UCLouvain

LOUVAIN
School of Management

Louvain School of Management

Analytical versus machine learning calibration: Which one works best in portfolio optimization?

Author: Arthur Rigal
Supervisor: Pr. Nathan Lassance
Academic year 2022-2023
Dissertation for the Master 120 in Business Engineering
Financial Engineering
Daytime schedule

Abstract

This master's thesis conducts empirical research on the estimation via machine learning of the decision parameters of different financial portfolio optimization strategies. Using the cross-validation method, the aim of the dissertation is to see whether the out-of-sample performance of portfolios estimated via machine learning is superior to that of conventional, analytically estimated, with strong statistical assumptions, portfolios.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm

Acknowledgments

I would particularly like to thank my supervisor, Professor Nathan Lassance from the Louvain-School of Management, who helped and guided me throughout the writing of my thesis. His expertise, availability and professionalism were the most valuable help I could have received.

I would also like to thank Professor Majeed Simaan from Stevens Institute of Technology, without whom the development of the empirical research code would have been much more difficult. His replication of Kan, Wang and Zhou's strategy was a great help.

Finally, I would like to thank the entire teaching staff of my Financial Engeneering major, who passed on to me its passion for Financial mathematics and were able to transfer all its academic dedication during these two brilliant years.

TABLE OF CONTENTS

Introduction

This master's thesis deals with the optimization of financial portfolios. Over the last few decades, several specialists have proposed mathematical optimization models. These models all share the same objective: to generate an optimal portfolio allocation (weight), i.e., the one generating the highest out-of-sample performance.

Among this literature, several authors have published articles on a very specific subject: portfolio shrinkage, i.e., methods for combining and improving basic portfolios such as those related to Harry Markowitz's modern portfolio theory (1952).

All these portfolio optimizations are based on an analytical methodology, with optimal parameter estimation required to calculate optimal weights. The aim of this thesis is to see whether, instead of estimating these parameters analytically, it would be possible to outperform the portfolios by optimizing the parameters via machine learning.

Indeed, estimation by analytical methods generates an exact mathematical solution, but under the constraints of strong statistical assumptions. In contrast, estimation using machine learning methods makes very few assumptions, but does not produce an exact mathematical solution. It only generates an approximation using numerical methods, and therefore allows for a high degree of flexibility.

Consequently, this master’s thesis will attempt to answer the question: "Analytical versus machine learning calibration: Which one works best in portfolio optimization?"

In this paper, the machine learning method chosen will be cross-validation, which will be applied to 4 different portfolio optimization methods: Covariance matrix shrinkage by Ledoit & Wolf (2004), portfolio shrinkage by Kan & Zhou (2007), portfolio shrinkage by Tu & Zhou (2010) and portfolio shrinkage by Kan, Wang & Zhou (2022).

This master's thesis will show that, for certain cases, machine learning calibration of financial portfolios can be very useful and efficient, and that the subject deserves to be investigated in greater depth using methods other than those used in this thesis.

Indeed, the empirical study will show that many times, calibration of shrinkage parameters via machine learning can generate better performance than analytical calibration. This was notably the case for a particular dataset, although the study also showed that calibration via machine learning.

As a result, the essay will proceed as follows:

First, a brief review will be given on the basic concepts of modern portfolio theory, without which the various optimization methods could not exist. An objective critique of their characteristics will attempt to explain the reason for the emergence of shrinkage methods.

Next, the four optimization methods investigated will be presented, along with their mathematical formulae for optimal portfolios.

After that, the features and results of a parallel empirical investigation using R software will be presented, suggesting several possible answers to the thesis.

Finally, an analysis of the results will be conducted, with the aim of answering the initial research question.

Theoritical study

1. Classical Methods

1.1. Markowitz

In 1952, in his article entitled "Portfolio selection"[1], Harry Markowitz laid the foundations for modern techniques in the mathematical optimization of financial asset portfolios. His method, called "Mean-Variance Portfolio", seeks to calculate the optimal asset allocation for which the portfolio will deliver a sufficient average return (defined according to an acceptability threshold) while presenting minimal volatility, and therefore risk.

We will see that, from a multitude of possible asset combinations, only one asset allocation will deliver the desired average return with the lowest volatility.

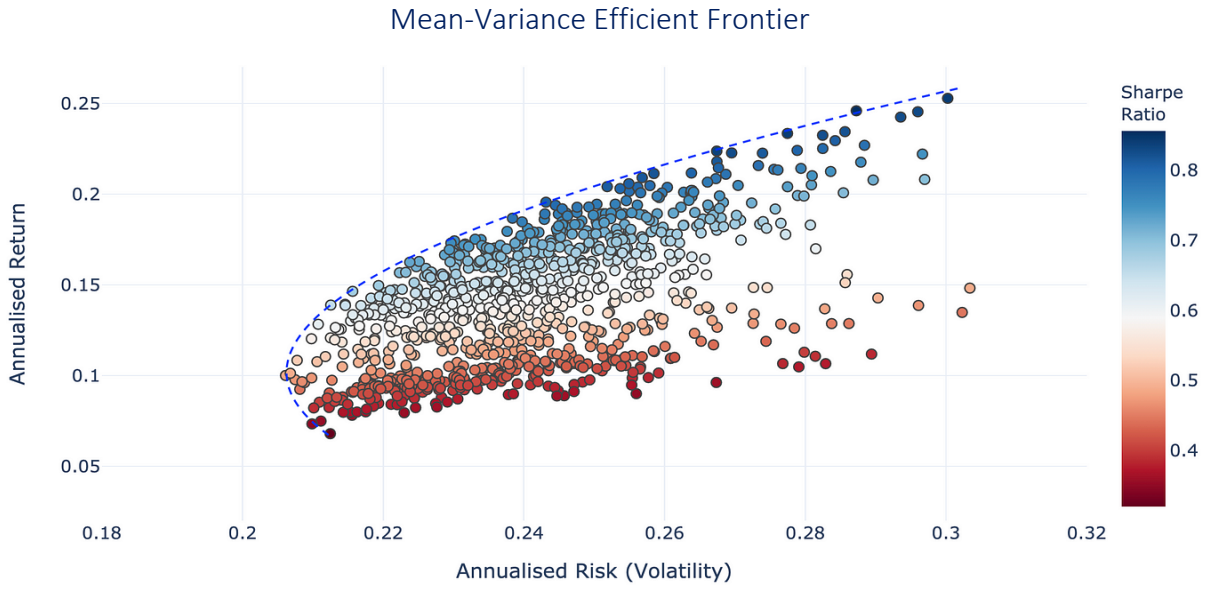

Figure 1: The Mean-Variance efficient frontier (in dotted blue line) of a portfolio made of daily returns for a universe of 6 stocks constructed by the most efficient portfolio (giving the highest Sharpe ratio for each level of risk).

