# Mandatory Assignment 1

FMW786 & GHC278

2024-03-08

## 1 Efficient portfolios and estimation uncertainty:

We retrieve the daily adjusted prices of every stock in the Dow Jones 30 index spanning from January 1st, 2000 to December 31st, 2023. Using the `TidyQuant` functions: `tq_index` to get the 30 tickers of the Dow Jones Index and `tq_get` to get their adjusted return history.

The number of unique symbols in the returns data frame is 27, and we have thus removed the three tickers that did not have a continuous trading history for the entire sample period.

For the 27 stocks, we compute the covariance matrix, $\Sigma$ and mean return, $\mu$ for each asset. With the risk-free rate of zero we find the sharpe ratio for each asset by normalizing with the asset's volatility. As we have monthly data the moments are annualized.

From the sharpe ratios, we find that the stock with the highest Sharpe-ratio is: UNH. The performance of the stock from

Table 1: Highest performing Sharpe-ratios and Return, annualized.

| Ticker | Sharpe | Return |
|---|---|---|
| UNH | 0.929 | 0.223 |

We define a function `compute_efficient_frontier` which take two inputs: a (N ×N) variance-covariance matrix `Sigma_est` and a vector `mu_est`.

To begin, the function calculates the weights for the minimum variance portfolio (MVP), which represents the portfolio with the lowest possible risk given the selected assets.

Next, the function proceeds to compute the weights for an efficient portfolio that delivers twice the expected return of the MVP. This step involves uses the covariance matrix and expected returns vector.

Utilizing the principles of the two-mutual fund theorem, the function characterizes a range of portfolio weights representing combinations of the MVP weights and the efficient portfolio weights. These weights are systematically computed for various values of `c`, where `c` ranges from -0.1 to 1.2. The resulting data structure, `portfolio_matrix`, encapsulates the weights for each portfolio.

Depending on the specified `return_type`, the function offers flexibility in the output format. It can return a tibble containing the weights for each portfolio (`"weights"`), a tibble with the expected return and standard deviation of each portfolio (`"efficient_frontier"`), as well as the weights of the tangency portfolio, using (`"tangency_weights"`).

Using the `ggplot2` package, the efficient frontier is plotted, with the x-axis representing the annualized standard deviation and the y-axis representing the annualized expected return. Each point on the plot corresponds to a specific portfolio configuration, with the MVP-portfolio and the efficient portfolio that delivers twice the return of the MVP highlighted, corresponding to extreme values of `c` (0 and 1), and smaller points denoting individual assets within the portfolio.

