = \Sigma^{-1}\hat{\mu}$ and $\hat{\theta}^*(i) = \Sigma^{-1}\hat{\mu}^*(i)$ where $\hat{\mu}^*(i)$ is the vector of the mean returns of the $i-th$ bootstrap sample. We can determine the true bias $b(G)$, the bias analytically $b_A(G)$ and the bootstrap bias by

$$b(G) = \frac{\hat{\mu}^T \hat{\theta}}{\sqrt{\hat{\theta}^T \Sigma \hat{\theta}}} - \frac{\mu^T \hat{\theta}}{\sqrt{\hat{\theta}^T \Sigma \hat{\theta}}} = \hat{\rho} - \hat{\tau} \tag{15}$$

$$b_A(G) = \frac{k}{T\hat{\rho}} \quad , \quad k = n - 1 \tag{16}$$

$$b_B(\hat{G}) = \frac{1}{B} \sum_{i=1}^{B} D^*(i) \quad , \quad D^*(i) = \frac{\hat{\mu}^*(i)^T \hat{\theta}^*(i)}{\sqrt{\hat{\theta}^*(i)^T \Sigma \hat{\theta}^*(i)}} - \frac{\hat{\mu}^T \hat{\theta}^*(i)}{\sqrt{\hat{\theta}^*(i)^T \Sigma \hat{\theta}^*(i)}} \tag{17}$$

Because the bootstrap method is based on simulated data from a multivariate normal distribution, we also need to consider the randomness inherent in the simulated data itself. Therefore, since there is randomness in the simulated data, we will need to repeat the entire bootstrap process multiple times. Let A be the number of simulated multivariate normally distributed samples, and B be the number of bootstrap samples per simulated data. Errors in the approximation gets smaller if B becomes infinitely large, so we will set $B = 1000$ in this study. To determine A, we should look at the graph of the true bias for different A with different T. Consider the data simulated by μ and Σ described in Table 6.1. For different A we obtain



When comparing the results, we observe that for $A = 10$ and $A = 50$, the bias line still fluctuates significantly compared to $A = 100$ and $A = 1000$. Since it takes a considerable amount of time to generate 1000 bootstrap samples for each data set in 1000 simulated data sets, we will now choose $A = 100$.

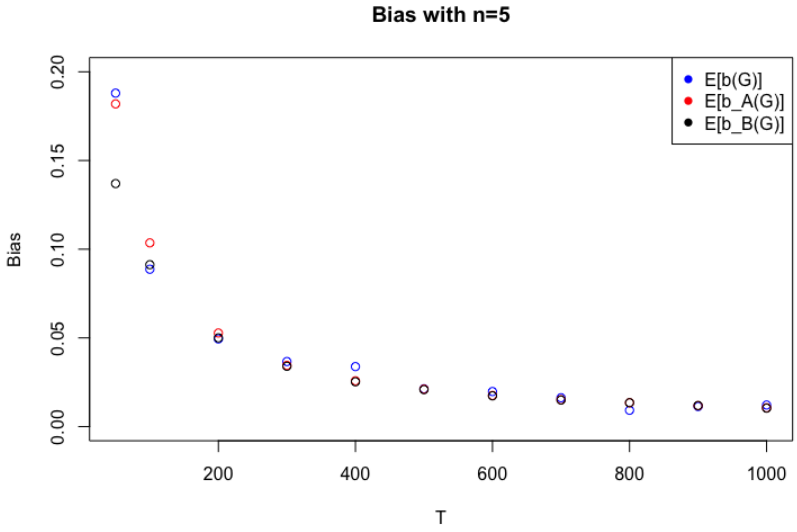
For given μ and Σ, then we have $\theta^* = \Sigma^{-1}\mu$. The out-of-sample Sharpe ratio of optimal parameter applied to out-of-sample data is

$$\tau^* = \tau(\theta^*) = \frac{\mu^T \theta^*}{\sqrt{\theta^* \Sigma \theta^*}}. \tag{18}$$

Let $n = 5$, for these simulations we will set $A = 100$ and $B = 1000$. The vector μ and matrix Σ are given in Table 6.1 and $\tau(\theta^*) = 0.373$. For different T we have

T	50	100	500	1000
$E[b(G)]$	0.18798357	0.08870314	0.02130247	0.01215171
$E[b_A(G)]$	0.18190248	0.10356744	0.02111789	0.01059187
$E[b_B(\hat{G})]$	0.13698760	0.09127060	0.02083634	0.01038829
$E[b_A(G) - b(G)]$	-0.0060810913	0.0148642926	-0.0001845812	-0.0015598427
$E[b_B(G) - b(G)]$	-0.0509959701	0.0025674580	-0.0004661327	-0.0017634268
$Var(b(G))$	0.0121420539	0.0096372878	0.0020956847	0.0008120217
$Var(b_A(G) - b(G))$	0.0246147640	0.0156246108	0.0023272278	0.0008572548
$Var(b_B(\hat{G}) - b(G))$	0.0163976531	0.0134063846	0.0023026829	0.0008634795

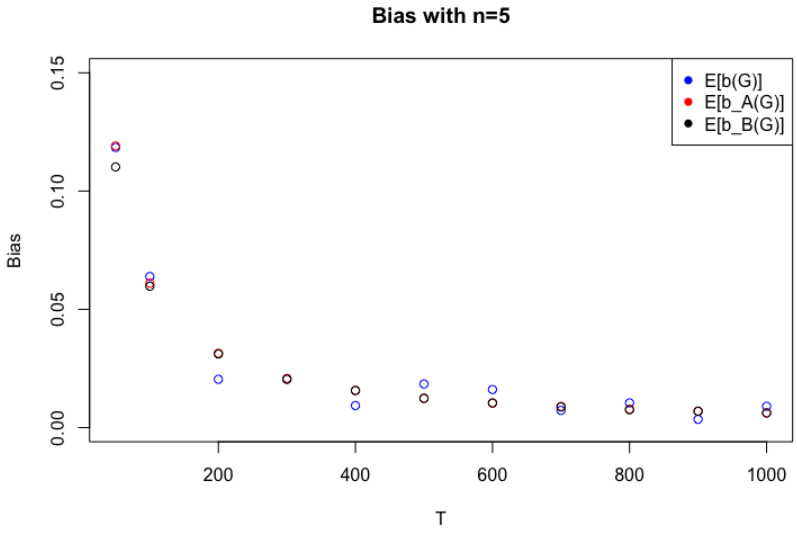
The graph of the expectations of the true bias, the bias determined analytically and the bootstrap estimates of $b(G)$ for different T is



Now with $n = 5$ again, but $\tau(\theta^*) = 0.639$ with μ and Σ given in Table 6.2 , we get

T	50	100	500	1000
$E[b(G)]$	0.118449198	0.063864710	0.018439971	0.008999992
$E[b_A(G)]$	0.119081434	0.061031618	0.012340676	0.006206526
$E[b_B(\hat{G})]$	0.110203248	0.059778875	0.012384546	0.006191265
$E[b_A(G) - b(G)]$	0.000632236	-0.002833092	-0.006099294	-0.002793466
$E[b_B(G) - b(G)]$	-0.008245950	-0.004085835	-0.006055425	-0.002808728
$Var(b(G))$	0.0170670606	0.0092963760	0.0019834704	0.0008156174
$Var(b_A(G) - b(G))$	0.0232022581	0.0113535830	0.0020607965	0.0008313181
$Var(b_B(\hat{G}) - b(G))$	0.0218060986	0.0110296799	0.0020538077	0.0008269363

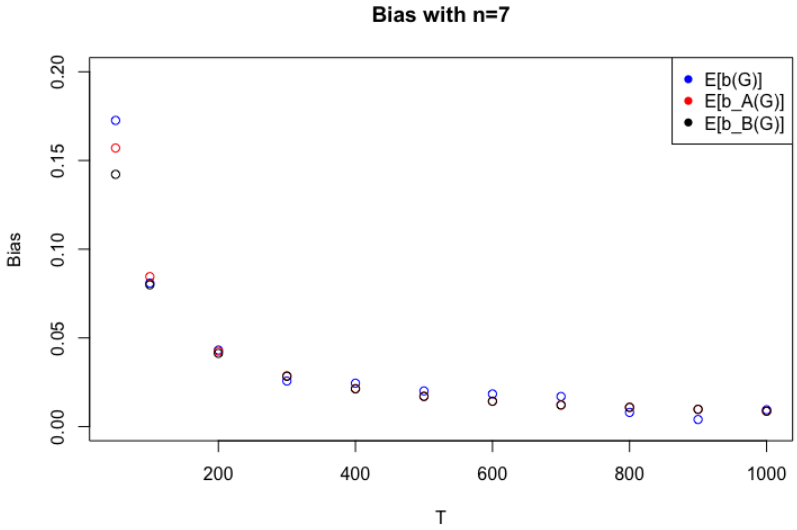
The graph for different T is



Let $n = 7$. The vector μ and matrix Σ given in Table 6.3. Here $\tau(\theta^*) = 0.692$ and we have

T	50	100	500	1000
$E[b(G)]$	0.172598863	0.080858114	0.020078872	0.009452923
$E[b_A(G)]$	0.157033706	0.084548507	0.017155271	0.008627436
$E[b_B(\hat{G})]$	0.142190753	0.079895199	0.016972179	0.008727723
$E[b_A(G) - b(G)]$	-0.0155651561	0.0036903936	-0.0029236010	-0.0008254863
$E[b_B(G) - b(G)]$	-0.0304081100	-0.0009629150	-0.0031066924	-0.0007251995
$Var(b(G))$	0.0168894399	0.0111699971	0.0020454595	0.0008012633
$Var(b_A(G) - b(G))$	0.0247100195	0.0143077051	0.0021504106	0.0008208858
$Var(b_B(\hat{G}) - b(G))$	0.0220951196	0.0138568447	0.0021405644	0.0008185002

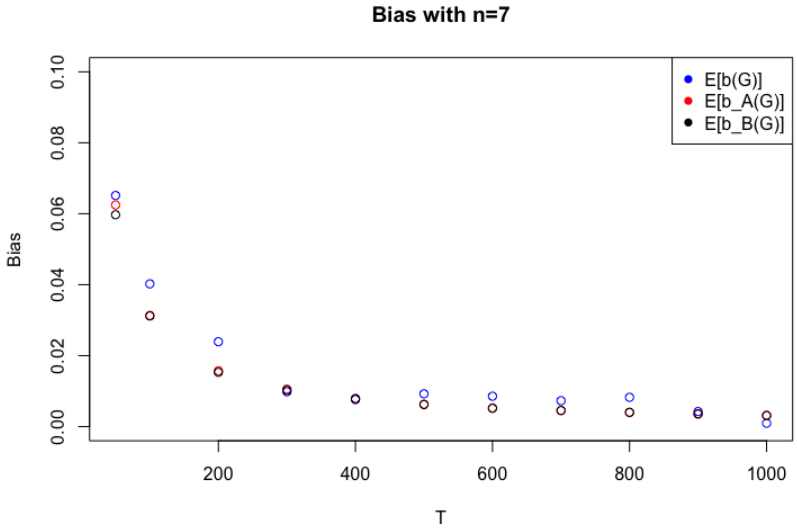
For different T we have



Now for the same $n = 7$, but $\tau(\theta^*) = 1.893$ with different μ and Σ as in Table 6.4.