Source: Toward Data Science[2]

Each point on the graph represents a specific asset allocation, or portfolio, which is therefore linked to a specific risk. The best-optimized allocations then form the efficient frontier, i.e., all allocations with the lowest risk rate for a given return - or, conversely, the highest returns for a

[1] Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), 77–91. https://doi.org/10.2307/2975974

[2] Dolphin, R. (2022, March 11). Efficient Frontier in Python | towards Data science. Medium. https://towardsdatascience.com/efficient-frontier-in-python-detailed-tutorial-84a304f03e79

given level of risk. Naturally, a rational investor will not be interested in the points below the curve, as they do not represent optimal allocations.

It is important to note that it is impossible to have points beyond the curve using only risky assets.

Here, the average portfolio return, R_p, is defined by:

$$E(R_p) = \sum_i w_i \, E(R_i)$$

where R_i and w_i are respectively the return and weight associated with asset i.

In Markowitz's theory, two different cases can be considered: either the universe includes a risk-free asset, and the investor can therefore decide to allocate part of his portfolio to it, or there is no risk-free asset, and his portfolio will be entirely made up of risky assets, leading as a result to a sum of weights equal to 1.

In both cases, the portfolio will consist of N risky returns. The difference is that, if a risk-free asset exists, an additional investment can be made in it.

In case of a risk-free asset universe, the investor will have the choice to invest a certain proportion of his portfolio in it. The resulting optimal portfolio will therefore be on the tangent point linking the risk-free asset amount to the efficient frontier such as:

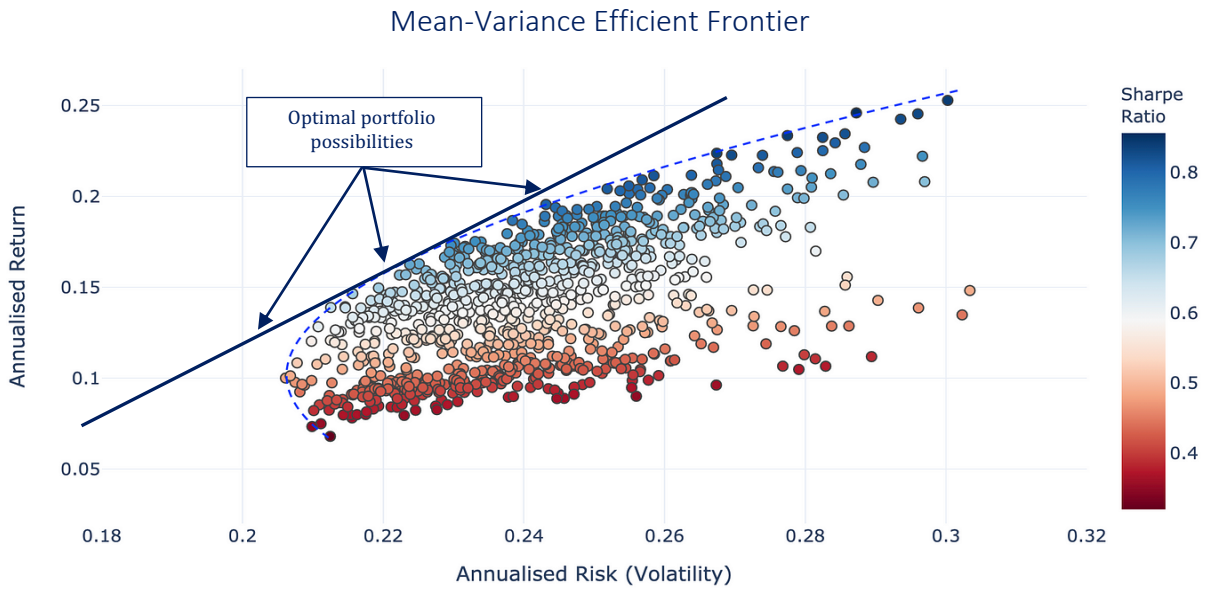

Figure 2: The Mean-Variance efficient frontier with the optimal portfolio combining risk-free and risky assets.

Source: Toward Data Science

The location of the optimal portfolio on the tangent will be defined by the investor's risk aversion.

By defining μ as the vector of mean excess returns and Σ as the covariance matrix of returns and γ the risk aversion rate (varying from 1 to 5), it is possible to generate the optimal weight equations as:

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

in a risk-free asset universe, with $1 - w'1_N$ investments in it
where γ is the risk-aversion coefficient, Σ the sample covariance matrix and
μ the vector of means.

The equation for the weights of the fully invested Mean-Variance portfolio (without investing in the risk-free asset) will be reviewed in [1.2] as it needs beforehand other computation.

Efficient frontiers are obtained by varying risk-aversion coefficient, where a new optimal allocation w is created for each variation in γ.

Although it is the basis of modern financial mathematics, several criticisms have been levelled at the Mean-Variance portfolio theory:

- **Sensitivity to parameter estimates:** Markowitz optimization relies on the estimation of expected returns and asset variance-covariance matrices. However, the theory relies only on past observations to estimate future values of μ and Σ. These estimates can be sensitive to fluctuations in the historical data used and cannot consider parameters or characteristics that are not reflected in the past, which can introduce some uncertainty into the results. This is a major concern in practice, since the entire out-of-sample portfolio will only be an estimate of the in-sample portfolio based on these past data.

  Thus, since Mean-Variance optimization is based solely on past observations, certain one-off events, which may reflect abnormally ideal characteristics (abnormally large return, negative correlation, small variance, etc.), will tend to cause observation means to vary considerably, leading to erroneous optimization choices. Furthermore, as Chopra & Ziemba (1993)[3] show, errors of mean are even more worrying since they are up to 10 times more impactful than errors of variance.

3 Chopra, V., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. The Journal of Portfolio Management, 19(2), 6–11. https://doi.org/10.3905/jpm.1993.409440

By way of illustration, in the case where an asset has in the past presented extremely large returns, Mean-Variance theory will base its optimization without really considering the exceptionality of such an event, and will therefore generate a relatively erroneous mean and covariance matrix for predicting the asset's future behavior.