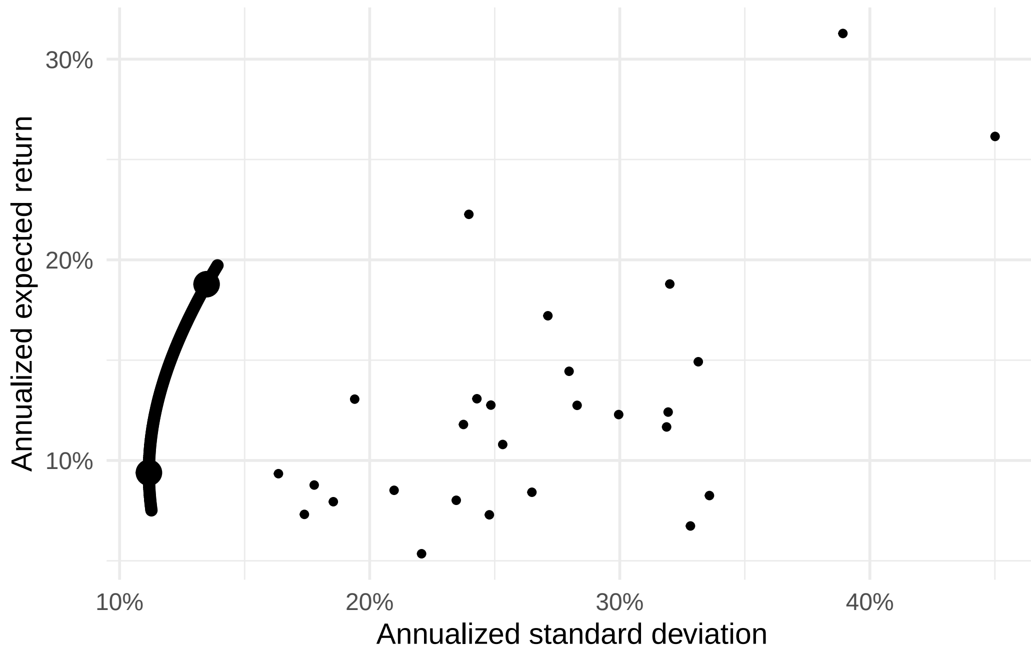


Figure 1: Efficient frontier for DOW index constituents

Now we want to find the efficient tangency portfolio weights $\omega_{tgc}$, with a risk-free rate of zero, based on our parameters, $\mu$ and $\Sigma$.
We have implemented the solution for the weights to the sharpe ratio maximizing problem, $\omega_{tgc} = \frac{\Sigma^{-1} \cdot (\mu - r_f \cdot 1)}{1' \Sigma^{-1} \cdot (\mu - r_f \cdot 1)}$ in the function `compute_efficient_frontier`, and thus compute the tangency portfolio weights. These are given by:

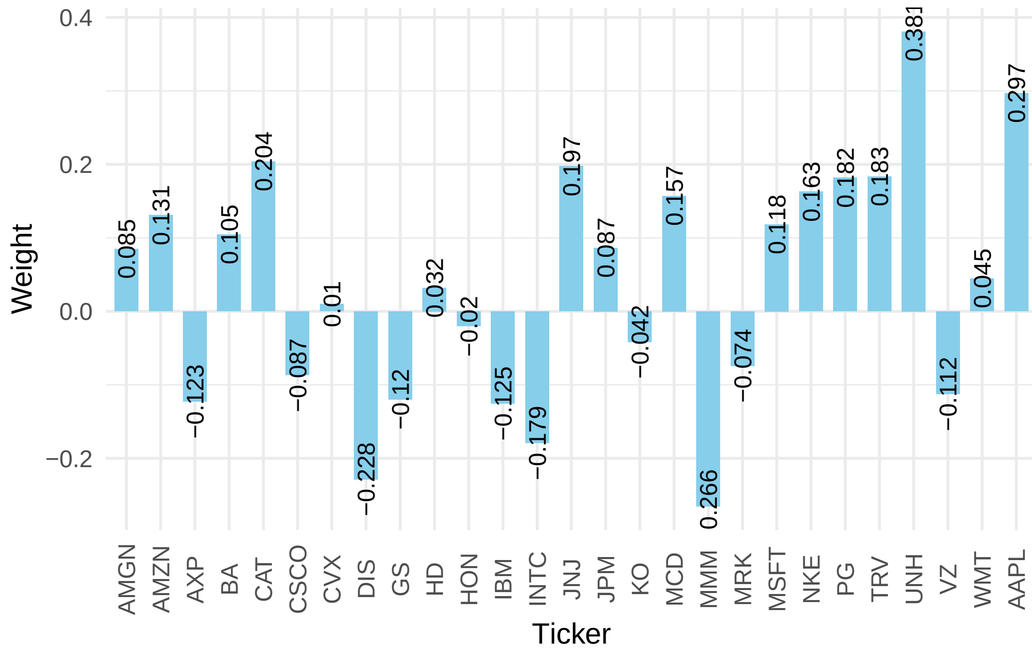


Figure 2: The weights of the Efficient Tangency Portfolio

Furthermore, the annualized mean return, standard deviation, and sharpe ratio of the tangency portfolio is given by:

Table 2: Tangency Portfolio Statistics, annualized

| Measure | Value |
|---|---|
| Mean Return | 0.300 |
| Standard Deviation | 0.200 |
| Sharpe Ratio | 1.503 |

Thus we can see, that the sharpe ratio of the tangency portfolio, which is 1.503, is significantly higher than that of the highest performing individual stock, which is UNH . This is to be expected as the tangency portfolio is constructed by optimizing the trade-off between risk and return across multiple assets. As UNH is included in this portfolio, we can always get at least a sharpe ratio of 0.929, if there were no value in diversifying by only holding UNH.

