TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Equity returns and sample size - impact on optimal porfolios

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

Mechanical Engineering - Management

by

Paul Maximilian Setinek
Registration Number 11711314

to the Instute of Management Science

at the TU Wien

Advisor: Ao.Univ.Prof. Mag. Dr. Dr. Thomas Dangl

Vienna, 8th June, 2022

Paul Maximilian Setinek   Thomas Dangl

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Stichprobengröße und Aktienmärkte — die Auswirkung auf optimale Portfolios

BACHELORARBEIT

zur Erlangung des akademischen Grades

Bachelor of Science

im Rahmen des Studiums

Wirtschaftsingenieurswesen - Maschinenbau

eingereicht von

Paul Maximilian Setinek
Matrikelnummer 11711314

am Institut für Managementwissenschaften

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Mag. Dr. Dr. Thomas Dangl

Wien, 8. Juni 2022

Paul Maximilian Setinek | Thomas Dangl

Erklärung zur Verfassung der Arbeit

Paul Maximilian Setinek

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 8. Juni 2022

Paul Maximilian Setinek

Danksagung

Bei der Erstellung dieser Bachelorarbeit habe ich sehr viel Unterstützung und Hilfe erhalten, für die ich mich an dieser Stelle gerne bedanken würde. Ein besonderer Dank gilt meinem Betreuer, Ao.Univ.Prof. Mag. Dr. Dr. Thomas Dangl. Das schnelle und aufschlussreiche Feedback hat diese Arbeit auf ein höheres Niveau gebracht. Weiters möchte ich mich bei meiner Familie für die bedingungslose Unterstützung während meines gesamten Bachelorstudiums bedanken.

Acknowledgements

Throughout the writing of this bachelor thesis, I have received a great deal of support and assistance. Therefore I would like to express my thanks at this point. First, I would like to express my gratitude to my supervisor, Ao.Univ.Prof. Mag. Dr. Dr. Thomas Dangl. The quick and insightful feedback brought this thesis to a higher level. Furthermore, I would like to thank my family for the unconditional support throughout my whole bachelor's degree.

Kurzfassung

In dieser Arbeit wird untersucht, wie sich die Größe der für die Parameterschätzung verwendeten Stichprobe auf die out-of-sample-Performance optimaler Portfolios auswirkt. Ich vergleiche den klassischen Ansatz der Mean-Variance Optimierung mit einem Resampling Ansatz, der versucht die Portfolioauswahl bei Parameterunsicherheit zu verbessern. Das hierzu gewählte Verfahren ist eine Monte-Carlo-Simulationsstudie, die auf historischen Renditedaten von 11 Sektoren des S&P500 von Oktober 1989 bis November 2021 basiert. Das Maß für die out-of-sample-Performance ist das Sicherheitsäquivalent der erwarteten Portfoliorendite, das aus der Nutzenfunktion eines Anlegers mit konstanter relativer Risikoaversion abgeleitet wird. Die Ergebnisse zeigen, dass bei kleinen Stichprobengrößen die mit dem Resampling Ansatz ausgewählten Portfolios ein signifikant höheres erwartetes Sicherheitsäquivalent der Portfoliorendite im Vergleich zu den mit der klassischen Mean-Variance Methode konstruierten Portfolios liefern. Mit zunehmender Stichprobengröße nimmt die Performance der klassischen Mean-Variance Methode jedoch zu, und schlussendlich, bei Stichproben mit ausreichender Größe, übertrifft sie die des Resampling-Ansatzes.

Abstract

This thesis investigates the impact of the sample size used for parameter estimation on the out-of-sample performance of optimal portfolios. I compare the classical Mean-Variance optimisation framework to a resampling approach that tries to improve portfolio choice when facing parameter uncertainty. The chosen procedure is a Monte Carlo simulation study based on a time series of historical return data of 11 sectors of the S&P500 from October 1989 until November 2021. The out-of-sample performance measure is the certainty equivalent of the expected portfolio return that is derived from an investor's utility function with constant relative risk aversion. The obtained results show that for small sample sizes, portfolios chosen with the resampling approach yield a significantly higher expected out-of-sample certainty equivalent of portfolio return compared to portfolios constructed using the classical Mean-Variance framework. However, as sample size increases, the performance of the classcial Mean-variance method increases and eventually it outperforms the resampling approach.