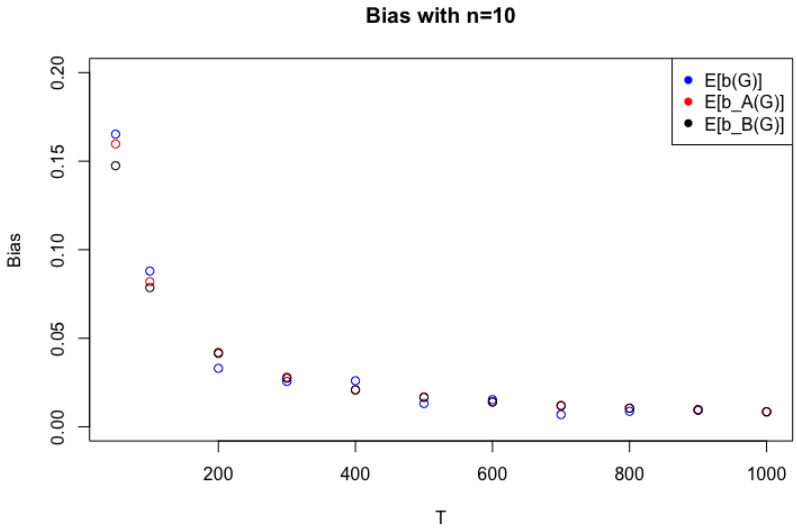
T	50	100	500	1000
$E[b(G)]$	0.0651319261	0.0402263691	0.0092300500	0.0009463701
$E[b_A(G)]$	0.062475119	0.031342895	0.006322203	0.003171175
$E[b_B(\hat{G})]$	0.059707624	0.031169910	0.006187428	0.003110910
$E[b_A(G) - b(G)]$	-0.002656807	-0.008883474	-0.002907847	0.002224805
$E[b_B(G) - b(G)]$	-0.005424302	-0.009056459	-0.003042622	0.002164539
$Var(b(G))$	0.016553071	0.010870797	0.002257263	0.001034536
$Var(b_A(G) - b(G))$	0.017675174	0.011232939	0.002272195	0.001038016
$Var(b_B(\hat{G}) - b(G))$	0.017685806	0.011325991	0.002255490	0.001040529

For different T we have



Last, we set $n = 10$ with μ and Σ given in Table 6.5 , $\tau(\theta^*) = 1.062$ and for different T we have

T	50	100	500	1000
$E[b(G)]$	0.165234589	0.087959603	0.013164947	0.008566129
$E[b_A(G)]$	0.159739061	0.081930363	0.016876840	0.008440529
$E[b_B(\hat{G})]$	0.147442243	0.078610189	0.016561705	0.008409919
$E[b_A(G) - b(G)]$	-0.0054955275	-0.0060292402	0.0037118938	-0.0001256003
$E[b_B(G) - b(G)]$	-0.0177923459	-0.0093494139	0.0033967582	-0.0001562107
$Var(b(G))$	0.018172842	0.007373774	0.001895096	0.001048880
$Var(b_A(G) - b(G))$	0.024093572	0.008571411	0.001956823	0.001065460
$Var(b_B(\hat{G}) - b(G))$	0.023212075	0.008534129	0.001953808	0.001063178



As the number of observations, denoted by T, increases towards infinity, it appears that the bias and the bootstrap bias tends to converge towards 0. This convergence is in line with the analytic bias. Specifically, as $T \to \infty$, the ratio $\frac{k}{T\hat{\rho}} \to 0$. This indicates that the bias decreases as the sample size grows larger. So the more data samples we have, the smaller the difference between the in-sample Sharpe Ratio and the out-of-sample Sharpe Ratio. However we see that the expectation of the real bias oscillates every now and then. We can prevent this by increasing the value of A. However, this would require a significantly larger number of simulations, resulting in long computation times.

Moreover, the expectation of the difference between the true bias, represented by $b(G)$, and the bias estimated through bootstrapping, denoted as $b_B(\hat{G})$, also tends to approach 0 as the sample size increases. This suggests that the bootstrap bias becomes increasingly accurate and aligns closely with the true bias as T becomes larger.

3.2 Bias with Unknown Covariance Matrix

Since real data works with an unknown μ and Σ, we want to determine the bootstrap bias for $\hat{\mu}$ and $\hat{\Sigma}$. But first we will look at an situation where Σ is known, but proceed as if Σ was unknown. In this way we can compare the real bias $B(G)$ and the bootstrap estimate of the bias $b_G(\hat{G})$. Note that since Σ is unknown, we can not use $b_A(G) = \frac{k}{T\hat{\rho}}$ since it works only for a known Σ.

Denote $\phi(\theta)$ as the in-sample Sharpe ratio with the estimated covariance matrix $\hat{\Sigma}$ as

$$\phi(\theta) = \frac{\hat{\mu}^T\theta}{\sqrt{\theta^T\hat{\Sigma}\theta}}$$

and

$$\bar{\theta} \in \arg\max_{\theta\in\Theta} \phi(\theta) \qquad \Longleftrightarrow \qquad \bar{\theta} = \hat{\Sigma}^{-1}\hat{\mu}$$

So the in-sample Sharpe ratio of $\bar{\theta}$ is given by

$$\bar{\phi} = \phi(\bar{\theta}) = \frac{\hat{\mu}^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}\bar{\theta}}} \tag{19}$$

The replacements are given as follows

$$G(x) \to \hat{G}(x)$$
$$X_\alpha \sim G(x) \to X^*_\alpha \sim \hat{G}(x)$$
$$\mu \to \hat{\mu}$$
$$\hat{\mu} \to \hat{\mu}^*$$
$$\Sigma \to \hat{\Sigma}$$
$$\hat{\Sigma} \to \hat{\Sigma}^*$$
$$\bar{\theta} \to \bar{\theta}^*$$

We can determine the real bias $b(G)$ and the bootstrap bias $b_B(\hat{G})$ by

$$b(G) = \frac{\hat{\mu}^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}\bar{\theta}}} - \frac{\mu^T\bar{\theta}}{\sqrt{\bar{\theta}^T\Sigma\bar{\theta}}} \tag{20}$$

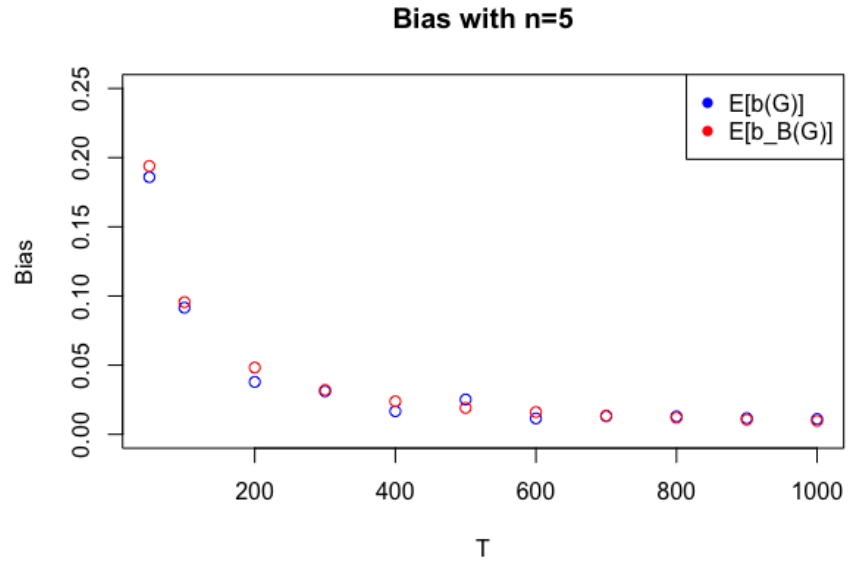
$$b_B(\hat{G}) = \frac{1}{B}\sum_{i=1}^{B} D^*(i) \quad , \quad D^*(i) = \frac{\hat{\mu}^*(i)^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}^*(i)\bar{\theta}^*(i)}} - \frac{\hat{\mu}^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}\bar{\theta}^*(i)}} \tag{21}$$

Since $\bar{\theta} = \hat{\Sigma}^{-1}\hat{\mu}$, it follows that $\bar{\theta}^*(i) = \hat{\Sigma}^{*-1}(i)\hat{\mu}^*(i)$ where $\hat{\mu}^*(i)$ is the vector of the mean returns of the $i-th$ bootstrap sample and $\hat{\Sigma}^{*-1}(i)$ is the inverse matrix of the estimated covariance matrix of the $i-th$ sample.