We can observe that the tangency portfolio weights allocate significant proportions to certain stocks, indicating a potential lack of diversification. Additionally, it's important to recognize that the mean returns and variance-covariance matrix used for optimization are subject to estimation uncertainty, which can introduce additional risk and potential deviation from expected outcomes. Thus, the implementation could be very sensitive to estimation uncertainty.

## 1.1 The efficient frontier under estimation uncertainty

Next we wish to simulate efficient frontiers under estimation uncertainty. We assume that the returns are identically and independently multivariate normal distributed with a vector of expected returns, $\mu$ and the variance-covariance matrix, $\Sigma$. We simulate hypothetical return samples and compute the sample moments $\hat{\mu}$ and $\hat{\Sigma}$. From the estimated efficient frontiers we can analyze how much they deviate from the true efficient frontier.

First, we define a function `simulate_returns` which generates a matrix of simulated returns for a specified number of periods and distribution. The generated hypothetical returns are i.i.d. $N(\mu, \Sigma)$ and the sample moments are thus the expected returns, $\hat{\mu}$ and variance-covariance matrix, $\hat{\Sigma}$. Each column corresponds to one of the assets and each row the returns of a single period. Our simulated returns consider 200 periods.

We can now plot the simulated efficient frontier using data from `simulate_returns`.

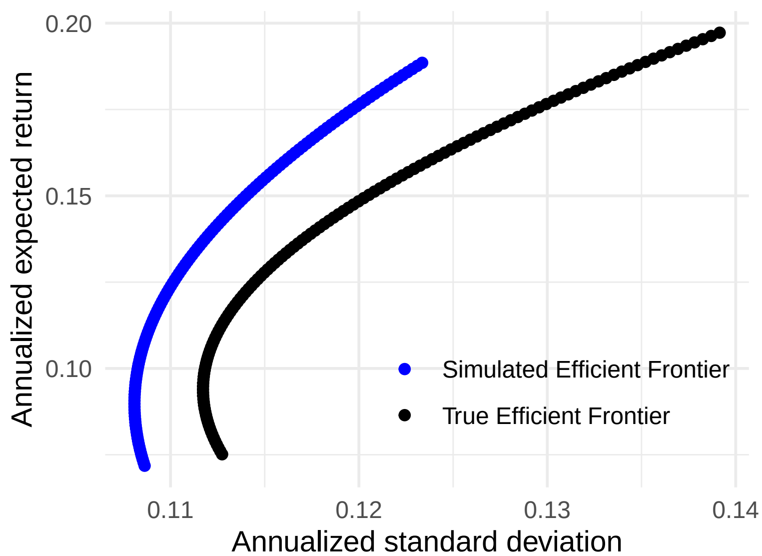


Figure 3: The Efficient Frontier under Estimation Uncertainty

Upon examining the above figure, we observe that the simulated efficient frontier diverges from the 'true' efficient frontier. Notably, the simulated frontier seems to dominate the 'true' frontier, suggesting that the simulation may overestimate the returns for a given level of risk. This discrepancy has significant implications for portfolio construction, potentially leading to misallocation of funds across the 27 assets under consideration. This phenomenon may stem from a limited sample size; a hypothesis supported by the statistical principle of the law of large numbers which posits that estimation errors decrease as sample size increases. Hence, the deviation between the simulated and 'true' frontiers might be expected to diminish with a larger dataset that offers a more robust estimate of the underlying return distributions.

We now extend our analysis by simulating 100 different efficient frontiers for two separate cases: one with the number of periods equal to 200, and the other with 10,000 periods.