As a result, certain past assumptions will lead to highly concentrated weights on assets that will no longer have the same characteristics in the future, multiplying the damage caused by such estimation errors.

Indeed, the Mean Variance portfolio often tends to maximize the effects of estimation errors in input assertions, leading to a portfolio allocation that is falsely optimal, or at least financially irrelevant, since the optimization will tend to assign extreme weights to assets with ideal characteristics (abnormally large return, negative correlation, and small variance), and which are of course those most prone to large maximization errors.

- **Instability of the optimal solution**: As Best & Grauer (1991)[4] demonstrate, Markowitz optimization is in some cases extremely unstable, due in particular to the fact that a tiny change in input assumptions can lead to a totally different optimal solution.

  This problem is explained by the fact that the covariance matrix can sometimes be very difficult to invert, leading to an "ill-conditioned" character that may be caused by too many inputs.

- **Lack of interest in higher moments**: As Lassance (2021)[5] explains in his article, Mean-Variance theory assumes that investors' preferences only concern the mean and the v variance, leading to a certain lack of interest in higher moments (skewness and kurtosis).

  Furthermore, due to the large estimation errors in the upper moments, it is often preferable to stay on the efficient frontier even if you have preferences towards the upper moments.

4 Best, M.J. and Grauer, R.R. (1991) On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. The Review of Financial Studies, 4, 315-342. https://doi.org/10.1093/rfs/4.2.315

5 Lassance, N. (2022). Reconciling mean-variance portfolio theory with non-Gaussian returns. European Journal of Operational Research, 297(2), 729–740. https://doi.org/10.1016/j.ejor.2021.06.016

1.2. Minimum-Variance

Markowitz's Mean-Variance theory implies the construction of an efficient frontier generated by the optimal allocations for a given level of risk. If an investor (calculating his portfolio based on Mean-Variance theory) decides to minimize the risk associated with his portfolio, he will choose the allocation represented by the point furthest to the left of the efficient frontier. This portfolio is called the Minimum-Variance portfolio and, as its name suggests, has the smallest variance - and therefore the smallest risk.

We will illustrate it by the Mean-Variance graph used previously:

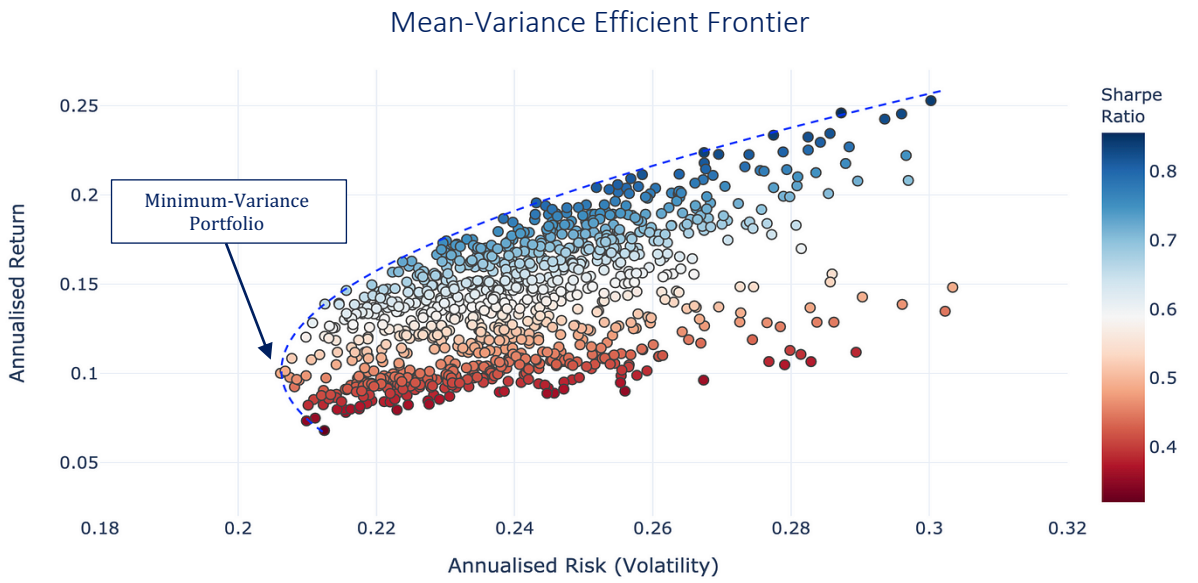

Figure 3: The Mean-Variance efficient frontier with the highlighted Minimum-Variance Portfolio
Source: Toward Data Science

We can see on the graph that the Minimum-Variance portfolio is the leftmost point, reflecting the lowest risk.

This allocation quickly became popular with investors, given the problems associated with estimating errors in Mean-Variance portfolios. Choosing to use such a portfolio means having the smallest variance, and therefore the lowest risk associated with it, regardless of expected returns, and therefore μ.

Indeed, as Chopra & Ziemba (1993)[3] show in their paper where they quantify that errors in μ have a much bigger impact (up to 10 times bigger) than in Σ, Markowitz portfolios generally perform worse than the Minimum-Variance portfolio because the latter depends only on Σ and not on μ.

To achieve this minimum volatility, the Minimum Variance portfolio will tend to favor asset allocations with negative correlations, thereby reducing total risk as the loss on one asset is cancelled out by the gain on the other.

We can construct the Global Minimum-Variance portfolio as follows:

$$w_{GMV} = \frac{\Sigma^{-1} 1_N}{1'_N \Sigma^{-1} 1_N}$$

And its mean will be equal to:

$$\mu_{GMV} = w_{GMV}{}' \mu$$

Thanks to these computations, we can then generate the fully invested Mean-Variance portfolio such as:

$$w = w_{GMV} + \frac{1}{\gamma} w_z$$

in a no risk-free asset universe, under the constraint that $w'1_N = 1$
where w_{GMV} is the Global Minimum-Variance portfolio and $w_z = \Sigma^{-1}(\mu - \mu_{GMV}\ 1_N)$ the plug-in portfolio.

Somewhat simple in its implementation, this technique offers several advantages:

- **Less sensitive to estimates:** Because Minimum-Variance optimization is not concerned with asset performance, it will not require precise estimates of expected asset returns, making the approach less sensitive to estimation errors.