Let $n = 5$ and $\tau(\theta^*) = 0.639$ like in the previous section with μ and Σ given in Table 6.2 , we get the following results

T	50	100	500	1000
$E[b(G)]$	0.18592260	0.09155274	0.02520011	0.01108002
$E[b_B(\hat{G})]$	0.193940913	0.095573047	0.019108900	0.009741311
$E[b_B(\hat{G}) - b(G)]$	0.008018310	0.004020307	-0.006091210	-0.001338706
$Var(b(G))$	0.038383073	0.012137729	0.002016526	0.001193710
$Var(b_B(\hat{G}) - b(G))$	0.038899947	0.012641336	0.002029076	0.001203932

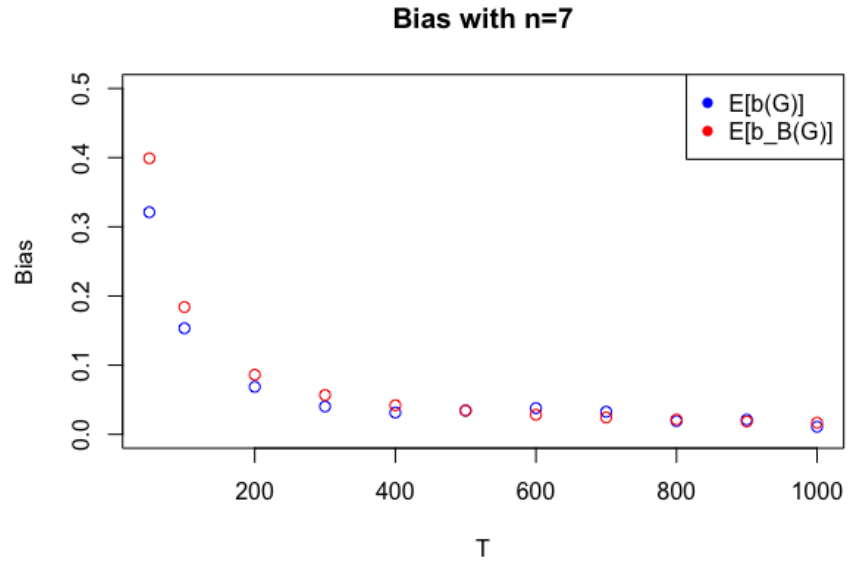
with the graph for different T.



Let $n = 7$ and $\tau(\theta^*) = 1.893$ and μ and Σ as Table 6.4 , we get

T	50	100	500	1000
$E[b(G)]$	0.32106749	0.15329509	0.03445561	0.01095473
$E[b_B(\hat{G})]$	0.39901946	0.18398893	0.03432741	0.01689747
$E[b_B(\hat{G}) - b(G)]$	0.077951965	0.030693841	-0.000128204	0.005942742
$Var(b(G))$	0.069198737	0.036262468	0.005706364	0.002904131
$Var(b_B(\hat{G}) - b(G))$	0.053253785	0.032811071	0.005600765	0.002835116

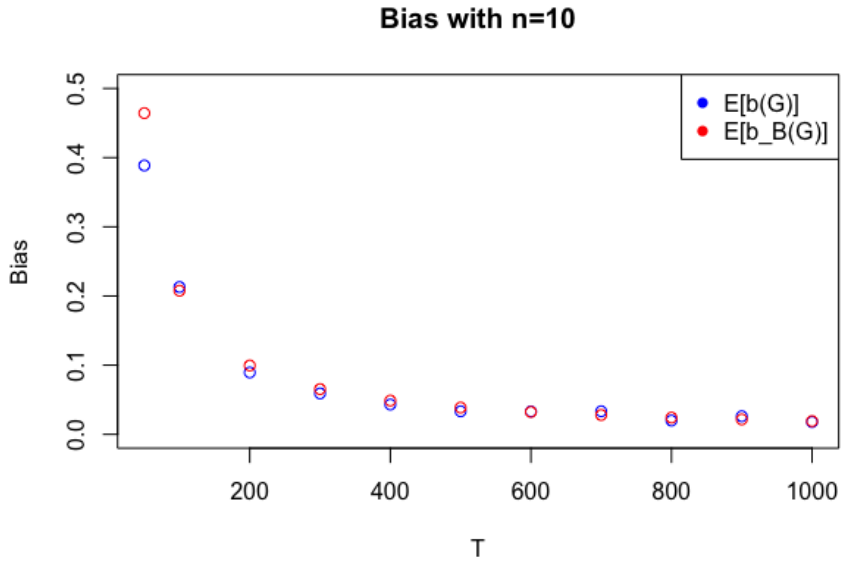
with the graph for different T.



And for $n = 10$ with $\tau(\theta^*) = 1.062$ and μ and Σ as Table 6.5

T	50	100	500	1000
$E[b(G)]$	0.38866244	0.21292585	0.03328613	0.01785017
$E[b_B(\hat{G})]$	0.46431969	0.20758395	0.03869519	0.01912915
$E[b_B(\hat{G}) - b(G)]$	0.075657252	-0.005341894	0.005409066	0.001278981
$Var(b(G))$	0.051864881	0.019598431	0.002975479	0.001563956
$Var(b_B(\hat{G}) - b(G))$	0.039046105	0.017959168	0.002978843	0.001552197

with the graph for different T



Compared to the tables in Section 3.1, $E[b_B(\hat{G}) - b(G)]$ is generally higher. This is because we are estimating two parameters $\hat{\mu}$ and $\hat{\Sigma}$ instead of just one like before where we estimated the mean returns $\hat{\mu}$, but Σ was known. However, the differences are still very small, although not as small as in Section 3.1.

3.3 Bias with real data

For so far we've looked at situations with simulated data generated by a multivariate normal distribution. We had a situation where we assumed that Σ was known and a situation where Σ was unknown, although we knew it to simulate the data. Because we knew it we were able to measure the accuracy of the bootstrap bias with an unknown covariance matrix, situations like in the stock market. We are able to estimate $\hat{\mu}$ and $\hat{\Sigma}$, but the true μ and Σ remains unknown. Therefore we are only able to estimate the bias through the bootstrap method. The replacements for the bootstrap method becomes

$$G(x) \to \hat{G}(x)$$

$$X_\alpha \sim G(x) \to X^*_\alpha \sim \hat{G}(x)$$

$$\mu \to \hat{\mu}$$

$$\hat{\mu} \to \hat{\mu}^*$$

$$\Sigma \to \hat{\Sigma}$$

$$\hat{\Sigma} \to \hat{\Sigma}^*$$

$$\bar{\theta} \to \bar{\theta}^*$$

The bootstrap estimate of the bias $b(G)$ is given by

$$b_B(\hat{G}) = \frac{1}{B}\sum_{i=1}^{B} D^*(i) \quad , \quad D^*(i) = \frac{\hat{\mu}^*(i)^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}^*(i)\bar{\theta}^*(i)}} - \frac{\hat{\mu}^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}\bar{\theta}^*(i)}} \tag{22}$$

Using the *quantmod* mode in R , we are able to download the stock returns from different stocks on Yahoo.com. We will be using the monthly stock returns for 10 companies in the $S\&P500$ over the last 15 years (T = 180) with the following ticker symbols :
Apple Inc. (AAPL), Amazon.com Inc. (AMZN), Alphabet Inc. (GOOGL), Microsoft Corporation (MSFT), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), Visa Inc. (V), Procter & Gamble Company (PG), NVIDIA Corporation (NVDA) and Coca-Cola Company (KO). Then the mean monthly returns $\hat{\mu}$ and the weights $\bar{\theta}$ of the stocks is given by

	Mean return	Weights
AAPL	0.022681496	0.15312343
AMZN	0.024058524	0.25226688
GOOGL	0.015162039	-0.15646389
MSFT	0.015965240	0.08915606
JNJ	0.006089021	-0.08510370
JPM	0.011240314	0.01287974
V	0.015362046	0.28126941
PG	0.005921317	0.35711495
NVDA	0.034145774	0.14641688
KO	0.005823066	-0.05065977

When the portfolio weights have negative values, it suggests that you should consider going short for those particular stocks. Going short involves borrowing shares of a stock, selling them at the current market price, and repurchase them back when the stock price has lowered. The difference in the price you sold it for and the price you bought it back again, is your profit per share. By selling high and buying back at a lower price, you could profit from a decline in the stock's price.

Set $B = 1000$, now we are able to calculate the in-sample Sharpe ratio with the estimated covariance matrix $\hat{\Sigma}$, the bootstrap bias and the estimated out-of-sample Sharpe ratio in terms of $\hat{\mu}$ and $\hat{\Sigma}$. We get the following results

$\phi(\theta)$	0.3512865
$b_B(\hat{G})$	0.1508145
$\bar{\tau}$	0.200472

So $\bar{\tau} \approx 0.200472$. This suggests that the investment may not be favorable since its Sharpe ratio is significantly lower than 1.

3.4 Parametric Bootstrapping

There are also other methods available for creating a bootstrap sample. In our previous approach to create a bootstrap sample, we randomly selected data points from the original data set with equal probability $\frac{1}{T}$ and repeated this T times. However, we can also create a parametric bootstrap sample. Specifically, we can do it in a similar way to how we simulated data using a multivariate normal distribution in Section 3.1. We will repeat this process again, but this time the bootstrap sample is generated using a multivariate normal distribution where the parameters are the estimated mean returns $\hat{\mu}$ and the estimated covariance matrix $\hat{\Sigma}$ from the original data sample. This assumes that the distribution of stock returns follows a normal distribution.