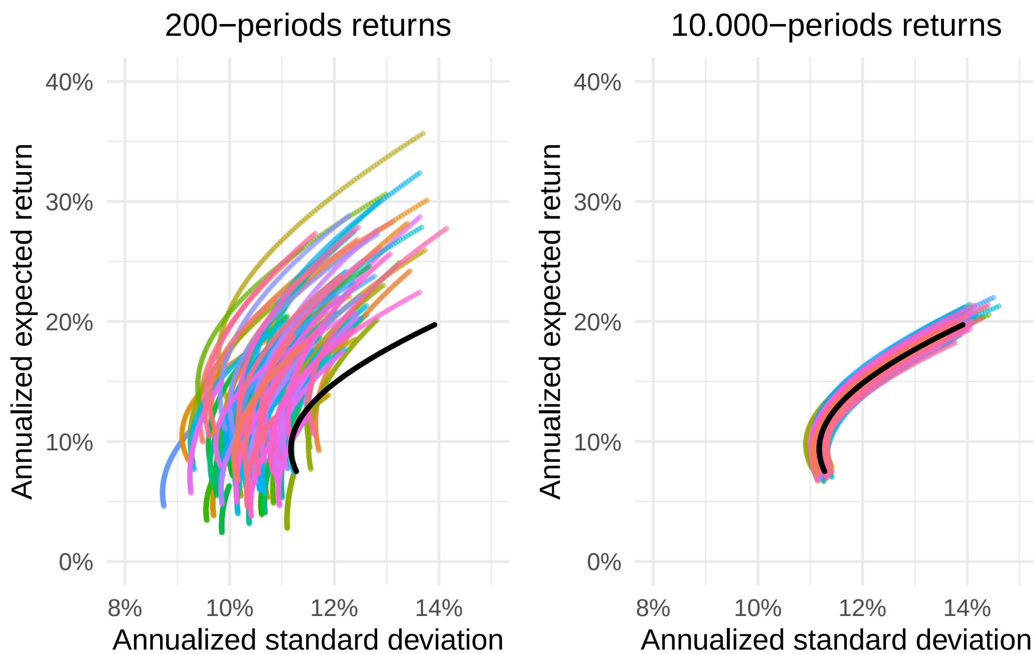


Figure 4: DGP with 100 simulated Efficient Frontiers

In the 200-period simulations, there is a noticeable tendency for the simulated efficient frontiers to outperform the 'true' efficient frontier, indicated by the black line. This divergence may be attributable to our underlying assumption that the data-generating process follows an i.i.d. $N(\mu, \Sigma)$. This assumption presumes symmetry in the error terms, which might not accurately reflect the actual return distributions that could exhibit skewness and non-normality.

From the 10.000 periods returns, we see that the variability in the efficient frontiers disappear. Additionally the 'true' efficient frontier seems to be in the average. Thus, if the DGP is correctly normal-distributed estimation uncertainty will decrease in the sample size.

Considering the above results for the efficient frontier, we now turn toward portfolio performance under estimation uncertainty.

## 1.2 Portfolio performance under estimation uncertainty

In this code segment, we conduct a Monte Carlo simulation to examine the behavior of tangency portfolio weights and sharpe ratios under estimation uncertainty. We begin by running simulations to generate synthetic return data, simulating potential real-world scenarios. Two empty data frames are initialized to store the results: `tangency_weights_df` holds tangency portfolio weights for each simulation, while `sharpe_ratios_df` stores the corresponding sharpe ratios. Within a loop for each simulation, we extract returns data and estimate parameters, including the covariance matrix and mean returns. Tangency portfolio weights are calculated using the `compute_efficient_frontier` function. We then compute the annualized sharpe ratio of the tangency portfolio using: $\sqrt{12}\frac{\hat{\omega}'_{\text{tgc}}\mu}{\sqrt{\hat{\omega}'_{\text{tgc}}\Sigma\hat{\omega}'_{\text{tgc}}}}$. Results, including weights and sharpe ratios, are stored in the respective data frames.

