Stockholms universitet

Comparing the Global Minimum Variance portfolio to an equal weights benchmark in terms of risk and expected return

Hampus Olander

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2021:5
Matematisk statistik
Juni 2021

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2021:5**
http://www.math.su.se

Comparing the Global Minimum Variance portfolio to an equal weights benchmark in terms of risk and expected return

Hampus Olander*

June 2021

Abstract

In this thesis, the Global Minimum Variance (GMV) portfolio is compared to a bench mark portfolio, consisting of stocks held with equal weights. This is done by deriving two statistical tests that allows us to test whether the risk (or variance) and expected return of the two portfolios are significantly different.

The conclusion is firstly that there is no significant difference in the expected return; in fact the simple bench mark strategy outperformed the GMV portfolio during one of the three observed years. When it comes to the variances however, the GMV portfolio did as it is purposed to do by achieving a lower risk compared to the benchmark, for all three years under observation.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: hampus@ohlander.eu. Supervisor: Taras Bodnar.

Contents

1 Introduction

On one hand, the financial markets of the world exists to grease the wheel of the economies that allow for our way of life, by allocating capital to the most promising ventures and away from inefficient businesses. On the other hand, its purpose is to maximize profit; by generating the most return possible for its investors while simultaneously minimizing risk. Billions if not trillions of dollars are at stake, and it is therefore not surprising that the question of how to construct optimal portfolios is of immense interest.

There are many different ways of constructing financial portfolios; passive ones that are not based in data at all but simply consist of holding many different assets and hoping for the best; simple ones that use historic data to make educated guesses on what assets to buy; to funds relying on extremely complex mathematical models and highly paid individuals to do the requisite research. This thesis will compare a simple so-called bench mark portfolio, similar to an index fund, with a data-based portfolio known as the Global Minimum Variance portfolio. By accessing historical data and deriving a rigorous statistical test, we will investigate whether the more complex portfolio can outperform the simple one, both in terms of its return and its risk. This will be done for the years 2018, 2019 and 2020; the analysis done separately for each year.

Since not every reader may be familiar with the fundamentals of finance, the first few pages of this thesis will contain a very brief introduction to the subject of portfolio theory. However, basic knowledge in statistics, probability theory and linear algebra is expected from the reader.

2 Fundamentals of Portfolio Theory

In finance, a portfolio is a selection of financial assets that can be held in specific amounts. By constructing a portfolio, an investor chooses what assets to hold, and how much of each, based on his or her knowledge and beliefs of the market.

The total value of all the assets in a portfolio is less interesting than the relative portions in which they are held. An investor holding 30 \$ in asset A and 70 \$ in asset B can be said to hold the same portfolio as an investor holding 30 million \$ in asset A and 70 million \$ in asset B. Both investor operate under the belief that asset B is relatively better to hold than A than B (which can mean an expectation of higher return or lower risk for asset B, or a combination of these), but also that both A and B are, in combination, a better choice than all other assets that may be available to them. In essence, a portfolio is a choice of weights by which an investor decides how to invest his or her capital. A decision to not hold a particular asset can be viewed as that asset having been assigned a weight of 0.

One might ask why investors usually hold many different assets in their portfolio, and why not they are simply buying the asset they believe the most in. The answer to this is that by constructing a portfolio of many assets, investors take advantage of risk reduction by what is known as diversification. One can show that whenever two assets are not perfectly correlated, the lowest risk achievable will be a portfolio holding some combination of them. By "not putting all eggs in the same basket" an investor can reduce the total risk of their investments. Notably, this reduction in risk does not imply a corresponding reduction in expected returns, and it is this fact that makes diversification indispensable for any serious investor.

2.1 Return and Risk

Two fundamental concepts in finance are return and risk. Return is perhaps the most important metric that any investor concerns himself or herself with. Return is the percentage growth in value of an asset (note that a combination of assets, i.e. a portfolio, is itself an asset) over some period of time, and mathematically the return between periods 0 and t can be stated as

$$R = \frac{P_t - P_0}{P_0}, \text{ where } P_i \text{ is the price of the asset (or the portfolio) at time } i.$$

Since future prices are uncertain, this means that future returns is also uncertain and has to be modelled as a random variable. For that reason a major topic of interest is how to determine the expected return of assets, bearing in mind that bigger is better when speaking of returns.

But expected return is not the only important metric when comparing assets. In economics it is often assumed that rational agents are risk averse[1], meaning that the less risky an asset is, the better it is. Upon this basic assumption relies the

[1] [4], p. 160

entire insurance industry, in which the business idea is to allow customers to reduce their financial risk by paying a premium to an insurance company.