The parametric bootstrap method can be applied to both simulated data and real data. We'll start by looking at the simulated first.
Let $\boldsymbol{X} \sim N(\mu, \Sigma)$ be an n dimensional multivariate normal distribution where $\mu \in R^n$ and $\Sigma \in R^{n\times n}$ is a symmetric, positive definite $n \times n$ matrix. Now again we are able to simulate the returns of n stocks using the *mvrnorm* package in R with an given T. Assume Σ is known and μ is unknown so that we can compare the results with the tables from Section 3.1.
So for B bootstrap samples we have $X_1^*, X_2^*, ..., X_B^* \sim N(\hat{\mu}, \Sigma)$ where $\hat{\mu}$ is the estimated mean returns of the original data set that was multivariate normally distributed with mean μ and covariance matrix Σ. Denote $b_{B_P}(\hat{G})$ as the parametric bootstrap bias and $b_{B_{NP}}(\hat{G})$ as the bootstrap bias where the bootstrap sample is made by randomly selecting data points from the original data set as described earlier this chapter. Let

$\Delta_P = E[b_{B_P}(\hat{G}) - b(G)]$ and $\Delta_{NP} = E[b_{B_{NP}}(\hat{G}) - b(G)]$. Now we are able to compare both the expectations of the given μ and Σ given in the Appendix 6.

For the same vector μ and matrix Σ given in Table 6.1 and $n = 5$ with $\tau(\theta^*) = 0.373$, for different T we have [5]

T	50	100	500	1000
$E[b(G)]$	0.161438402	0.121482645	0.029211961	0.006953545
$E[b_{B_P}(\hat{G})]$	0.14537207	0.08738081	0.02020537	0.01066994
$E[b_{B_{NP}}(\hat{G})]$	0.14272527	0.08571980	0.02035783	0.01058429
$Var(b_{B_P}(\hat{G}) - b(G))$	0.019883738	0.013008789	0.002365005	0.001024831
$Var(b_{B_{NP}}(\hat{G}) - b(G))$	0.020048627	0.013053651	0.002388197	0.001020464
Δ_P	-0.016066328	-0.034101836	-0.009006590	0.003716392
Δ_{NP}	-0.018713130	-0.035762847	-0.008854128	0.003630747
$\lvert\Delta_P\rvert \leq \lvert\Delta_{NP}\rvert$?	yes	yes	no	no

Now with $n = 5$ and μ and Σ as in Table 6.2 with $\tau(\theta^*) = 0.639$ we have

T	50	100	500	1000
$E[b(G)]$	0.102555996	0.073888945	0.006435471	0.004927047
$E[b_{B_P}(\hat{G})]$	0.114963694	0.058065894	0.012431604	0.006427476
$E[b_{B_{NP}}(\hat{G})]$	0.112109330	0.057476420	0.012627387	0.006146473
$Var(b_{B_P}(\hat{G}) - b(G))$	0.0213450919	0.0111561158	0.0018134659	0.0009754792
$Var(b_{B_{NP}}(\hat{G}) - b(G))$	0.0213733193	0.0112661018	0.0018082516	0.0009977858
Δ_P	0.012407698	-0.015823051	0.005996133	0.001500430
Δ_{NP}	0.009553334	-0.016412525	0.006191916	0.001219427
$\lvert\Delta_P\rvert \leq \lvert\Delta_{NP}\rvert$?	no	yes	yes	no

For $n = 7$ and $\tau(\theta^*) = 0.692$ with μ and Σ as in Table 6.3 , we have

T	50	100	500	1000
$E[b(G)]$	0.15065272	0.10134822	0.01844502	0.01421800
$E[b_{B_P}(\hat{G})]$	0.148507588	0.078436755	0.016926924	0.008582008
$E[b_{B_{NP}}(\hat{G})]$	0.14395746	0.07737558	0.01702628	0.00850104
$Var(b_{B_P}(\hat{G}) - b(G))$	0.0239630822	0.0120946178	0.0020251103	0.0008141151
$Var(b_{B_{NP}}(\hat{G}) - b(G))$	0.0243591136	0.0121813489	0.0020471102	0.0008127827
Δ_P	-0.002145136	-0.022911462	-0.001518100	-0.005635997
Δ_{NP}	-0.006695264	-0.023972638	-0.001418743	-0.005716965
$\lvert\Delta_P\rvert \leq \lvert\Delta_{NP}\rvert$?	yes	yes	no	yes

Now for the same $n = 7$, but $\tau(\theta^*) = 1.893$ and μ and Σ as in Table 6.4 , we obtain

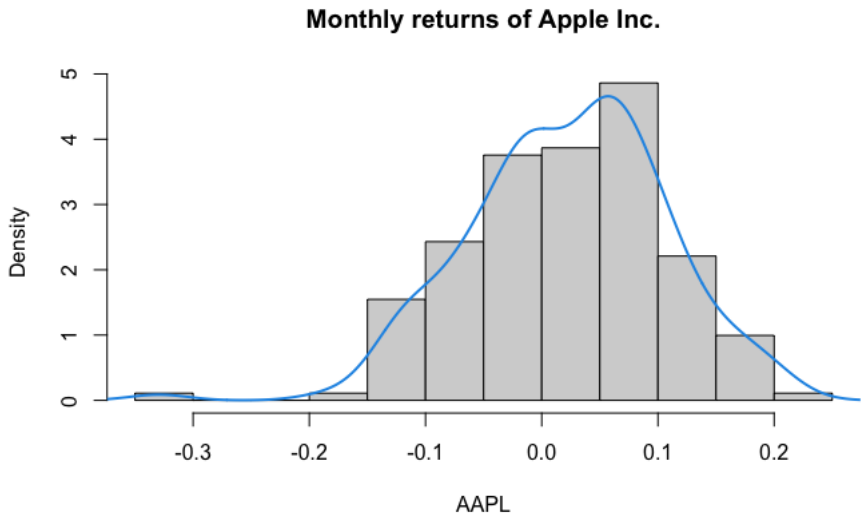
T	50	100	500	1000
$E[b(G)]$	0.082405655	0.023672404	-0.001584687	0.003369389
$E[b_{B_P}(\hat{G})]$	0.061744905	0.031690093	0.006163020	0.003117028
$E[b_{B_{NP}}(\hat{G})]$	0.060631251	0.031585741	0.006358751	0.002930593
$Var(b_{B_P}(\hat{G}) - b(G))$	0.0213035798	0.0127760426	0.0025203315	0.0009934036
$Var(b_{B_{NP}}(\hat{G}) - b(G))$	0.0212727407	0.0128651896	0.0025345261	0.0009794762
Δ_P	-0.020660750	0.008017689	0.007747707	-0.000252361
Δ_{NP}	-0.0217744043	0.0079133371	0.0079434376	-0.0004387955
$\lvert\Delta_P\rvert \leq \lvert\Delta_{NP}\rvert$?	yes	no	yes	yes

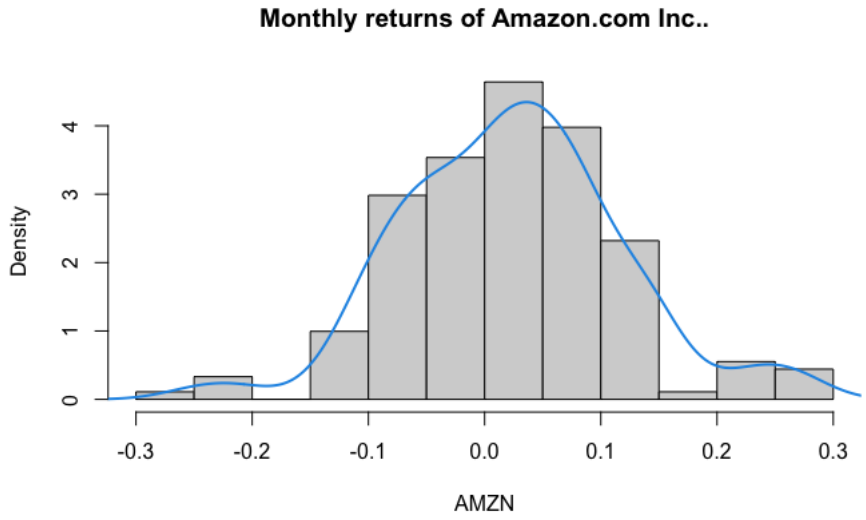
[5] Note that $E[b(G)]$ is different than in 3.1 , as we simulate the data A = 100 times again resulting in different $\hat{\mu}$.

Last, with $n = 10$ and $\tau(\theta^*) = 1.062$ and with μ and Σ as in Table 6.5

T	50	100	500	1000
$E[b(G)]$	0.15878676	0.07173546	0.01647524	0.00472808
$E[b_{B_P}(\hat{G})]$	0.150372015	0.080372687	0.016813668	0.008490658
$E[b_{B_{NP}}(\hat{G})]$	0.148216456	0.078741650	0.016753313	0.008454651
$Var(b_{B_P}(\hat{G}) - b(G))$	0.025029831	0.010218825	0.002248660	0.001095269
$Var(b_{B_{NP}}(\hat{G}) - b(G))$	0.024801466	0.010376836	0.002244633	0.001085448
Δ_P	-0.0084147426	0.0086372223	0.0003384249	0.0037625782
Δ_{NP}	-0.0105703011	0.0070061855	0.0002780703	0.0037265706
$\lvert\Delta_P\rvert \leq \lvert\Delta_{NP}\rvert$?	yes	no	no	no

The results indicate that the parametric bootstrap simulation is not always effective. Intuition would make you think it should have been much more effective because the simulated data itself is already normally distributed, but 11 out of the 20 results have been improved. This also shows that the non-parametric bootstrap method is also very effective to estimate a bias for the out-of-sample Sharpe Ratio with an unknown covariance matrix.
However, for real data we cannot assume that the distribution of the returns are normally distributed. Take a look at the distribution of the monthly returns of Apple Inc. (AAPL) and Amazon.com Inc. (AMZN) over the last 15 years.





Since using the bootstrap method is almost as effective as the parametric bootstrap, knowing that the distribution is normal, the parametric bootstrap seems not to be a good choice when working with real data. It seems that working with the non-parametric bootstrap method will result in a more accurate bias estimation for real data.