- **Greater robustness against systemic risk:** Thanks to its minimal risk and its inverse correlation, Minimum-Variance optimization will generally be more resilient to periods of stress or crisis, when correlations between assets, and therefore possible losses, can increase significantly.

As mentioned above, it can be seen in absolute terms that the Minimum-Variance portfolio can often outperform (or lose less) over the long term than other Mean-Variance portfolios. However, such a conclusion is not always true, and it is preferable first to refer to the individual preferences of investors, who may in other cases be preferable to first refer to Markowitz-determined portfolios if asset returns play a crucial role in decision-making.

On the other hand, although the Minimum-Variance portfolio has considerable advantages, it also has some disadvantages to consider:

- **Low performance:** As the optimization does not take returns into account, these are not guaranteed, and the portfolio may therefore show low performance despite very low risk. Such an effect is augmented within the context of a market upswing.

- **Concentration bias:** As the optimization only relies on risk minimization, it will tend to highly concentrate its weights on uncorrelated assets, leading therefore to a non-diversified portfolio.

1.3. Naive 1/N

Over the last few decades, several naive portfolio optimization methods have been developed, in an attempt to understand the conditions under which optimal Mean-Variance (including Minimum-Variance) portfolio models can perform well even in the presence of estimation risks. Several "naive" strategies, i.e., those that are extremely easy to compute, are used to compare performance with optimization models. The best-known of these is the 1/N strategy, which, as its name suggests, assigns equal weights to each asset, i.e., 1 divided by the number of assets in the portfolio.

In their 2009 paper, DeMiguel, Garlappi & Uppal[6] show that when portfolio optimizations are based on estimated elements, the naive 1/N strategy often works better than more complex portfolios. Indeed, since the two major problems with optimal portfolios are the difficulty of estimating parameters and computational instability (relating to the covariance matrix), they will in some cases be inefficient, compared with naive strategies, which do not encounter these two problems. The 1/N portfolio also offers other advantages over optimal portfolios, including maximum diversification and reduced operational costs.

However, it is important to note that the naive 1/N portfolio remains a (sometimes overly) simplistic method, and can also have its drawbacks, notably real underperformance compared to properly optimized portfolios under certain conditions, such as when assets have different characteristics or significant correlations. In addition, it does not consider the investor's individual preferences or specific objectives.

Ultimately, the choice between a Markowitz portfolio and a naive portfolio depends on the preferences and constraints of each investor. Some may prefer the simplicity and robustness of the naive portfolio, while others may opt for the more sophisticated, customized approach offered by portfolio optimization.

6 DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/NPortfolio Strategy? Review of Financial Studies, 22(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

2. Shrinkage Methods

The literature offers a multitude of improvements to these rather simple methods. Much of it concerns shrinkage methods, i.e., methods for obtaining better estimates of the parameters involved. Often based on a combination of several basic methods, they deliver better results by combining the advantages of several basic techniques, while in most cases retaining a certain computational ease.

4 shrinkage methods will be presented in this section:

- **Ledoit & Wolf shrinkage**, which focuses on the shrinkage of the sampling covariance matrix.
- **Tu & Zou shrinkage**, which focuses on the combination of Markowitz and 1/N.
- **Kan & Zou shrinkage**, which is concerned with the combination of Markowitz and Min-Var in the presence of a risk-free asset.
- **Kan, Wang & Zou shrinkage**, which looks at the combination of Markowitz and Min-Var without a risk-free asset.

2.1. Ledoit-Wolf

In 2004 Ledoit & Wolf [7] published an article describing a methodology for shrinking the covariance matrix. This approach aims at improving covariance matrix estimation by reducing its volatility and improving its stability, particularly when the number of observations is limited compared to the number of assets, resulting in a highly unstable covariance matrix.

Indeed, in Markowitz theory, there may be situations where the sample covariance matrix contains too many estimation errors, which disrupt the Mean-Variance optimizer and leave room for a low-quality optimization.
The aim of Ledoit & Wolf's work is to propose a method for shrinking this covariance matrix to produce a new one whose extreme values are corrected, thus greatly reducing the amount of estimation error.

This shrinkage method will therefore take the extreme values (which are lacking in the model) and bring them back towards the center (hence the word "shrinkage", since the spectrum of estimates will be narrowed). Ledoit & Wolf will propose an algebraic solution to define the optimal shrinkage target - i.e., a threshold in the spectrum beyond which estimates are qualified as extreme and therefore require correction - and at what intensity to perform the refocusing.

Let us return to the Markowitz portfolio, which defines the optimal portfolio as:

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

As mentioned above, the main problem with the Mean-Variance portfolio is its difficulty in obtaining good results due to estimation errors reflected in the covariance matrix. Ledoit & Wolf therefore propose a new shrinkage covariance matrix to correct for extreme values, called the Σ_{Shrink} and defined by:

$$\Sigma_{Shrink} = \delta F + (1-\delta)S$$

where S is the sample covariance matrix, $F = \mu I$ is a structured estimator representing the target matrix where μ is the mean of the elements of the diagonal of S and I the identity matrix (see Ledoit & Wolf 2004 bis)[8], and δ is the shrinkage intensity between 0 and 1, found analytically under strong statistical assumptions such as independent and identically distributed returns (i.i.d.).

7 Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, *30*(4), 110–119. https://doi.org/10.3905/jpm.2004.110

8 Ledoit, O., & Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365–411. https://doi.org/10.1016/s0047-259x(03)00096-4

We can see that Σ_{Shrink} is a linear combination of the sampling covariance matrix S and the target matrix F. The proportion of each is defined by the shrinkage intensity δ.
Ledoit & Wolf estimate this combination parameter δ analytically by minimizing the quadratic loss function characterized by the Frobenius norm of the difference between the shrinkage estimator and the sampling covariance matrix *S*:

$$L(\delta) = ||\delta F + (1-\delta)S - \Sigma||^2$$

For additional information on the computation of δ, please refer to Ledoit & Wolf (2004) [2.1].

Therefore, the new optimal portfolio based on Ledoit & Wolf's shrink-wrapped covariance matrix can be defined as:

$$w_{LW} = \frac{1}{\gamma} \Sigma_{Shrink}^{-1} \mu$$

Before delving further into the exploration of other shrinkage methods, it is important to mention the basis on which the optimal combination parameter is chosen. Indeed, each of the following methods will combine "classic" portfolios according to a specific distribution illustrated by a parameter.