In a perfect world where we knew the exact probability distribution of the price (and hence the return) of an asset at some future time, we would have a perfect picture of the riskiness of that asset over that time period. This is a rather daunting task, so instead investors often just tackle the simpler problem which is to attempt to estimate the variance of the asset rather than its entire probability distribution. For this reason the variance, or the standard deviation, is the simplest and most commonly used measure of risk.

Note that assets are often correlated with one another, especially so if they for instance are stocks of companies on the same market. For an investor, this correlation is a good thing. While diversification is possible even in cases where individual assets are uncorrelated, with correlation it is possible to go one step further and balance a portfolio in such a way as to have one asset partially off-set the loss in a different one. This process is known as hedging, in which negative correlations are sought in order to greatly reduce risk without incurring a corresponding loss in expected return. We will return to this idea shortly when defining the GMV (Global Minimum Variance) portfolio, upon which this thesis is centered.

Assuming that we know the variance and covariances of individual assets, how do we then compute the portfolio variance? Note that the portfolio return is a linear combination of the asset returns, which are (generally) correlated random variables. The variance of a linear combination $Z = aX + bY$ can be written

$$V(Z) = a^2V(X) + b^2V(Y) + 2abCov(X,Y) =$$

$$(a,b)\begin{pmatrix} V(X) & Cov(X,Y) \\ Cov(X,Y) & V(Y) \end{pmatrix}(a,b)^T \text{ on matrix form.}$$

For the general case, given a portfolio p with covariance matrix Σ and with a set of weights w the variance of p is then

$$V_p = w^T \Sigma w.$$

An investor that has already decided on a portfolio (i.e. a set of weights) can of course also observe the performance of the portfolio periodically, take note of what the return has been and calculate the sample variance of those observations. This is an alternate, empirical way of estimating the portfolio variance. The issue with such a method is that it introduces another layer of uncertainty. In the method derived above, the only source of uncertainty in estimating the portfolio variance is how well we have estimated the covariance matrix. If we now construct a portfolio using the estimated covariance matrix and observe its performance, the randomness in the underlying asset performances causes a second round of uncertainty in our estimate, making it less certain.

2.2 Portfolios used for this thesis

2.2.1 GMV portfolio

So far we have concluded that different assets have different expected returns and different risks associated. For a rational and risk averse investor, higher expected return is always preferable, and so is lower risk. It is usually found that one can achieve higher expected return at the cost of higher risk, and the question that every investor has to face is how much risk they are willing to accept, knowing that higher risk means higher expected rewards.

One extreme strategy is to minimize the risk as much as possible. If one has knowledge of all the variances and covariances associated to some set of assets, finding the weights that minimizes the portfolio variance is a minimization problem with a known solution. If w are the weights, and C is the covariance matrix, the problem is to find $\min w^T C w$, which is solved using Lagrangian multipliers.

The solution to this is the weights of the so-called Global Minimum Variance (GMV) portfolio. If there are p assets to choose from the weights are found to be[2]

$$w_{GMV} = \frac{1_p^T C^{-1}}{1_p^T C^{-1} 1_p} \text{ where } 1_p^T \text{ is the } p\text{-dimensional vector of ones.}$$

Of course, the covariance matrix C is not known and has to be estimated, usually using historical data, in any practical situation. This is the approach taken later in this thesis. Once the GMV portfolio weights have been found one can once again use historical data to estimate the expected return of the portfolio.

The GMV portfolio is one extreme in which the investor tries to keep the variance at its lowest. It can be viewed as the ideal portfolio for the infinitely risk averse investor[3]. If we instead allow for more risk, other weights rendering higher returns can be chosen. This is the basis for the concept of the efficient frontier. The efficient frontier are all the points in the return-variance space (the points on the return vs variance graph) that are efficient, i.e. where the expected return is maximized given a particular level of variance, or equivalently where the variance is minimized given a particular level of expected return. It is hence a set of optimal portfolios, with different combinations of risk and returns. It has been shown mathematically that the efficient frontier is parabola shaped[4]. The efficient frontier is related to the parameter s which will be used later in the statistical test.

Finally, it is possible to compute the variance of the GMV portfolio directly using the covariance matrix (upon which the weights are based). This theoretical derivation of the portfolio variance is used later for the statistical test. While we do not provide any proof, it can be shown that

$V_{GMV} = \dfrac{1}{1_p^T \Sigma^{-1} 1_p}$, with Σ being the covariance matrix and 1_p being a p-dimensional

[2] [1], p. 73
[3] [2], p. 1
[4] [3], p. 318

matrix of ones.