3.5 Efficient Bootstrap Simulation

We're going to look at an situation again where both μ and Σ are unknown and let $A = 100$ and $B = 1000$. Then the expectations of the true bias $B(G)$, the bootstrap bias $b_B(\hat{G})$ and the variance of the bootstrap bias of μ and Σ as in Table 6.2 for different T is

T	50	100	500	1000
$E[b(G)]$	0.182091381	0.080742047	0.019995929	0.005194821
$E[b_B(\hat{G})]$	0.193029532	0.096274737	0.019104932	0.009564266
$E[b_B(\hat{G}) - b(G)]$	0.0109381506	0.0155326894	-0.0008909971	0.0043694456
$Var(b_B(\hat{G}) - b(G))$	0.024647856	0.010543000	0.002199383	0.001563571
$Var(b_B(\hat{G}))$	0.0002554725	0.00005197233	0.000002852164	0.00000117283

We can reduce the variance of D^* significantly with an effective, yet extremely simply method called the efficient bootstrap simulation method.
Let $\boldsymbol{X}_T = (X_1, X_2, ..., X_T)$ be a random sample of size T of stock returns with unknown probability function $G(x)$. Let f be the Sharpe ratio function. Set the difference between the in-sample Sharpe ratio of $\bar{\theta}$ and T times the out-of-sample Sharpe ratio of every stock based on weight $\bar{\theta}$ as

$$D(\boldsymbol{X}_T; G) = f(\boldsymbol{X}_T|\bar{\boldsymbol{\theta}}) - T\int f(z|\bar{\boldsymbol{\theta}})dG(z), \tag{23}$$

where $f(\boldsymbol{X}_T|\bar{\theta}) = \sum_{\alpha=1}^{T} f(X_\alpha|\bar{\boldsymbol{\theta}})$. We can decompose $D(\boldsymbol{X}_T; G)$ into three terms:

$$D(\boldsymbol{X}_T; G) = D_1(\boldsymbol{X}_T; G) + D_2(\boldsymbol{X}_T; G) + D_3(\boldsymbol{X}_T; G), \tag{24}$$

where

$$\begin{aligned} D_1(\boldsymbol{X}_T; G) &= f(\boldsymbol{X}_T|\bar{\boldsymbol{\theta}}) - f(\boldsymbol{X}_T|\boldsymbol{\theta}) \\ D_2(\boldsymbol{X}_T; G) &= f(\boldsymbol{X}_T|\boldsymbol{\theta}) - T\int f(z|\boldsymbol{\theta})dG(z) \\ D_3(\boldsymbol{X}_T; G) &= T\int f(z|\boldsymbol{\theta})dG(z) - T\int f(z|\bar{\boldsymbol{\theta}})dG(z). \end{aligned} \tag{25}$$

Since the bias $b(G)$ is the expectation of $D(\boldsymbol{X}_T; G)$ with respect to $G(x)$, we can take the expectation of the right-hand side of equation 24. For the expectation of $D_2(\boldsymbol{X}_T; G)$ we have

$$\begin{aligned} E_G\Big[D_2(\boldsymbol{X}_T; G)\Big] &= E_G\Big[f(\boldsymbol{X}_T|\boldsymbol{\theta}) - T\int f(z|\boldsymbol{\theta})dG(z)\Big] \\ &= \sum_{\alpha=1}^{T} E_G\Big[f(X_\alpha|\boldsymbol{\theta})\Big] - TE_G\Big[f(Z|\theta)\Big] \\ &= 0. \end{aligned}$$

Therefore, the expectation in the second term can be removed from the bias for the Sharpe ratio. The expectation of $D(\boldsymbol{X}_T; G)$ is then given by

$$E_G\Big[D(\boldsymbol{X}_T; G)\Big] = E_G\Big[D_1(\boldsymbol{X}_T; G) + D_3(\boldsymbol{X}_T; G)\Big]. \tag{26}$$

Similarly, for the bootstrap estimate, we have

$$E_{\hat{G}}\Big[D(\boldsymbol{X}_T^*; \hat{G})\Big] = E_{\hat{G}}\Big[D_1(\boldsymbol{X}_T^*; \hat{G}) + D_3(\boldsymbol{X}_T^*; \hat{G})\Big]. \tag{27}$$

For each bootstrap sample we have

$$D_1(\boldsymbol{X}_T^*(i); \hat{G}) + D_3(\boldsymbol{X}_T^*(i); \hat{G}) = f(\boldsymbol{X}_T^*(i)|\bar{\boldsymbol{\theta}}^*(i)) - f(\boldsymbol{X}_T^*(i)|\bar{\boldsymbol{\theta}}) \tag{28}$$

$$+ f(\boldsymbol{X}_T|\bar{\boldsymbol{\theta}}) - f(\boldsymbol{X}_T|\bar{\boldsymbol{\theta}}^*(i)). \tag{29}$$

So for B bootstrap samples with replacement, we can use the following as the bootstrap estimate of the bias for the Sharpe ratio with unknown μ and Σ where the variance of $b_b(\hat{G})$ is reduced

$$b_B(\hat{G}) = \frac{1}{B}\sum_{i=1}^{B}\Big[D_1(\boldsymbol{X}_T^*(i); \hat{G}) + D_3(\boldsymbol{X}_T^*(i); \hat{G})\Big]. \tag{30}$$

Let's look back at the previous situation where μ and Σ are unknown, but we simulate data with μ and Σ. The replacements for the bootstrap method becomes

$$\begin{gathered} G(x) \to \hat{G}(x) \\ X_\alpha \sim G(x) \to X_\alpha^* \sim \hat{G}(x) \\ \mu \to \hat{\mu} \\ \hat{\mu} \to \hat{\mu}^* \\ \Sigma \to \hat{\Sigma} \\ \hat{\Sigma} \to \hat{\Sigma}^* \\ \bar{\theta} \to \bar{\theta}^* \end{gathered}$$

We can determine the true bias $b(G)$, the bootstrap bias $b_B(\hat{G})$ and the bootstrap bias based on the efficient bootstrap simulation $b_{B_{D_1+D_3}}(\hat{G})$ by

$$b(G) = \frac{\hat{\mu}^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}\bar{\theta}}} - \frac{\mu^T\bar{\theta}}{\sqrt{\bar{\theta}^T\Sigma\bar{\theta}}} \tag{31}$$

$$b_B(\hat{G}) = \frac{1}{B}\sum_{i=1}^{B} D^*(i) \quad , \quad D^*(i) = \frac{\hat{\mu}^*(i)^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}^*(i)\bar{\theta}^*(i)}} - \frac{\hat{\mu}^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}\bar{\theta}^*(i)}} \tag{32}$$

$$b_{B_{D_1+D_3}}(\hat{G}) = \frac{1}{B}\sum_{i=1}^{B} D_1^*(i) + D_3^*(i) \quad ,$$

$$D_1^*(i) + D_3^*(i) = \frac{\hat{\mu}^*(i)^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)^T\hat{\Sigma}^*(i)\bar{\theta}^*(i)}} - \frac{\hat{\mu}^*(i)^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}^*(i)\bar{\theta}}} + \frac{\hat{\mu}^T\bar{\theta}}{\sqrt{\bar{\theta}^T\hat{\Sigma}\bar{\theta}}} - \frac{\hat{\mu}^T\bar{\theta}^*(i)}{\sqrt{\bar{\theta}^*(i)\hat{\Sigma}\bar{\theta}^*(i)}} \tag{33}$$

Let $A = 1$ and $B = 1000$ and μ and Σ as in Table 6.2 with $n = 5$ and $\tau(\theta^*) = 0.639$. Then we have

T	50	100	500	1000
$E[b(G)]$	0.1830706	0.09991509	0.02801865	0.008603255
$E[b_B(\hat{G})]$	0.1977577	0.0860242	0.01889751	0.008836818
$E[b_{B_{D_1+D_3}}(\hat{G})]$	0.1766645	0.0747394	0.01859225	0.00887424
$Var(D^*)$	0.03624417	0.01249258	0.002556437	0.00137943
$Var(D_1^* + D_3^*)$	0.01514015	0.002970085	0.0001907571	3.808036e-05
$\frac{Var(D^*)}{Var(D_1^*+D_3^*)}$	2.39391	4.206136	13.40153	36.22418

So the variance clearly decreases for the efficient bootstrap method. However, if we use the same μ and Σ, but set $A = 100$ and we consider the variance of $b_B(\hat{G}) - b(G)$ and the variance of $b_{B_{D_1+D_3}}(\hat{G}) - b(G)$ instead of the variance of D^* and $D_1^* + D_3^*$. We have

T	50	100	500	1000
$E[b(G)]$	0.18461463	0.08232263	0.01534991	0.01364866
$E[b_B(\hat{G})]$	0.19397732	0.09604567	0.01927288	0.00953835
$E[b_{B_{D_1+D_3}}(\hat{G})]$	0.175067736	0.086928329	0.017577812	0.008803043
$Var(b_B(\hat{G}) - b(G))$	0.02070216	0.01382024	0.00254931	0.00150082
$Var(b_{B_{D_1+D_3}}(\hat{G}) - b(G))$	0.02161585	0.01406552	0.00254228	0.00149871

And for $n = 10$ with μ and Σ as in Table 6.5

T	50	100	500	1000
$E[b(G)]$	0.40940322	0.18569519	0.03595673	0.02203501
$E[b_B(\hat{G})]$	0.46608338	0.20434356	0.03874754	0.01944856
$E[b_{B_{D_1+D_3}}(\hat{G})]$	0.43279702	0.19075269	0.03615625	0.01808009
$Var(b_B(\hat{G}) - b(G))$	0.036378997	0.016582132	0.003614419	0.001564121
$Var(b_{B_{D_1+D_3}}(\hat{G}) - b(G))$	0.038593564	0.016825635	0.003615264	0.001566377

When we examine the efficient bootstrap simulation method on the real data used in 3.3 of the 10 stocks of the $S\&P500$, we get the following results.
For data from the last 15 years[6] (T=180), 10 years (T=120) and 5 years (T=60) we get

T	50	100	500
$\phi(\bar{\theta})$	0.4684776	0.4402991	0.3535254
$E[b_B(\hat{G})]$	0.2892111	0.1913182	0.1478384
$E[b_{B_{D_1+D_3}}(\hat{G})]$	0.270292	0.1846712	0.1451355
$Var(b_B(\hat{G}))$	0.08166466	0.01140712	0.006426418
$Var(b_{B_{D_1+D_3}}(\hat{G}))$	0.0745123	0.008628737	0.004570688

It is remarkable that although the variance of D^* significantly decreases when working with $D_1 + D_3$, it has little impact on the variance of the difference between the bootstrap bias and true bias when using $D_1 + D_3$ instead of $D_1 + D_2 + D_3$. This is partly due to the high values of A and B, which compensate for outliers. Since both are high, we observe that the expectation of the bootstrap bias with $D_1 + D_2 + D_3$ remains fairly stable, resulting in minimal change in the variance of the difference between bootstrap bias and true bias. Since the performance of the expectation of the bias does not deteriorate significantly, we will continue using the efficient bootstrap method going forward.