Although analytically different, the equations for the optimal combination parameters are based on a maximization of the portfolio's expected out-of-sample utility, calculated as:

$$U(w) = \mu_p - \frac{\gamma}{2} \sigma_p^2$$

where $\mu_p = w'\mu_t$ is the average portfolio return and $\sigma_p^2 = w'\Sigma w$ its variance.

It is also important to point out that various authors make two assumptions about excess returns. First, the standard assumption about their distribution is that excess returns are independent and identically distributed (i.i.d.). Second, they assume that excess returns follow a multivariate normal distribution, with μ and Σ as maximum likelihood estimators of the mean and covariance matrix respectively.

2.2. Kan-Zhou

In 2007, Kan & Zhou[9] demonstrated that the portfolio composed of the risk-free asset and its relative tangent to the Markowitz efficient frontier can be fully outperformed by combining it with the Minimum-Variance portfolio, suggesting that the presence of estimation risk still enables better performance than an risk-free portfolio.

Indeed, as explained above, Mean-Variance theory assumes that an investor who knows the true value of the parameter will only invest in a combination of the risk-free asset and the tangential portfolio. The reality is that the true value of the parameter is unknown and must therefore be estimated, leading to estimation errors. Combining with another risky portfolio can diversify the estimation risk of the tangential portfolio, thereby lowering the risk of the overall portfolio. This effect is due to the fact that the estimation errors of the two portfolios are not perfectly correlated.

Their choice to combine the tangential portfolio with the Minimum-Variance portfolio stems from the fact that the weights of the latter depend only on the sample covariance matrix.

Thus, by adding the minimum-variance constraint, Kan and Zhou construct the three-fund portfolio by combining risk-free asset, tangential Mean-Variance and Minimum-Variance portfolios.

The optimal weights for the Kan & Zhou portfolio will therefore be:

$$w_{KZ} = c_3(c\, w_{MV} + (1-c)\left(\frac{\mu_{GMV}}{\gamma}\right)\Sigma^{-1}1_N)$$

where $c_3 = \frac{(T-N-1)(T-N-4)}{T(T-2)}$ is a constant scalar, c is the combining parameter, w_{MV} is the Mean-Variance portfolio and $\mu_{GMV} = \frac{1'_N \Sigma^{-1}\mu}{1'_N \Sigma^{-1}1_N}$ is the mean of excess return of the Minimum-Variance Portfolio.

And the optimal value of c as computed by Kan & Zhou is:

$$c^* = \frac{k\Psi^2}{\Psi^2 + \frac{N-1}{T}}$$

where k is a scalar constant and Ψ the ex-ante Minimum-Variance frontier. For additional information on the computation of c, please refer to Kan & Zhou (2007) [III].

[9] Kan, R., & Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, *42*(3), 621–656. https://doi.org/10.1017/s0022109000004129

2.3. Tu-Zhou

In 2011, Tu & Zhou[10] proposed another shrinkage method. Its principle is simple: combine the Markowitz method with the naive 1/N portfolio (DeMiguel, Garlappi & Uppal, 2009) according to a combination coefficient δ.

Indeed, as explained above, the naive portfolio often generates better results than sophisticated investment strategies, due to the need to estimate optimal portfolios. The idea of combining it with the Mean-Variance portfolio may therefore make sense to solve the fatalistic aspect of the 1/N portfolio, having no strategy at all, and thus may generate better results than the naive strategy, while reaffirming the usefulness of the theory in practice.

As a result, the optimal weights for the Tu & Zhou portfolio will be:

$$w_{TZ} = \delta\, w_{MVtz} + (1-\delta) w_{1/N}$$

where δ is the combination parameter, $w_{MVtz} = \frac{(T-N-2)}{T} w_{MV}$ is the Tu & Zhou-adjusted Mean-Variance portfolio and $w_{1/N} = \frac{1_N}{N}$ is the naive 1/N portfolio.

And the optimal value of δ, as computed by Tu & Zhou is equal to:

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

where π_1 measures the impact from the bias of the 1/N rule and π_2 measures the impact from the variance from w_{MVtz}. For additional information on the computation of c, please refer to Tu & Zhou (2010) [2.1].

10 Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. Journal of Financial Economics, 99(1), 204–215. https://doi.org/10.1016/j.jfineco.2010.08.013

2.4. Kan-Wang-Zhou

In 2022, Kan, Wang & Zhou[11] address the same problem as Kan & Zhou (2007), but in the case where risk-free assets are not available. Indeed, their postulate stems from the fact that in many investment cases, such as investment funds, the portfolio could only focus on risky assets, as risk-free assets did not exist.

In their research, they explain that the idea of adapting Kan and Zhou's (2007) three-fund portfolio by simply reweighting the weights of risky assets so that their sum equals 1 does not result in an optimal portfolio. Their aim is then to remodel an equation combining Markowitz and Minimum-Variance portfolios, but this time in a universe devoid of risk-free assets.

This results in an optimal weight equation as follows:

$$W_{KWZ} = w_{GMV} + \frac{c}{\gamma} w_z$$

where w_{GMV} is the global Minimum-Variance portfolio, c the shrinkage parameter and $w_z = \Sigma^{-1}(\mu - \mu_{GMV}\ 1_N)$ the sample zero-investment portfolio with $\mu_{GMV} = \frac{1'_N \Sigma^{-1}\mu}{1'_N \Sigma^{-1} 1_N}$ the mean of excess return of the Minimum-Variance Portfolio.

The shrinkage parameter c is estimated by Kan, Wang & Zhou as:

$$c^* = \frac{k\Psi^2}{\Psi^2 + \frac{N-1}{T}}$$

where k is a scalar constant and Ψ the ex-ante Minimum-Variance frontier. For additional information on the computation of c, please refer to Kan, Wang & Zhou (2022) [2.3].

[11] Kan, R., Wang, X., & Zhou, G. (2022). Optimal Portfolio Choice with Estimation Risk: No Risk-Free Asset Case. *Management Science, 68*(3), 2047–2068. https://doi.org/10.1287/mnsc.2021.3989

Empirical study

The empirical part of this dissertation will focus on judging the optimization capability of machine learning methods in the context of estimating the parameters of the equations in the theoretical part, which were previously found analytically.