2.2.2 Benchmark portfolio

Perhaps the most basic type of portfolio is one that relies on no analysis in its choice of weights, but instead relies upon the plurality of assets that causes diversification to make the portfolio worthwhile. Index funds are the most important example; portfolios that for instance incorporate all the stocks in a stock market with weights often chosen according to the relative size of the companies involved. The benefit of this is that no technical analysis of what the expected returns and covariances are is needed, making it a simple and low cost strategy for portfolio construction. For that reason it is often referred to as a passive investment strategy (an active strategy then being a strategy that relies upon actual analysis of individual assets).

These types of passive strategies are often considered a base line of investments; their returns are available to anyone because they require no knowledge of the market or of the individual assets. The challenge is to find strategies that improve upon this base line. The GMV portfolio is one such strategy, with the attempted improvement being a lower risk than any other portfolio including the base line.

A particular passive investment strategy is investment in a group of stocks with equal weights. This is one of the portfolios this thesis will analyze, hereafter referred to as the Benchmark portfolio. This will be compared to the other strategy that has been discussed above, namely the Global Minimum Variance portfolio. The expected return and variance of these portfolios will be compared by the use of statistical tests that are derived in later sections of this thesis.

Similar to the variance of the GMV portfolio, there is a theoretical result that allows for computation of the bench mark portfolio variance using the covariance matrix of the underlying assets. From the discussion in 2.1 the variance of the benchmark is

$V_b = b^T \Sigma b$, with Σ being the covariance matrix and b being the bench mark weights. This result is used to estimate $\hat{V}_b$ later when performing the statistical test.

3 Methodology

The idea of this thesis is to construct a GMV portfolio using historical data and compare it to a passive investment strategy; the bench mark portfolio, using actual real world returns. More precisely we are using data for four consecutive years, 2017 up to and including 2020. Data has been downloaded using the R-package tidyquant. It utilizes an API to allow us easy access to data from Yahoo Finance. The assets used in the analysis has been selected by picking 50 stocks randomly from among the 100 largest companies (by market cap) in the US as of today.

There are potential issues with this approach. By using historic data for the largest companies today, there is a selection bias since today's largest companies on average have been rather successful in order to reach that position. On the other hand, a top-100 company in 2017 that has not had the same stock price growth would not be part of the selection if it has dropped out of the top-100. If the purpose of the analysis were to analyze stock market performance in general, this bias would be detrimental. However, since we are interested in comparing two different portfolios, this is a minor issue.

When the 50 companies have been randomly selected, we proceed to download daily stock price data for the four year period. Using tidyquant, we can also access daily returns for each stock (recall that it is calculated from opening and closing prices by $R_{\text{daily}} = \dfrac{P_{\text{closing}} - P_{\text{opening}}}{P_{\text{opening}}}$).

The data on daily returns is then used both to construct the global mean variance portfolio (choosing the weights), and to analyze the portfolio performances. Note that the benchmark portfolio is not constructed using historic data at all, hence why it is called a passive strategy. The years for which we are testing the portfolios are 2018, 2019 and 2020. The GMV portfolio is constructed using data from the preceding year. For instance, using data of daily returns from 2017, we estimate the return co-variance matrix and the expected return vector for the 50 assets, and this is the basis for selecting the GMV portfolio weights for the entirety of the year 2018.

3.1 Why update weights only once a year?

One might question the choice to only update the portfolio once a year. After all, the portfolio we hold December 30th 2018 will be based entirely on 2017 data, and not include more recent data from earlier in the year 2018.

There are admittedly more sophisticated methods to choose portfolio weights. For instance one could use data from the preceding 365 days to calculate the weights of today, a so-called rolling window estimator. This would be a computationally more demanding method, but the more important reason for not choosing the method is that while it is theoretically feasible it is not practical. Updating the data set every day would cause the optimal portfolio weights to change every day, if only ever so slightly. This would mean that every day an investor would have to either buy or sell some portion of every asset they are holding, to adjust the portfolio according to the new weights. In a real world situation this would cause large transaction costs, including the fees paid to the marketplace for executing the transactions, and the value of the time spent updating the portfolio.