[6] Written in June 2023, the data is from July 2008 up until June 2023

4 Optimal out-of-sample performance on the stock market

In the previous chapter we've looked at the bootstrap method to provide a solution for our problem, to find the bias for the Sharpe ratio in terms of $\hat{\mu}$ and $\hat{\Sigma}$. Now we can apply this to find an out-of-sample Sharpe ratio as high as possible.

4.1 AEX index

Whereas we used some stocks of the $S\&P500$ to provide an illustration in Section 3.3, we will now evaluate the stocks included in the AEX index. The AEX index (Amsterdam Exchange index) is a stock market index of the largest and most traded Dutch companies on the stock market. The Amsterdam Exchange, formerly known as the Amsterdam Stock Exchange , consists out of 25 companies as of today. It consist multinationals like Shell (SHELL) , Heineken (HEIA) and Philips (PHIA) . Since it is an index, every company has its own weight, based on multiple factors like the number of shares or the stock price.

We will use 21 out of the 25 stocks on the AEX index, as we will exclude DSM FIRMENICH AG (DSM), EXOR NV (EXOR), Unilever PLC (UNA) and UNIVERSAL MUSIC GROUP NV (UMG) because of limited data.

The stocks of the AEX are divided into 4 groups and we will use the monthly returns from October 2019 until March 2023 ($T = 42$).

Company Name	**Ticker Symbol**
ABN AMRO Bank	ABN
Adyen	ADYEN
Aegon	AGN
Akzo Nobel	AKZA
ArcelorMittal	MT
ASM International	ASM
ASML Holding	ASML
ASR Nederland	ASRNL
BE Semiconductor Industries	BESI
Heineken	HEIA
IMCD Group	IMCD
ING Groep	INGA
Koninklijke Ahold Delhaize	AD
Koninklijke KPN	KPN
NN Group	NN
Koninklijke Philips	PHIA
Prosus	PRX
RANDSTAD NV	RAND
RELX NV	REN
Royal Dutch Shell	SHELL
Wolters Kluwer	WKL

For every k, we can determine the in-sample Sharpe ratio $\phi(\bar{\theta})$ and the estimated out-of-sample Sharpe ratio of $\bar{\theta}$, the $SRIC_c$. For Group 1 and $B = 1000$ we have

k =	0	1	2	3	4
ϕ	0.2374982	0.2908931	0.4109325	0.4223135	0.4369413
$SRIC_c$	0.2029535‘	0.1844124	**0.2758241**	0.2340135	0.1989806

With ADYEN, AKZA and MT as the stocks in the optimal portfolio of $k = 2$.
For Group 2 we have

k =	0	1	2	3	4
ϕ	0.3472794	0.3663406	0.3684859	0.3704620	0.3705338
$SRIC_c$	**0.3381414**	0.2848894	0.2274709	0.1751970	0.1138997

With ASM as the only stock in the optimal portfolio for Group 2 with $k = 0$.

For Group 3 we have

k =	0	1	2	3	4
ϕ	0.2561472	0.2775962	0.3020766	0.3108165	0.3127524
$SRIC_c$	**0.22601033**	0.16883125	0.12507073	0.07784122	0.03122697

With IMCD as the only stock in the optimal portfolio for Group 3 with $k = 0$ as well.

And for Group 4 we have

k =	0	1	2	3	4	5
ϕ	0.2507819	0.3271396	0.3972709	0.4365822	0.4442045	0.4538303
$SRIC_c$	0.2200048	0.2380637	**0.2705799**	0.2513734	0.2004801	0.1613454

With PHIA, PRX and WKL as the optimal portfolio for $k = 2$.
Combining these stocks gives us a set of 8 stocks. For every k we can determine the optimal portfolio for the in-sample Sharpe ratio and thus determine the estimated out-of-sample Sharpe ratio.

k =	0	1	2	3	4	5	6	7
ϕ	0.3472794	0.5854177	0.6048200	0.6332137	0.6628492	0.6926390	0.7025422	0.7026793
$SRIC_c$	0.3372524	**0.5405120**	0.5011670	0.4757376	0.4613320	0.4388097	0.3987629	0.3430310

The 2 stocks are ASM and PHIA, with weights

Stock	Weight
ASM	3.300712
PHIA	-4.300712

4.2 Dow Jones

We will now turn our attention to stocks in another index, the Dow Jones Index. It is an index comprising 30 of the most traded stocks on the American stock exchange. This index was created by the editors of the Wall Street Journal and Charles Dow, the founder of Dow Jones & Company. The list of the 30 stocks is divided into three groups.

Company Name	**Ticker Symbol**
American Express	AXP
Amgen	AMGN
Apple	AAPL
Boeing	BA
Caterpillar	CAT
Cisco Systems	CSCO
Chevron	CVX
Goldman Sachs	GS
Home Depot	HD
Honeywell International	HON
IBM	IBM
Intel	INTC
Johnson & Johnson	JNJ
Coca-Cola	KO
JPMorgan Chase	JPM
McDonald's	MCD
3M Company	MMM
Merck & Co.	MRK
Microsoft	MSFT
Nike	NKE
Procter & Gamble	PG
Travelers Companies	TRV
UnitedHealth Group	UNH
Salesforce.com	CRM
Verizon Communications	VZ
Visa	V
Walgreens Boots Alliance	WBA
Walmart	WMT
Disney	DIS
Dow Inc.	DOW

We will be using the monthly returns of the past 5 years (from July 2018-June 2023) with $T = 60$. For Group 1 we get

$\mathbf{k} =$	0	1	2	3	4
$\bar{\phi}$	0.2989935	0.3092388	0.3271078	0.3364129	0.3468378
$SRIC_c$	**0.29026984**	0.24882782	0.21719532	0.18010834	0.15167386

$\mathbf{k} =$	5	6	7	8	9
$\bar{\phi}$	0.3555418	0.3590934	0.3613501	0.3631258	0.3640199
$SRIC_c$	0.12052628	0.08855036	0.05069839	0.02848946	-0.01053263

With AAPL as the only stock.

For Group 2 we have

$\mathbf{k} =$	0	1	2	3	4
$\bar{\phi}$	0.3577082	0.4937761	0.5597171	0.5920486	0.6104849
$SRIC_c$	0.3569828	0.4509397	0.4803914	**0.4850859**	0.4656908

$\mathbf{k} =$	5	6	7	8	9
$\bar{\phi}$	0.6251110	0.6364925	0.6397944	0.6427450	0.6453592
$SRIC_c$	0.4408035	0.4224554	0.3952310	0.3514836	0.3157805

The optimal group for $k = 1$ is MCD, MMM, MRKK and MSFT.

For Group 3 we get

$\mathbf{k} =$	0	1	2	3	4
ϕ	0.2388362	0.3439759	0.3885305	0.4103151	0.4357754
$SRIC_c$	0.22044014	0.28434715	**0.28520458**	0.25687398	0.21891846

$\mathbf{k} =$	5	6	7	8	9
ϕ	0.4569858	0.4662683	0.4765018	0.4881302	0.4910108
$SRIC_c$	0.20844617	0.17762202	0.16560048	0.13485215	0.09807598

where the 3 stocks are PG, VZ and V.

We will combine the 8 stocks together. We get

$\mathbf{k} =$	0	1	2	3
ϕ	0.3577082	0.4937761	0.5619183	0.5964770
$SRIC_c$	0.3514779	0.4470615	0.4826738	0.4695378

$\mathbf{k} =$	4	5	6	7
ϕ	0.6445280	0.6696924	0.6761052	0.6790434
$SRIC_c$	**0.4896885**	0.4761158	0.4557754	0.4217120

where MMM, MRK, MSFT, PG and VZ are the stocks for $k = 4$ with weights

Stock	Weight
MMM	-1.4767443
MRK	0.9247129
MSFT	1.5172696
PG	1.2050789
VZ	-1.1703171

The estimated out-of-sample Sharpe ratio's from both cases are not very good and it is nowhere near an estimated out-of-sample Sharpe ratio of 1. The $SRIC_c$ heavily penalizes portfolios with a large number of parameters, which, when combined with the high level of volatility observed in the stock market in recent years, makes it challenging to find an optimal portfolio with an expected out-of-sample Sharpe ratio above 1 as of right now.

5 Conclusion

We first examined the concept of the Sharpe ratio and its mathematical formulation. In doing so, we also explored the Sharpe Ratio Information Criterion, which led us to our research problem. To find a bias correction for the Sharpe ratio when the covariance matrix Σ is unknown. For this method, we looked into the Akaike Information Criterion, which selects the model with the best log-likelihood in a similar manner.