A comparison will therefore be made between the performance of portfolios where the shrinkage/combination parameter is found analytically under strong statistical assumptions (normal distribution, i.i.d.), as in the case of the original papers, and by a machine learning method.

In this way, performance relative to the returns of portfolios found analytically will be compared with the performance of portfolios under machine learning calibration, leading to an objective study of the ability of machine learning to estimate optimal parameters.

1. Datasets

In order to make the best possible comparison between the two techniques, I have chosen to use 5 different datasets, namely:

- **10 Industry Portfolio**, composed of 10 assets representing a summary of the 10 largest industries.
- **48 Industry Portfolio**, composed of 48 assets representing a summary of the 48 largest industries.
- **25 Portfolios Formed on Size and Book-to-Market**, composed of the intersection of 5 portfolios formed on size (market equity, ME) and 5 portfolios formed on the book equity to market equity ratio (BE/ME).
- **25 Portfolios Formed on Operating Profitability and Investment**, made up of the intersections of 5 portfolios formed on profitability (OP) and 5 portfolios formed on investment (Inv).
- **10 Portfolios Formed on Momentum**, made up of the 10 NYSE prior (2-12) return decile breakpoints.

Although completely arbitrary, this choice of datasets makes it possible to investigate the subject on several samples that are totally different from one another in their composition.

For more information, all these portfolios and their characteristics can be found on the Kenneth R. French website[12].

To these portfolios, it is important to subtract the risk-free rate, which can also be found on the same site.

For the sake of accuracy and quality of comparison, the datasets will have the same monthly observation windows, i.e., from July 1963 to May 2023, a period of 719 observations.

2. Methodology

Both the analytical and machine learning parts of the process will be based on the same methodology. Portfolio performance will be evaluated according to their out-of-sample utility, which in turn will be calculated using the rolling windows technique. For the latter, I have chosen a 10-year window, i.e., 120 months. The rolling windows technique is performed as follows:

Each 120-month window generates a specific portfolio. This portfolio is applied to the following month, the 121st, to obtain out-of-sample returns for each asset, which when added together form the overall return for the month according to the optimal allocation calculated over the previous 120 months. For each dataset, the first 120 months (July 1963 to June 1973) will not produce any out-of-sample returns, as they will serve as the first training window.

In the end, the rolling windows technique will have generated 599 out-of-sample returns (from July 1973 to May 2023), based on which the overall utility will be calculated.
The two final utilities will then be compared qualitatively, to see which method works best.

On the analytical side, the methodology is relatively straightforward: optimal weights are calculated by first calculating the optimal parameter using analytical resolution.

[12] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Once the optimal weights, and therefore the optimal portfolio, have been found, applying them to the asset returns enables us to calculate the out-of-sample returns over the entire test period via rolling windows, and thus find the portfolio's overall utility.

In contrast, the machine learning technique tests all possible values of the optimal parameter, resulting in a variety of different portfolios. The choice of the optimal parameter is once again based on utility, but is tested via cross-validation, enabling greater precision of choice:

Within each window, all parameter values are tested (in 0.01 steps), from 0 to 1.
For each parameter value, the window is divided into a precise number of folds. This thesis will carry out three different cross-validations: 10 folds, 3 folds and k folds (k being the number of observations). Each fold will be removed one-by-one from the window, so that the model can be trained to calculate the optimal weights on the remaining folds. The weights will then be applied to the removed fold to generate out-of-sample yields. When all the folds have been tested one by one, the cross-validation will have generated a total of 120 out-of-sample returns. The overall utility will then be calculated based on these 120 returns.

In short, each value of the optimal parameter will generate 120 returns, which in turn will generate a utility. The utilities for each parameter value are finally compared, with the highest utility representing the optimal parameter.

Once the optimal parameter, and therefore the optimal weights, have been found, the method of evaluating the machine learning portfolio is similar to that of the analytical portfolio, namely the calculation of the utility of out-of-sample returns via rolling windows.

3. Optimizations & Results

While each optimization method has its own resolution for calculating the optimal shrinkage parameter value, calibration via machine learning remains identical and common to all methods. Indeed, as explained above, the latter tests all parameter values and selects the one that generates the highest out-of-sample utility.

Each window, whether analytical or machine learning, will therefore be characterized by an optimal portfolio allocation determined by the optimal parameter. As explained above, each optimal allocation will then generate out-of-sample returns over the month following the window, and the sum of all the returns will be used to calculate the portfolio's overall utility.

Once the two utilities (analytical and machine learning) are found for each of the 4 portfolios, an objective comparison between the two techniques can be carried out.

For the sake of uniformity, the various utilities can be identified by the different acronyms that make up their names.

The first acronym stands for the calculated portfolio, with LW for Ledoit & Wolf, KZ for Kan & Zhou, TZ for Tu & Zhou and KWZ for Kan, Wang & Zhou.

The second acronym stands for the dataset used, with IN10 for 10 industry, IN48 for 48 industry, BM25 for 25 formed on size and Book-to-Market, OP25 for 25 formed on Operating Profitability and Investments, and MO10 for 10 momentum.

The third acronym stands for the calibration technique, with A for analytics and ML for machine learning.

Finally, for the machine learning technique, the fourth acronym represents the number of folds in the cross-validation, with 3 for 3-folds, 10 for 10-folds and K for K-fold.

Here are the results:

3.1. Ledoit-Wolf

Table of Utilities

Utility_LW_BM25_A	-0.04384721156495 64
Utility_LW_BM25_ML_10	0.00310368855780088
Utility_LW_BM25_ML_3	-0.00340897542944535
Utility_LW_BM25_ML_K	-0.0956590821823659
Utility_LW_IN10_A	-0.0248530739048456
Utility_LW_IN10_ML_10	-0.016277621464045 7
Utility_LW_IN10_ML_3	-0.0258792131051373
Utility_LW_IN10_ML_K	-0.049502327087306
Utility_LW_IN48_A	-0.208849104575171
Utility_LW_IN48_ML_10	-0.0668797377146921
Utility_LW_IN48_ML_3	-0.102879157921109
Utility_LW_IN48_ML_K	-0.248858830441134
Utility_LW_MO10_A	-0.011805487257766
Utility_LW_MO10_ML_10	0.00215509692758 91
Utility_LW_MO10_ML_3	-0.0076181926183029
Utility_LW_MO10_ML_K	-0.0374404568239489
Utility_LW_OP25_A	-0.0454062656217163
Utility_LW_OP25_ML_10	-0.0869551910108252
Utility_LW_OP25_ML_3	-0.0117309961033828
Utility_LW_OP25_ML_K	-0.115442257420239