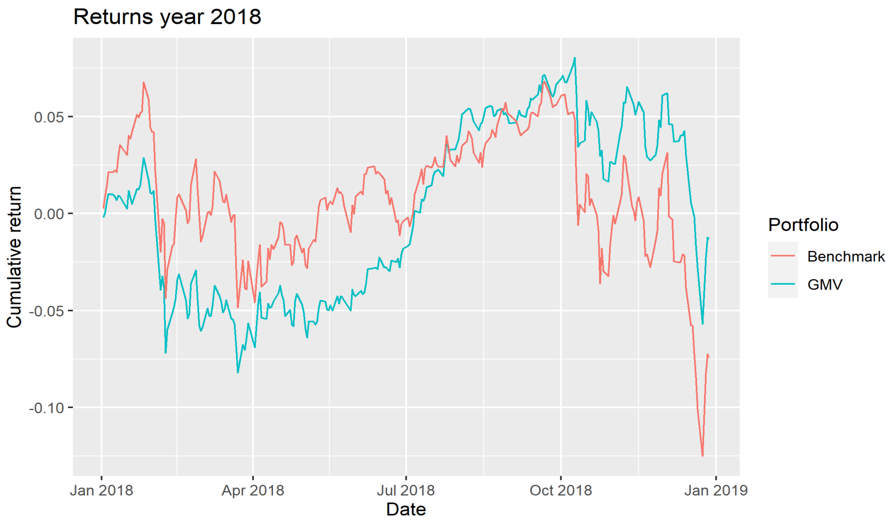
Of course, one could account for this by introducing more complex models such as only changing the portion allocated in a certain asset once the divergence between the optimal weights and what we currently hold reaches a certain threshold. If an asset initially makes up 0.1 of our portfolio, we could update the amount held once the optimal weight diverges from this by for instance 10 %, i.e. once the data dictates that we should hold less then 0.09 or more than 0.11 in the optimal portfolio. This would introduce an amount of inertia in the process as one would only update the portfolio once the difference between the actual and the optimal portfolio becomes sufficiently large.

While transaction costs can be ignored in a theoretical investigation of this kind, this thesis aims to stay somewhat true to the practical realities of trading. Also, as the objective of this thesis isn't to construct the most effective portfolio possible, but rather to compare two different and rather simple portfolios, the choice to update the portfolio once a year is deemed sufficient for this purpose.

4 Visualizing the data

4.1 Visualizing the cumulative return

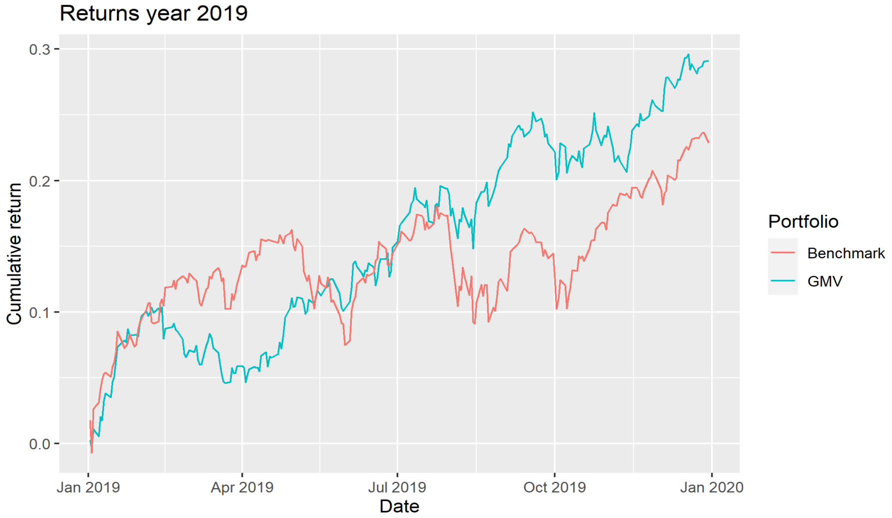
Before performing the statistical tests, it is always a wise idea to take a brief visual look at the data. The most common type of graph in terms of asset returns is cumulative returns over some period of time. For the year 2018, the two portfolios had the following cumulative return.



During the year 2018, both portfolios had total returns in the negative. This might come as a surprise given that in general, the stock market have had extraordinary returns in the last couple of years, but that is only true on average. In fact, the year 2018 was a remarkably bad year for the stock market, with the commonly used index Dow Jones being down by 5.6 % [5] during 2018, in line with our portfolios. Also note that the GMV did outperform the benchmark portfolio overall.