In Chapter 3, we utilized the bootstrap method to estimate a bias correction for the Sharpe ratio, both with and without assuming known Σ. Since we were working with simulated data, we knew the true values of μ and Σ. Therefore, we could assess the accuracy of the expected bootstrap bias and analytical bias compared to the expectation of the true bias. The analytical and bootstrap biases closely aligned and consistently remained close to each other. For the simulated data where Σ is assumed unknown, we could only compare the expectation of the true bias and the bootstrap bias for different values of T. The bootstrap method proved to be highly effective in estimating a bias correction for the out-of-sample Sharpe ratio using only the estimated covariance matrix $\hat{\Sigma}$. We also explored various extensions of the bootstrap method, such as the parametric bootstrap and the efficient bootstrap simulation method, but found limited improvement. Finally, we applied these techniques to stocks in the Dutch and American markets and found that achieving an optimal out-of-sample Sharpe ratio above 1 is quite challenging. In-sample Sharpe Ratios with high k values are heavily penalized.

Overall, we succeeded in making the Sharpe ratio more applicable by removing the assumption of known Sigma, which also makes it more difficult to find an out-of-sample Sharpe ratio above 1.

6 Appendix

.1 Table 1

$$\text{mean\_returns} = \begin{bmatrix} 0.05 \\ 0.07 \\ 0.09 \\ 0.08 \\ 0.06 \end{bmatrix}$$

$$\text{cov\_mat} = \begin{bmatrix} 0.04 & 0.02 & 0.01 & 0.03 & 0.02 \\ 0.02 & 0.09 & 0.03 & 0.01 & 0.04 \\ 0.01 & 0.03 & 0.16 & 0.02 & 0.01 \\ 0.03 & 0.01 & 0.02 & 0.09 & 0.03 \\ 0.02 & 0.04 & 0.01 & 0.03 & 0.12 \end{bmatrix}$$

.2 Table 2

$$\text{mean\_returns} = \begin{bmatrix} 0.11 \\ 0.09 \\ 0.10 \\ 0.11 \\ 0.13 \end{bmatrix}$$

$$\text{cov\_mat} = \begin{bmatrix} 0.04 & 0.02 & 0.01 & 0.03 & 0.02 \\ 0.02 & 0.09 & 0.03 & 0.01 & 0.04 \\ 0.01 & 0.03 & 0.10 & 0.02 & 0.01 \\ 0.03 & 0.01 & 0.02 & 0.04 & 0.03 \\ 0.02 & 0.04 & 0.01 & 0.03 & 0.12 \end{bmatrix}$$

.3 Table 3

$$\text{mean\_returns} = \begin{bmatrix} 0.12 \\ 0.10 \\ 0.08 \\ 0.15 \\ 0.11 \\ 0.13 \\ 0.09 \end{bmatrix}$$

$$\text{cov\_mat} = \begin{bmatrix} 0.06 & 0.03 & 0.02 & 0.01 & 0.02 & 0.03 & 0.02 \\ 0.03 & 0.08 & 0.04 & 0.02 & 0.03 & 0.04 & 0.02 \\ 0.02 & 0.04 & 0.07 & 0.01 & 0.02 & 0.03 & 0.01 \\ 0.01 & 0.02 & 0.01 & 0.09 & 0.02 & 0.03 & 0.02 \\ 0.02 & 0.03 & 0.02 & 0.02 & 0.08 & 0.03 & 0.02 \\ 0.03 & 0.04 & 0.03 & 0.03 & 0.03 & 0.10 & 0.04 \\ 0.02 & 0.02 & 0.01 & 0.02 & 0.02 & 0.04 & 0.06 \end{bmatrix}$$

.4 Table 4

$$\text{mean\_returns} = \begin{bmatrix} 0.07 \\ 0.08 \\ 0.09 \\ 0.10 \\ 0.11 \\ 0.11 \\ 0.05 \end{bmatrix}$$

$$\text{cov\_mat} = \begin{bmatrix} 0.023 & 0.013 & 0.009 & 0.008 & 0.003 & 0.005 & 0.007 \\ 0.013 & 0.026 & 0.011 & 0.009 & 0.001 & 0.001 & 0.007 \\ 0.009 & 0.011 & 0.013 & 0.008 & 0.002 & 0.006 & 0.005 \\ 0.008 & 0.009 & 0.008 & 0.010 & 0.002 & 0.005 & 0.005 \\ 0.003 & 0.001 & 0.002 & 0.002 & 0.004 & 0.003 & 0.002 \\ 0.005 & 0.001 & 0.006 & 0.005 & 0.003 & 0.016 & 0.005 \\ 0.007 & 0.007 & 0.005 & 0.005 & 0.002 & 0.005 & 0.008 \end{bmatrix}$$

.5 Table 5

$$\text{mean\_returns} = \begin{bmatrix} 0.15 \\ 0.08 \\ 0.06 \\ 0.13 \\ 0.06 \\ 0.14 \\ 0.07 \\ 0.03 \\ 0.12 \\ 0.05 \end{bmatrix}$$

$$\text{cov\_mat} = \begin{bmatrix} 0.05 & 0.02 & 0.01 & 0.00 & 0.01 & 0.02 & 0.01 & 0.00 & 0.01 & 0.01 \\ 0.02 & 0.08 & 0.03 & 0.01 & 0.02 & 0.01 & 0.02 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.03 & 0.06 & 0.02 & 0.01 & 0.00 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.00 & 0.01 & 0.02 & 0.04 & 0.01 & 0.01 & 0.01 & 0.00 & 0.00 & 0.00 \\ 0.01 & 0.02 & 0.01 & 0.01 & 0.07 & 0.01 & 0.02 & 0.01 & 0.01 & 0.01 \\ 0.02 & 0.01 & 0.00 & 0.01 & 0.01 & 0.05 & 0.03 & 0.02 & 0.01 & 0.01 \\ 0.01 & 0.02 & 0.01 & 0.01 & 0.02 & 0.03 & 0.09 & 0.02 & 0.01 & 0.01 \\ 0.00 & 0.01 & 0.01 & 0.00 & 0.01 & 0.02 & 0.02 & 0.06 & 0.02 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.00 & 0.01 & 0.01 & 0.01 & 0.02 & 0.05 & 0.02 \\ 0.01 & 0.01 & 0.01 & 0.00 & 0.01 & 0.01 & 0.01 & 0.01 & 0.02 & 0.04 \end{bmatrix}$$

.6 R Code for Chapter 3

```
library(MASS)
tijden <- c(50,100,500,1000)
tijdintervals <- length(tijden)

truebiasbvec <- rep(NA,tijdintervals)
bootsrapbiasvec <- rep(NA,tijdintervals)
efficientbootstrapvec <- rep(NA,tijdintervals)
verschivariance <- rep(NA,tijdintervals)
verschilvariance_efficient <- rep(NA,tijdintervals)

for (p in 1:tijdintervals){
  A <- 100
  rb <- rep(NA,A)
  bb <- rep(NA,A)
  bd1d3 <- rep(NA,A)
  for (j in 1:A) {
    mean_returns <- c(0.15, 0.08, 0.06, 0.13, 0.06, 0.14, 0.07, 0.03, 0.12, 0.05)

    # Covariance matrix
    cov_mat <- matrix(c(0.05, 0.02, 0.01, 0.00, 0.01, 0.02, 0.01, 0.00, 0.01, 0.01,
                        0.02, 0.08, 0.03, 0.01, 0.02, 0.01, 0.02, 0.01, 0.01, 0.01,
                        0.01, 0.03, 0.06, 0.02, 0.01, 0.00, 0.01, 0.01, 0.01, 0.01,
                        0.00, 0.01, 0.02, 0.04, 0.01, 0.01, 0.01, 0.00, 0.00, 0.00,
                        0.01, 0.02, 0.01, 0.01, 0.07, 0.01, 0.02, 0.01, 0.01, 0.01,
                        0.02, 0.01, 0.00, 0.01, 0.01, 0.05, 0.03, 0.02, 0.01, 0.01,
```

```
                    0.01, 0.02, 0.01, 0.01, 0.02, 0.03, 0.09, 0.02, 0.01, 0.01,
                    0.00, 0.01, 0.01, 0.00, 0.01, 0.02, 0.02, 0.06, 0.02, 0.01,
                    0.01, 0.01, 0.01, 0.00, 0.01, 0.01, 0.01, 0.02, 0.05, 0.02,
                    0.01, 0.01, 0.01, 0.00, 0.01 , 0.01 , 0.01, 0.01, 0.02, 0.04),
                  nrow = 10, ncol = 10, byrow = TRUE)



t <- tijden[p]
stocks <- length(mean_returns)
returns <- mvrnorm(t, mean_returns, cov_mat)

rf <- 0
means_sample <- colMeans(returns)
covariance_matrix <- cov(returns)
theta_sample <- solve(covariance_matrix) %*% means_sample
optimal_theta_sample <- theta_sample/sum(theta_sample)

#  The number of bootstrap replicates
B <- 1000

D <- rep(NA, B)
D1D3<- rep(NA,B)
for (i in 1:B) {
  # Generate a bootstrap sample
  boot_sample <- returns[sample(nrow(returns), replace = TRUE), ]
  meansboot_sample <- colMeans(boot_sample)
  covboot_sample <- cov(boot_sample)
  # Determine optimal Theta
  theta <- solve(covboot_sample) %*% meansboot_sample
  optimal_theta <- theta/sum(theta)

  sharpe_bootsample_optimaltheta <- (meansboot_sample %*% optimal_theta) /
  (sqrt(t(optimal_theta) %*%
  covboot_sample %*% optimal_theta))
  sharpe_sample_optimaltheta <- (means_sample %*% optimal_theta) /
  (sqrt(t(optimal_theta) %*% covariance_matrix %*% optimal_theta))

  d2 <- (meansboot_sample %*% optimal_theta_sample) /
  (sqrt(t(optimal_theta_sample) %*% covboot_sample %*% optimal_theta_sample))
  d3 <- (means_sample %*% optimal_theta_sample) /
  (sqrt(t(optimal_theta_sample) %*% covariance_matrix %*% optimal_theta_sample))
  D_i <- sharpe_bootsample_optimaltheta - sharpe_sample_optimaltheta
  D[i] <- D_i
  D_1_3 <- sharpe_bootsample_optimaltheta - d2 + d3 - sharpe_sample_optimaltheta
  D1D3[i] <- D_1_3
}



bias_bootstrap <- mean(D)
bias_efficientbootstrap <- mean(D1D3)
os  <- (mean_returns %*% optimal_theta_sample) /
(sqrt(t(optimal_theta_sample) %*% cov_mat %*% optimal_theta_sample))
ss <- (means_sample %*% optimal_theta_sample) /
(sqrt(t(optimal_theta_sample) %*% covariance_matrix %*% optimal_theta_sample))
realbias <- ss-os
```

```
    rb[j] <- realbias
    bb[j] <- bias_bootstrap
    bd1d3[j] <- bias_efficientbootstrap


}
  truebiasbvec[p] <- mean(rb)
  bootsrapbiasvec[p] <- mean(bb)
  efficientbootstrapvec[p] <- mean(bd1d3)
  verschivariance[p] <- var(bb-rb)
  verschilvariance_efficient[p] <- var(bd1d3-rb)

}

truebiasbvec
bootsrapbiasvec
efficientbootstrapvec
verschivariance
verschilvariance_efficient
```