Figure 4: Ledoit & Wolf shrinkage utility, via analytical and machine learning calibration

Source: RStudio

3.2. Kan-Zhou

Table of Utilities

Utility_KZ_BM25_A	0.010401615315987 7
Utility_KZ_BM25_ML_10	-0.0950073243751094
Utility_KZ_BM25_ML_3	-0.0921236607424604
Utility_KZ_BM25_ML_K	-0.129357995542018
Utility_KZ_IN10_A	-0.001164936928756
Utility_KZ_IN10_ML_10	0.003628043128169 13
Utility_KZ_IN10_ML_3	0.003609485713404 46
Utility_KZ_IN10_ML_K	0.003749351283893 47
Utility_KZ_IN48_A	-0.0087195460042247
Utility_KZ_IN48_ML_10	-1.93894801368622e-05
Utility_KZ_IN48_ML_3	-0.000191543418190855
Utility_KZ_IN48_ML_K	-0.000166302112430851
Utility_KZ_MO10_A	0.00688119566827 79
Utility_KZ_MO10_ML_10	-0.0529672429671083
Utility_KZ_MO10_ML_3	-0.0445219921727405
Utility_KZ_MO10_ML_K	-0.0587100196348771
Utility_KZ_OP25_A	0.003303291385034 15
Utility_KZ_OP25_ML_10	-0.0038327679502164 6
Utility_KZ_OP25_ML_3	-0.0023163125131667 2
Utility_KZ_OP25_ML_K	-0.0035227883823574 9

Figure 5: Kan & Zhou shrinkage utility, via analytical and machine learning calibration

Source: RStudio

3.3. Tu-Zhou

Table of Utilities

Utility_TZ_BM25_A	0.009950664323813 04
Utility_TZ_BM25_ML_10	-0.071531327838755 3
Utility_TZ_BM25_ML_3	-0.040223233335743 42
Utility_TZ_BM25_ML_K	-0.132085172817612
Utility_TZ_IN10_A	-0.001912458359674 75
Utility_TZ_IN10_ML_10	0.002285741866502 98
Utility_TZ_IN10_ML_3	0.002185561258635 37
Utility_TZ_IN10_ML_K	0.002349123263474 44
Utility_TZ_IN48_A	-0.007636623951042 23
Utility_TZ_IN48_ML_10	-0.000121387427257847
Utility_TZ_IN48_ML_3	-6.55879745411393e-05
Utility_TZ_IN48_ML_K	-0.000275400936172901
Utility_TZ_MO10_A	0.00566009579365 91
Utility_TZ_MO10_ML_10	-0.044398716996312 8
Utility_TZ_MO10_ML_3	-0.030478492629043 2
Utility_TZ_MO10_ML_K	-0.051087139089642 3
Utility_TZ_OP25_A	0.002768886688946 41
Utility_TZ_OP25_ML_10	-0.002430621200921 6
Utility_TZ_OP25_ML_3	-0.000681353706850198
Utility_TZ_OP25_ML_K	-0.002120640981386 02

Figure 6: Tu & Zhou shrinkage utility, via analytical and machine learning calibration

Source: RStudio.

3.4. Kan-Wang-Zhou

Table of Utilities

Utility_KWZ_BM25_A	0.0237958987571319
Utility_KWZ_BM25_ML_10	-0.0468413392883434
Utility_KWZ_BM25_ML_3	-0.0293290654097778
Utility_KWZ_BM25_ML_K	-0.0728469495406154
Utility_KWZ_IN10_A	0.00475183505244875
Utility_KWZ_IN10_ML_10	0.00315820872886092
Utility_KWZ_IN10_ML_3	0.00327870272515716
Utility_KWZ_IN10_ML_K	0.00321102358412386
Utility_KWZ_IN48_A	0.0157872843310147
Utility_KWZ_IN48_ML_10	0.000139267481705249
Utility_KWZ_IN48_ML_3	0.000222091947751086
Utility_KWZ_IN48_ML_K	-2.84623653030846e-05
Utility_KWZ_MO10_A	0.0119865246209332
Utility_KWZ_MO10_ML_10	-0.0461824526764581
Utility_KWZ_MO10_ML_3	-0.0330416980378054
Utility_KWZ_MO10_ML_K	-0.054238711895687
Utility_KWZ_OP25_A	0.0167808105647531
Utility_KWZ_OP25_ML_10	-0.00290900168851699
Utility_KWZ_OP25_ML_3	-0.00127209665795594
Utility_KWZ_OP25_ML_K	-0.00288079935640878

Figure 7: Kan, Wang & Zhou shrinkage utility, via analytical and machine learning calibration

Source: RStudio

4. Analysis

Once all the optimization methods have been applied, the various figures (R outputs) show whether the machine learning method performs better than the analytical method. The result is that for all three machine learning methods (3, 10, and K-folds cross-validation), the utility of the 10 Industry portfolio is always better than the utility of the analytical method for Kan & Zhou and Tu & Zhou shrinkage. Such an analysis can also be performed for the 48 Industry portfolio, but only with 2 out of 3 methods (3 and K-folds for Kan & Zhou, 10 and K-folds for Tu & Zhou). Furthermore, we can also see that the utility of some portfolio estimated by 3 and 10 folds cross-validation methods (except for OP25 where it is just 3 folds that outperforms) is higher than the analytical one for Ledoit & Wolf's shrinkage.

Kan Wang & Zhou's shrinkage, on the other hand, always delivers a better utility via analytical optimization. The machine learning method is therefore no better.

Finally, we can also see that there doesn't seem to be any correlation nor proportionality (increasing or decreasing) between the utility and the number of folds of the cross-validation.

Conclusion & extensions

In this study, we had the opportunity to investigate the ability of machine learning to optimize the parameters of 4 different shrinkage methods intended to improve the returns of a financial portfolio by combining several classical methods. The aim of this research was to see whether the optimization of parameters via machine learning could ultimately generate greater utility than that generated by analytical optimization, under strong statistical assumptions, of the parameters.