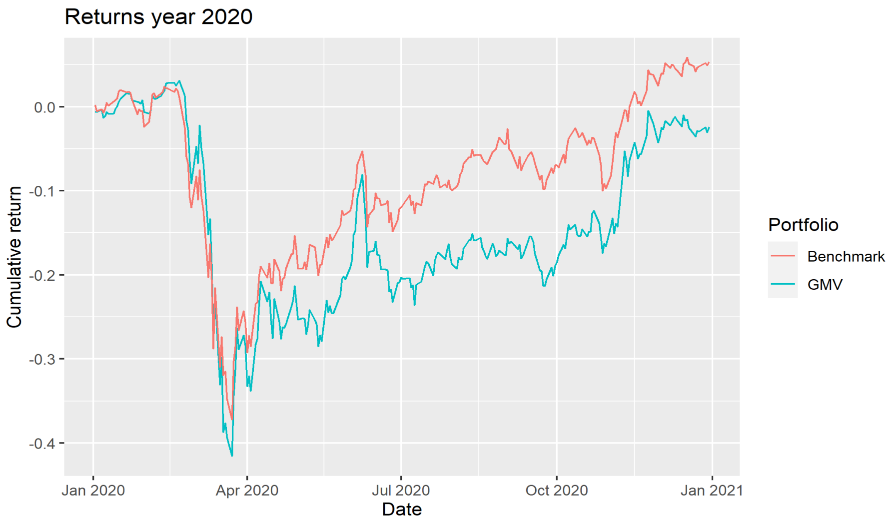
The same graph for 2019 looks like this

[5] https://www.macrotrends.net/1358/dow-jones-industrial-average-last-10-years



Once again the GMV portfolio has better returns on average. We will get back to analyzing whether this difference is statistically significant. Also note that 2019 was a remarkable year for the stock market as a whole.

2020 is the final year under observation.

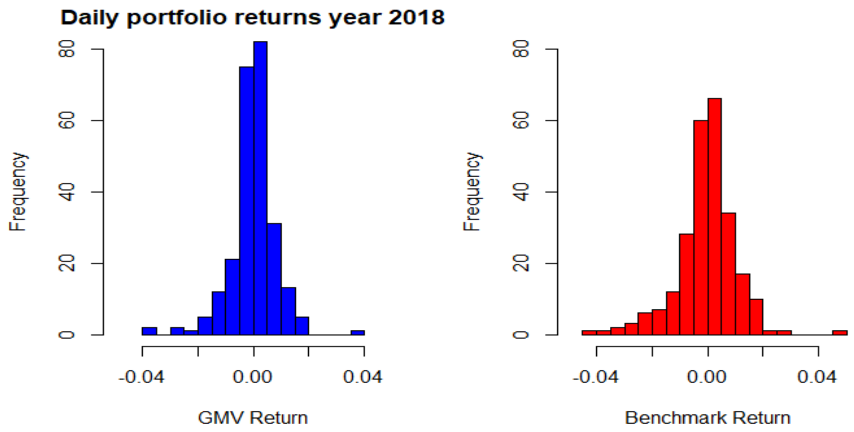


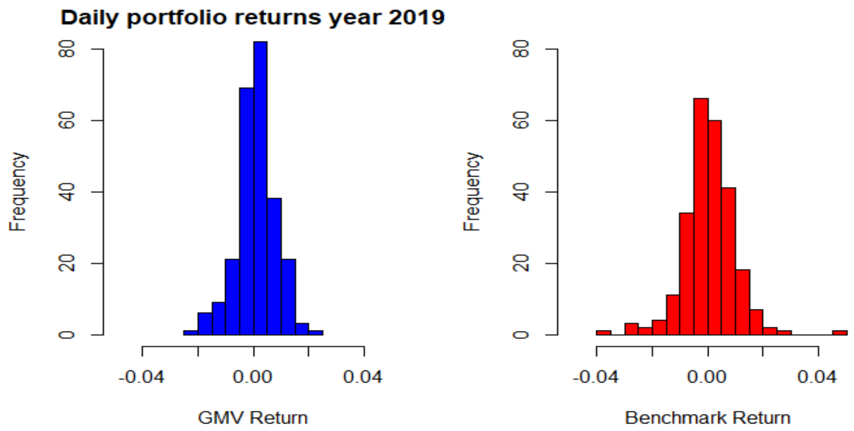
Of course, much can be said of the year 2020. As the world realized the severity of the COVID pandemic in March, the uncertainty caused the stock market to fall sharply, although it completely recovered during the remainder of the year. Interestingly enough, during 2020 the benchmark portfolio actually outperformed the GMV portfolio. We will get back to speculating why that is later on.

One observation is that from our data, it seems as if portfolio construction matters less than market performance when it comes to realized returns. This is a humbling fact worth bearing in mind. Even poorly planned portfolios perform well during good years, and the most sophisticated portfolios available should still be expected to lose value during bad years.

4.2 Visualizing the daily returns

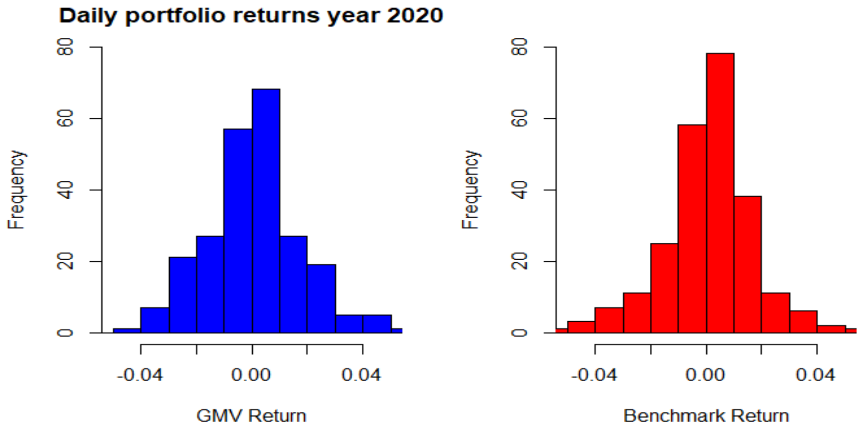
Another way to visualize the daily returns is with a histogram. This has the benefit of giving a sense of the distribution of daily returns, as well as allowing for comparison of differences in variance by looking at how wide or spread out the histograms are. For the years 2018 and 2019 we get the following graphs:





While the distributions of daily returns is not the focus of this thesis, as a curiosity it deserves mentioning that the daily returns appears to be roughly normally distributed. More importantly it seems like the GMV portfolio does have a somewhat smaller variance then the bench mark portfolio, because of its higher peak and lower spread. When it comes to comparing the returns the histogram is not the appropriate tool and we refer back to the previous section, or to the more rigorous analysis later in this thesis.

For the year 2020 these are the histograms:



The difference between the two portfolios is less pronounced and it is not possible to discern any difference in variance from these graphs.

After this brief visual inspection we turn to deriving actual tests for differences in expected return and variance for the two portfolios.

5 Deriving the statistical tests

In order to compare the portfolios in a meaningful manner, we need a test that can tell us if the portfolio returns and the portfolio risk (i.e. the variance) are significantly different. To do so, we need the distributions of the portfolio returns and the portfolio variances. Under the common assumption that the vector of asset returns is multivariate normal distributed, it follows that the daily return of a portfolio of assets is (univariate) normally distributed, since a portfolio is a simply a linear combination (or weighted average) of the individual assets. This is true for portfolios in general, and hence also for the benchmark and the GMV portfolios.

An exact expression for the distribution for the portfolio return and the variance can be derived from the paper *Statistical Inference for the Expected Utility Portfolio in High Dimensions* by Taras Bodnar et al., in which the authors prove the following asymptotic distribution.

Let R_{GMV}, R_b, V_{GMV} and V_b represent the expected daily returns and the variances in the daily returns for the GMV and the benchmark portfolios, respectively. Let the corresponding estimated values be denoted with hats. Furthermore, let s denote the slope of the efficient frontier. This will be discussed more in detail at the end of this section.

If we now define $\boldsymbol{t} = \begin{pmatrix} \hat{R}_{GMV} - R_{GMV} \\ \hat{V}_c - V_{GMV} \\ \hat{s}_c - s \\ \hat{R}_b - R_b \\ \hat{V}_b - V_b \end{pmatrix}$, then the following is true:

$\sqrt{n}\boldsymbol{t} \xrightarrow{d} \mathcal{N}_5(\mathbf{0}, \Omega_\alpha)$.

The $\boldsymbol{t}$ vector is asymptotically multivariate normally distributed, with some as of yet unspecified covariance matrix.

The covariance matrix of $\boldsymbol{t}$ depends on the same set of parameters as $\boldsymbol{t}$ does, so we cannot know it precisely. We can however use a consistent estimate $\hat{\Omega}_\alpha$. The parameters of this matrix will be defined and discussed in section 5.3.

$$\hat{\Omega}_\alpha = \begin{pmatrix} \frac{\hat{V}_c(\hat{s}_c+1)}{1-c} & 0 & 0 & \hat{V}_c & -2\hat{V}_c(\hat{R}_b - \hat{R}_{GMV}) \\ 0 & 2\frac{\hat{V}_c^2}{1-c} & 0 & 0 & 2\hat{V}_c^2 \\ 0 & 0 & 2\frac{(\hat{s}_c+1)^2+c-1}{1-c} & 2(\hat{R}_b - \hat{R}_{GMV}) & -2(\hat{R}_b - \hat{R}_{GMV})^2 \\ \hat{V}_c & 0 & 2(\hat{R}_b - \hat{R}_{GMV}) & \hat{V}_b & 0 \\ -2\hat{V}_c(\hat{R}_b - \hat{R}_{GMV}) & 2\hat{V}_c^2 & -2(\hat{R}_b - \hat{R}_{GMV})^2 & 0 & 2\hat{V}_b \end{pmatrix}$$

It deserves repeating that these results are true in the asymptotic case. More precisely, the vector $\sqrt{n}\boldsymbol{t}$ converges to the specified distribution when both n and p approaches infinity in such a way that the ratio tends to some constant c. This also the reason why the thesis focuses on daily returns, rather than for instance monthly. With daily returns, there are about 250 observations each year, which is deemed sufficiently large for the test to be applicable. This is also the reason why as many as 50 assets are used. It is also worth mentioning that using too many assets relative to the number of observations negatively impacts the estimation of the covariance matrix.