Below we show a histogram of the simulated sharpe ratio's with the sharpe ratio of the tangency portfolio with the true parameters, $\mu$ and $\sigma$:

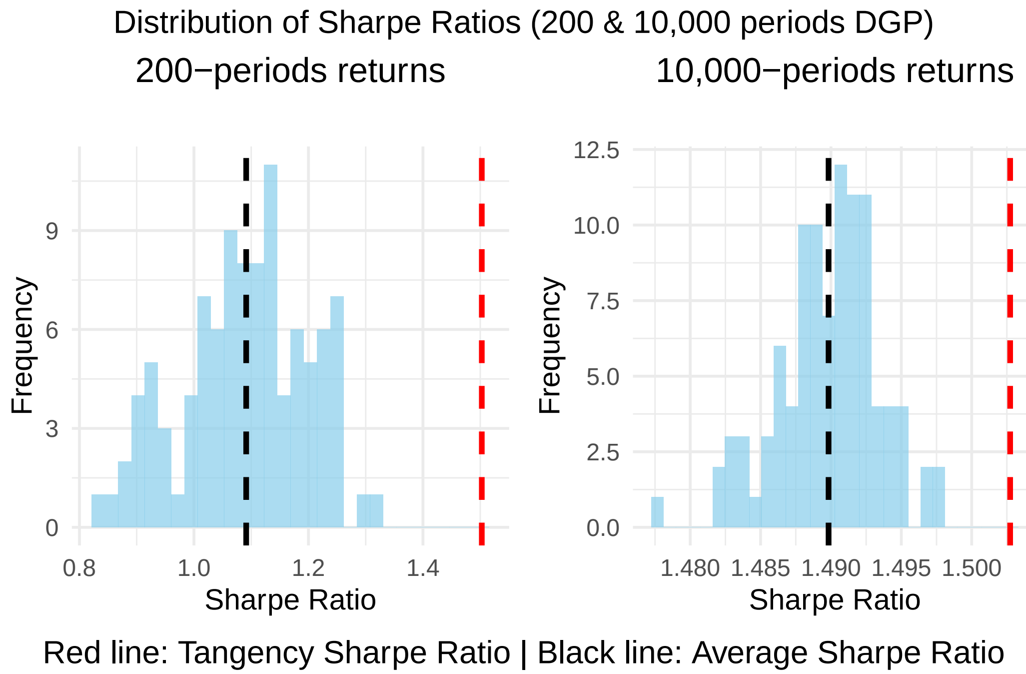


Figure 5: Distribution of Sharpe Ratios with 200 and 10.000 periods DGP

From the plotted histogram, it is obvious, that the share ratio of the tangency portfolio with the true parameters lies far above the average of the simulated sharpe ratios. The difference is 0.412. Clearly, this shows that estimation error can have big consequences for the implementation of the portfolio. Of further interest to be studied is, if a larger sample will reduce this uncertainty.

We show, that by increasing the sample period from 200 to 10.000, the simulated average sharpe ratio now moves much closer to that of the efficient tangency portfolio based on the true parameters. The difference falls from 0.412 to 0.013. Proving that the simulation sample size is very important, as it affects the accuracy and reliability of the estimated parameters and consequently the performance metrics of the portfolio. With a larger sample size, the estimation of expected returns and covariance matrix becomes more precise, leading to a distribution of simulated sharpe ratios that better reflects the true characteristics of the portfolio.

## 1.3 Alternative Portfolio Allocation Strategies

We note that as estimation uncertainty seems to be a significant issue, we could implement two estimation-free alternative allocation strategies.

1. An equally weighted portfolio does not suffer from estimation uncertainty, and if that is a major concern, then an equally weighted portfolio could be an option. However, it is unlikely that this would outperform the tangency portfolio, as it seems unreasonable to assume that historical covariance has no correlation with future covariance.

2. A market-weighted portfolio is another approach to address this issue, with the added bonus that, in theory, the tangency portfolio should be equal to the market portfolio. Therefore, theoretically, it should be as good or better than the tangency portfolio, as the market portfolio would be constructed with more sophisticated models that include expectations of future variance-covariance matrices.

Finally, it has been shown that variance-covariance matrices are very numerically unstable, which adds to the estimation uncertainty. One way to address this is to use covariance-shrinkage methods, which provide a biased but more stable estimator of variance-covariance matrices, potentially reducing estimation uncertainty. (We need a source for this and to implement it in the code).

## References