.7 R Code for Chapter 4

```
    library(MASS)
library(quantmod)
symbols <- c("AXP", "AMGN", "AAPL", "BA", "CAT", "CSCO", "CVX", "GS",
"HD", "HON", "IBM", "INTC", "JNJ", "KO", "JPM", "MCD", "MMM", "MRK",
"MSFT", "NKE", "PG","TRV", "UNH", "CRM", "VZ", "V", "WBA", "WMT",
"DIS","DOW")
getSymbols(symbols, src = 'yahoo')



AXP <- periodReturn(AXP, period = 'monthly', subset = '2018-07::')
AMGN <- periodReturn(AMGN, period = 'monthly', subset = '2018-07::')
AAPL <- periodReturn(AAPL, period = 'monthly', subset = '2018-07::')
BA <- periodReturn(BA, period = 'monthly', subset = '2018-07::')
CAT <- periodReturn(CAT, period = 'monthly', subset = '2018-07::')
CSCO <- periodReturn(CSCO, period = 'monthly', subset = '2018-07::')
CVX <- periodReturn(CVX, period = 'monthly', subset = '2018-07::')
GS <- periodReturn(GS, period = 'monthly', subset = '2018-07::')
HD <- periodReturn(HD, period = 'monthly', subset = '2018-07::')
HON <- periodReturn(HON, period = 'monthly', subset = '2018-07::')
IBM <- periodReturn(IBM, period = 'monthly', subset = '2018-07::')
INTC <- periodReturn(INTC, period = 'monthly', subset = '2018-07::')
JNJ <- periodReturn(JNJ, period = 'monthly', subset = '2018-07::')
KO <- periodReturn(KO, period = 'monthly', subset = '2018-07::')
JPM <- periodReturn(JPM, period = 'monthly', subset = '2018-07::')
MCD <- periodReturn(MCD, period = 'monthly', subset = '2018-07::')
MMM <- periodReturn(MMM, period = 'monthly', subset = '2018-07::')
MRK <- periodReturn(MRK, period = 'monthly', subset = '2018-07::')
MSFT <- periodReturn(MSFT, period = 'monthly', subset = '2018-07::')
NKE <- periodReturn(NKE, period = 'monthly', subset = '2018-07::')
PG <- periodReturn(PG, period = 'monthly', subset = '2018-07::')
TRV <- periodReturn(TRV, period = 'monthly', subset = '2018-07::')
UNH <- periodReturn(UNH, period = 'monthly', subset = '2018-07::')
CRM <- periodReturn(CRM, period = 'monthly', subset = '2018-07::')
```

```
VZ <- periodReturn(VZ, period = 'monthly', subset = '2018-07::')
V <- periodReturn(V, period = 'monthly', subset = '2018-07::')
WBA <- periodReturn(WBA, period = 'monthly', subset = '2018-07::')
WMT <- periodReturn(WMT, period = 'monthly', subset = '2018-07::')
DIS <- periodReturn(DIS, period = 'monthly', subset = '2018-07::')
DOW <- periodReturn(GS, period = 'monthly', subset = '2018-07::')


returns <- cbind(AXP,AMGN,AAPL,BA,CAT,CSCO,CVX,GS,HD,HON,
                 IBM,INTC,JNJ,KO,JPM,MCD,MMM,MRK,MSFT,NKE,
                 PG,TRV,UNH,CRM,VZ,V,WBA,WMT,DIS,DOW)
colnames(returns) <- symbols

rf <-  0


sharpe_ratio_calculator <- function(stockreturns) {

  means <- colMeans(stockreturns)
  cov_matrix <- cov(stockreturns)
  stocks <- ncol(stockreturns)
  t <- nrow(stockreturns)
  theta <- solve(cov_matrix) %*% means
  weights <- theta / abs(sum(theta))
  total <- means %*% weights
  portfolio_sd <- sqrt(t(weights) %*% cov_matrix %*% weights)
  sharpe_ratio <- (total - mean(rf)) / portfolio_sd

  result <- list(sharpe_ratio = sharpe_ratio , weights = weights)
  return(result)
}

optimaln <- function(n,stockreturns){
  num_stocks <- ncol(stockreturns)
  num_combinations <- choose(num_stocks, n)
  return_matrices <- rep(NA,num_combinations)
  combinations <- combn(num_stocks, n)
  for (i in 1:num_combinations) {

    subset_returns <- stockreturns[, combinations[, i]]
    return_matrices[[i]] <- sharpe_ratio_calculator(subset_returns)$sharpe_ratio
  }

  max_position <- which.max(return_matrices)
  hoogste <- max(return_matrices)
  combi <- combinations[,max_position]
  result <- list(hoogste = hoogste , combi = combi)
  return(result)
}

highestweight <- function(stockreturns){
  bestsharpe <- rep(NA,ncol(stockreturns))
  bestsharpecombi <- list()
  for (p in 1:ncol(stockreturns)){
    bestsharpe[p] <- optimaln(p,stockreturns)$hoogste
    bestsharpecombi[[p]] <- optimaln(p,stockreturns)$combi
  }
  result <- list(bestsharpe = bestsharpe , bestsharpecombi = bestsharpecombi)
```

```
  return(result)
}

outofsample_sharpe_ratio <- function(returns){
  means_sample <- colMeans(returns)
  covariance_matrix <- cov(returns)
  theta_sample <- solve(covariance_matrix) %*% means_sample
  optimal_theta_sample <- theta_sample/abs(sum(theta_sample))
  rfmean <- mean(rf)
  insample_sharpe_ratio <- ((means_sample %*% optimal_theta_sample)-rfmean) /
  (sqrt(t(optimal_theta_sample) %*% covariance_matrix %*% optimal_theta_sample))
  # Set the number of bootstrap replicates
  B <- 10000

  D <- rep(NA, B)
  for (i in 1:B) {
    # Generate a bootstrap sample
    boot_sample <- returns[sample(nrow(returns), replace = TRUE), ]
    meansboot_sample <- colMeans(boot_sample)
    covarianceboot_sample <- cov(boot_sample)
    # Determine optimal Theta
    theta <- solve(covarianceboot_sample) %*% meansboot_sample
    optimal_theta <- theta/abs(sum(theta))

    D1 <- ((meansboot_sample %*% optimal_theta)-rfmean) /
    (sqrt(t(optimal_theta) %*% covarianceboot_sample %*% optimal_theta))
    D4 <- ((means_sample %*% optimal_theta)-rfmean) /
    (sqrt(t(optimal_theta) %*% covariance_matrix %*% optimal_theta))

    D2 <- ((meansboot_sample %*% optimal_theta_sample)-rfmean) /
    (sqrt(t(optimal_theta_sample) %*% covarianceboot_sample %*% optimal_theta_sample))
    D3 <- ((means_sample %*% optimal_theta_sample)-rfmean) /
    (sqrt(t(optimal_theta_sample) %*% covariance_matrix %*% optimal_theta_sample))
    D_i <- D1 - D2 + D3 - D4

    D[i] <- D_i

  }
  return(insample_sharpe_ratio- mean(D))
}

best_out_of_sample_sharpe_ratio<-function(stockreturns){
  os_sharpe_values <- rep(NA,ncol(stockreturns))
  for (i in 1:ncol(stockreturns)){
    stockmatrix <- c()
    stockmatrix <- cbind(stockreturns[,highestweight(stockreturns)$bestsharpecombi[[i]]])
    os_sharpe_values[i] <- outofsample_sharpe_ratio(stockmatrix)
  }

  return(os_sharpe_values)
}

returns1 <- returns[,1:10]
returns2<- returns[,11:20]
returns3<- returns[,21:30]

highestweight(returns1)$bestsharpecombi
highestweight(returns1)$bestsharpe
```

```
best_out_of_sample_sharpe_ratio(returns1)

highestweight(returns2)$bestsharpecombi
highestweight(returns2)$bestsharpe
best_out_of_sample_sharpe_ratio(returns2)

highestweight(returns3)$bestsharpecombi
highestweight(returns3)$bestsharpe
best_out_of_sample_sharpe_ratio(returns3)
```

References

[1] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference*. Colorado State University: Springer, 2002.

[2] Kenneth P. Burnham and David R. Anderson. "Multimodel inference: understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2 (2004), pp. 261–304.

[3] Sadanori Konishi and Genshiro Kitagawa. *Information Criteria and Statistical Modeling*. New York, NY: Springer, 2008.

[4] Dirk Paulse and Jakob Söhl. "Noise fit, estimation error and a Sharpe information criterion". In: *Quantitative Finance* 20.6 (2020), pp. 1027–1043. DOI: 10.1080/14697688.2020.1718746.