Using this distribution the required tests can be derived. When that has been accomplished, a discussion on what the parameters in the test represent and how they are calculated will follow.

5.1 Test for equal returns

The first step is to define the matrix $M_1^T = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \end{pmatrix}$. This gives us

$$M_1^T \sqrt{n}\boldsymbol{t} = \hat{R}_{GMV} - R_{GMV} - \hat{R}_b + R_b \xrightarrow{d} \mathcal{N}(\mathbf{0}, M_1^T \Omega_\alpha M_1).$$

It is helpful to introduce the quantity $\Delta_R = R_{GMV} - R_b$ and its estimate $\hat{\Delta}_R = \hat{R}_{GMV} - \hat{R}_b$ as the excess return for the GMV portfolio relative to the benchmark portfolio. This is precisely the quantity that we aim to investigate. The previous result can be rewritten in a more illuminating form;

$$\hat{\Delta}_R - \Delta_R \xrightarrow{d} \mathcal{N}(\mathbf{0}, M_1^T \Omega_\alpha M_1).$$

By performing the matrix multiplications we find that $M_1 \Omega_\alpha M_1^T = \dfrac{\hat{V}_c(\hat{s}_c + 1)}{1 - c} - 2\hat{V}_c + \hat{V}_b$. This is a quantity that has to be calculated for the test.

The null hypothesis that the expected returns for both portfolios are the same can be stated as $\mathrm{H}_0 : \ R_{GMV} - R_b = \Delta_R = 0$.

The alternative hypothesis that we will test is that they are not the same; $\mathrm{H}_1 : \ R_{GMV} - R_b = \Delta_R \neq 0$.

Under this regime it is clear that a two-sided Z-test is the correct approach.

5.2 Deriving the test for the portfolio variances

The test for portfolio variances is derived in much the same way. This time, let us define the matrix

$M_2^T = \begin{pmatrix} 0 & 1 & 0 & 0 & -1 \end{pmatrix}$. It follows that

$$M_2^T \sqrt{n}\boldsymbol{t} = V_b - V_{GMV} - \hat{V}_b + \hat{V}_c \xrightarrow{d} \mathcal{N}(\mathbf{0}, M_2^T \Omega_\alpha M_2).$$

With the substitution $\Delta_V = V_b - V_{GMV}$ this can be rewritten as

$$\Delta_V - \hat{\Delta}_V \xrightarrow{d} \mathcal{N}(\mathbf{0}, M_2^T \Omega_\alpha M_2).$$

By performing the matrix multiplications we find that
$$M_2 \Omega_\alpha M_2^T = \frac{2\hat{V}_c^2}{1-c} - 4\hat{V}_c^2 + 2\hat{V}_b^2.$$

Similar to before, the null hypothesis that the portfolio variances are the same can be stated as $\mathrm{H}_0 : \ V_b - V_{GMV} = \Delta_V = 0$.

The alternative hypothesis that we will test is that they are not the same; $\mathrm{H}_1 : V_b - V_{GMV} = \Delta_V \neq 0$.

This means that a second Z-test will have to be performed in the next section.

5.3 The parameters of the tests

Some attention needs to be paid to the parameters of the matrix used in the tests, what they represent and how they are calculated.

$\hat{R}_{GMV}$ and $\hat{R}_b$ are the estimated daily return for the respective portfolios and are found simply by calculating the respective mean of all daily returns, for the respective year.

The attentive reader will note that $\hat{V}_{GMV}$ is not present in the distribution of the test statistic, but that there is instead the parameter $\hat{V}_c$. The reason for this is that estimating $\hat{V}_{GMV}$ by the estimated covariance matrix gives a inconsistent estimate. The details of why this is are complicated[6] and arguably out of the scope of undergraduate courses. Without providing further motivation it turns out that the following adjustment has to be made

$\hat{V}_c = \dfrac{\hat{V}_{GMV}}{1-c}$, in order to reach a consistent estimate of the portfolio variance ('c' for consistent).

$\hat{s}_c$ is a parameter related to the efficient frontier discussed previously. More precisely s is the slope coefficient of the efficient frontier and $\hat{s}_c$ is a consistent estimator of s. Again the details fall outside the scope of this paper but the formulas for the calculations are[7]:

$\hat{s} = \hat{x}^T \hat{Q} \hat{x}$, with $\hat{x}$ being the vector of daily returns and $\hat{Q}$ being the estimate of a matrix that is related to the return covariance matrix Σ through

$Q = \Sigma^{-1} - \dfrac{\Sigma^{-1} 1_p 1_p^T \Sigma^{-1}}{1_p^T \Sigma^{-1} 1_p}$, with 1_p being the vector of ones with p assets.