To achieve this, three cross-validation methods (3-folds, 10-folds and K-folds) were used to vary the shrinkage and combination parameters, ranging from 0 to 1 with a step size of 0.01. Thus, for each of the 599 10-year samples (120 monthly observations) based on 5 different portfolios from which the risk-free rate has been removed, 100 different parameter values have been tested per method, and the parameter value generating the highest out-of-sample utility the following month have been retained, thus forming the optimal portfolio - and therefore weights - for the sample. In this way, the rolling windows technique has been used to calculate the optimal weights of the 599 different sampling windows, generating a total of 599 returns, which led to compute the portfolio's overall utility, serving as a comparison criterion. In the end, the research establishes 4 different utility values: 1 via analytical methods, 3 via cross-validation.

Thus, the initial research question "Analytical versus machine learning calibration: which one works best in portfolio optimization?" can be answered accordingly:

In some cases, machine learning optimization methods can outperform the analytical method. However, as shrinkage methods are complex and require very precise analytical resolution of the optimal parameter, machine learning, although not relying on strong assumptions unlike the analytical method, cannot always generate a better performance than the analytical calibration of the parameters. It would therefore be wrong to conclude that machine learning can outperform analytics in every scenario.

So, in the end, one might think that the authors of optimization methods had found the perfect equation for choosing the optimum parameter. However, the perfect equation is more myth than reality, and perhaps other equations will find their way into the literature, surpassing the estimation methods proposed to date. A case in point is the recent Kan Wang & Zhou

optimization, in which Lassance, Martin-Utrera & Simaan (2023)[13] propose a different way of calibrating the shrinkage parameter, which turns out to give better out-of-sample results.

To conclude, it is important to note that this thesis has its own limitations.
Being a first draft on the subject, this master's thesis only focused on machine learning calibration via cross-validation. A multitude of other machine learning techniques can be applied to the subject, such as Bootstrap, Neural Network, Random Forest, and many others. This dissertation therefore leaves room for further investigation.

Similarly, only four estimation strategies are studied, leaving room for further investigation, such as an application of machine learning techniques to Kirby & Ostdiek's (2010)[14] Reward-to-Risk Timing (RRT) and Volatility Timing (VT) portfolios, Martellini & Ziemann (2009)[15], McKinley & Pastor (2000)[16] and many others.

Finally, machine learning calibration in this thesis is based on the utility of out-of-sample performance, a rather rational choice given the importance and significance of the concept in out-of-sample performance analysis. Other criteria can certainly supplement overall utility, such as cumulative return, Sharpe ratio or Certainty Equivalent Returns (CER), although utility remains in my view the most important and appropriate criterion in this context.

---

[13] Lassance, N., Martin-Utrera, A., & Simaan, M. (2021). A Robust Approach to Optimal Portfolio Choice with Parameter Uncertainty. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3855546

14 Kirby, C., & Ostdiek, B. (2010). It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naive Diversification. Social Science Research Network. https://doi.org/10.2139/ssrn.1530022

15 Martellini, L., & Ziemann, V. (2009). Improved estimates of Higher-Order comoments and implications for portfolio selection. Review of Financial Studies, 23(4), 1467–1502. https://doi.org/10.1093/rfs/hhp099

16 MacKinlay, A. C., & Pastor, L. (2000). Asset Pricing models: Implications for expected returns and portfolio selection. Review of Financial Studies, 13(4), 883–916. https://doi.org/10.1093/rfs/13.4.883

Bibliography

Scientific papers

Best, M.J. and Grauer, R.R. (1991) On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. The Review of Financial Studies, 4, 315-342. https://doi.org/10.1093/rfs/4.2.315

Chopra, V., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. The Journal of Portfolio Management, 19(2), 6–11. https://doi.org/10.3905/jpm.1993.409440

DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/NPortfolio Strategy? Review of Financial Studies, 22(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Kan, R., & Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. Journal of Financial and Quantitative Analysis, 42(3), 621–656. https://doi.org/10.1017/s0022109000004129

Kan, R., Wang, X., & Zhou, G. (2022). Optimal Portfolio Choice with Estimation Risk: No Risk-Free Asset Case. Management Science, 68(3), 2047–2068. https://doi.org/10.1287/mnsc.2021.3989

Kirby, C., & Ostdiek, B. (2010). It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naive Diversification. Social Science Research Network. https://doi.org/10.2139/ssrn.1530022

Lassance, N. (2022). Reconciling mean-variance portfolio theory with non-Gaussian returns. European Journal of Operational Research, 297(2), 729–740. https://doi.org/10.1016/j.ejor.2021.06.016

Lassance, N., Martin-Utrera, A., & Simaan, M. (2021). A Robust Approach to Optimal Portfolio Choice with Parameter Uncertainty. Social Science Research Network. https://doi.org/10.2139/ssrn.3855546

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management, 30(4), 110–119. https://doi.org/10.3905/jpm.2004.110

Ledoit, O., & Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365–411. https://doi.org/10.1016/s0047-259x(03)00096-4

MacKinlay, A. C., & Pastor, L. (2000). Asset Pricing models: Implications for expected returns and portfolio selection. Review of Financial Studies, 13(4), 883–916. https://doi.org/10.1093/rfs/13.4.883

Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), 77–91. https://doi.org/10.2307/2975974

Martellini, L., & Ziemann, V. (2009). Improved estimates of Higher-Order comoments and implications for portfolio selection. Review of Financial Studies, 23(4), 1467–1502. https://doi.org/10.1093/rfs/hhp099

Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. Journal of Financial Economics, 99(1), 204–215. https://doi.org/10.1016/j.jfineco.2010.08.013

Websites

GeeksforGeeks. (2021). Cross Validation in R programming. GeeksforGeeks. https://www.geeksforgeeks.org/cross-validation-in-r-programming/

Kałędkowski, D. (2023, March 21). Apply any R function on rolling windows. https://cran.r-project.org/web/packages/runner/vignettes/apply_any_r_function.html

Dolphin, R. (2022, March 11). Efficient Frontier in Python | towards Data science. Medium. https://towardsdatascience.com/efficient-frontier-in-python-detailed-tutorial-84a304f03e79

RPUBS - Reproducible Research in Portfolio Selection. (n.d.). https://rpubs.com/simaan84/KWZ_port

Rolling-Window Analysis of Time-Series Models - MATLAB & Simulink. (n.d.). https://www.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html

Team, D. (2023, March 31). Cross-Validation ou Validation croisée : définition et importance. Formation Data Science | DataScientest.com. https://datascientest.com/cross-validation