[6] [2], p. 4
[7] [2], p. 3-4

Again, it can been shown that $\hat{s}$ defined this way is an inconsistent estimate of s, and the following adjustment has to be made:

$$\hat{s}_c = (1 - c_n)\hat{s} - c_n.$$

It turns out that $\hat{s}_c \stackrel{a.s.}{\rightarrow} s$ which justifies its use as the appropriate estimator.

There is a final parameter c that has already been used in estimating the other parameters. It was mentioned earlier that c is the ratio of the number of assets and the number of observations, and that the results are true asymptotically as both these tends to infinity. In practice we replace p with the number of assets, which are 50 throughout, and n with the number of daily observations for each year, which varies but are around 250.

6 Performing the statistical test

6.1 Testing the returns

The thing that remains is calculating all the necessary parameters and then performing the test that was derived in the previous section. If the test for the difference in returns is performed for the three years, the results are as follows. Recall that $\hat{\Delta}_R = \hat{R}_{GMV} - \hat{R}_b$, and hence if this is positive it means that the GMV portfolio outperformed the benchmark during that year.

Testing difference in portfolio return				
Year	$\hat{\Delta}_R \cdot \sqrt{n}$	σ $(for\ \Delta_V)$	Z-score	p-value
2018	0.003680084	0.004193054	0.8776619619	0.380161
2019	0.009852573	0.008941371	1.101909	0.270505
2020	-0.003110534	0.007661338	-0.4060041	0.684743

The conclusion is that there is no significant difference between the two portfolio returns for any of the years. In fact, as we noted when looking at the graph of cumulative returns in a previous section, the benchmark portfolio actually outperformed the GMV portfolio for 2020 (which the negative coefficient in the table confirms).

6.2 Testing the variances

Performing the corresponding test for difference in variance the results are as follows. Recall that $\hat{\Delta}_V = \hat{V}_b - \hat{V}_c$. As the purpose of the GMV portfolio is to minimize the risk (variance) of the portfolio, we suspect that this quantity should be positive.

Testing difference in portfolio variance				
Year	$\hat{\Delta}_V \cdot \sqrt{n}$	σ $(for\ \Delta_V)$	Z-score	p-value
2018	0.0001410448	0.00002407661	5.858167	< 0.00001
2019	0.0009279642	0.000141669	6.550229	< 0.00001
2020	0.0005268116	0.00008373297	6.291568	< 0.00001

The conclusion is that the GMV portfolio indeed has a significantly lower variance that the benchmark portfolio for all years under observation. The small p-values indicates a very high degree of certainty. It would be very surprising if this was not the fact, given that the purpose of the GMV portfolio is to minimize its variance.

7 Discussion

This investigation ends with the conclusion that the Global Minimum Variance achieved its purpose in reducing the portfolio risk compared to a naive passive strategy, and that this was true for all the observed years. In terms of expected returns however there was no statistically significant difference between these two portfolios. These conclusions are in line with what we could guess from looking at graphs of the data in section 4.

One cannot avoid to wonder what would happen with a more frequent updating of the GMV portfolio, by constantly using the latest data. This could be compared to the benchmark portfolio, or you could compare these two GMV portfolios using the different methods of estimation. Is there a statistically significant benefit in using a rolling window estimator instead of the once-a-year portfolio construction? There are various further investigations possible.

Another interesting question is the performance of the GMV portfolio during various different market conditions. We noticed that the GMV portfolio underperformed the benchmark portfolio during the very volatile year 2020 (although the difference was not significant) and one might ask if this fact has to do with the volatility. By studying many different years, perhaps grouped into 'more volatile' and 'less volatile' one could observe the performance of the GMV portfolio under different conditions.

References

[1] Marek Capinski, Tomasz Zastawniak. *Mathematics for Finance, An Introduction to Financial Engineering*, Second Edition, Springer, London, 2010.

[2] Taras Bodnar, Solomiia Dmytriv, Yarema Okhrin, Nestor Parolya, and Wolfgang Schmid. *Statistical Inference for the Expected Utility Portfolio in High Dimensions*, EEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 69, 2021.

[3] Taras Bodnar Wolfgang Schmid, *Econometrical analysis of the sample efficient frontier*, The European Journal of Finance, 15:3, 317-335, https://doi.org/10.1080/13518470802423478

[4] Zvi Bodie, Alex Kane, Alan J. Marcus. *Investments*, Twelth Edition, McGraw Hill, New York, 2021.

[5] Victor DeMiguel, Lorenzo Garlappi, Raman Uppal. *Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?*. The review of Financial studies, 22(5), May 2009.