Contents

General Notation

x	A scalar
$\boldsymbol{x}$	A vector
$\boldsymbol{x'}$	The transpose of vector $\boldsymbol{x}$
$\boldsymbol{X}$	A matrix
$\boldsymbol{X}^{-1}$	The inverse of matrix $\boldsymbol{X}$
$\mathbf{1}$	A vector of ones, i.e. vector with all components being 1
$\mathbb{E}[x]$	The expected value of x

Abbreviations

ARA	Absolute risk aversion
CLT	Central limit theorem
CRRA	Constant relative risk aversion
FC	Forecast Confidence
GCP	Google Cloud Platform
GMV	Global minimum variance
MPT	Modern portfolio theory
MV	Mean-Variance
QP	Quadratic problem
REF	Resampled Efficiency Frontier
RRA	Relative risk aversion

CHAPTER 1

Introduction

1.1 History and state of the art

Modern portfolio theory (MPT) was first introduced by Harry M. Markowitz (1952, 1959). In his first publication about the topic, he rejects the hypothesis that an investor should simply maximise anticipated returns and argues that the variance of returns has to be taken into account, when building portfolios. Therefore this framework is often referred to as Mean-Variance (MV) framework, as it defines an efficient portfolio as one having minimum risk for an expected return, with the measures of return and risk being the mean and variance of the portfolio's return. The mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of returns, which are the two input parameters to this framework, need to be estimated precisely. Markowtiz himself already recognised this to be a crucial part of the portfolio choice and believed it should be achieved by a combination of statistical techniques and opinions of practical men (Markowitz, 1952, p. 91).

This issue of estimating these input parameters has been the main topic of research ever since Markowitz' first publication. Chopra and Ziemba (1993); Jobson and Korkie (1980); Frankfurter et al. (1971); Dickinson (1974) have all outlined the framework's high sensitivity to errors in the estimates of the inputs needed for optimisation. Therefore, many approaches have been developed in order to make the choice of optimal portfolios more robust and practically applicable. Approaches contain but are not limited to Bayesian-Stein shrinkage, Bayesian diffuse-prior, Ledoit-Wolf shrinkage (Ledoit and Wolf, 2004) or the framework introduced in Black and Litterman (1990) and later evolved in Black and Litterman (1991, 1992).

In this thesis, I want to lay the focus on a more heuristic approach introduced by Richard O. Michaud (1989). Instead of using a simple point estimate of the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ based on historical data, which I will refer to as the plug-in rule in this thesis, this alternative approach uses a resampling technique to simulate scenarios in order to obtain more robust parameter estimates. Michaud and Michaud (2007) provide a detailed explanation of the calculation of optimal portfolios using this resampling technique and also compare the out-of-sample performance to the simple plug-in rule. There are also further studies where Michaud's resampling framework is compared to the plug-in rule, such as e.g. Becker et al. (2009). However, the impact of sample size, risk aversion and Forecast Confidence (FC), a concept which will be explained later in this thesis, has not yet been discussed extensively in the literature. Therefore, this thesis tries to fill this gap and obtain insights in how these parameters influence the out-of-sample performance of the two strategies that are to be compared.

1.2 Research question and approach

In this thesis, I investigate the impact of sample size on the performance of two portfolio selection methods, namely the classical MV optimisation approach, also called plug-in rule, and Michaud's resampling approach. In order to compare these portfolio rules, I conduct a Monte Carlo simulation study where I simulate histories of returns, based on a multivariate normal distribution centred around the two true moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of 11 industries of the S&P500 from October 1989 until November 2021. For all these histories, optimal portfolios will be chosen based on the two portfolio rules. Afterwards, they are compared to each other via their out-of-sample performance. The out-of-sample performance measure is chosen to be the certainty equivalent of the portfolio return, r_{CE}. A certainty equivalent value represents the certain payout that makes the investor indifferent between receiving the certainty equivalent value or the uncertain outcome of the investment. By altering the length of simulated histories, i.e. the sample size T, this thesis analyses the effect of T on the resulting certainty equivalent.

Using this methodology, I find an answer to the research question which can be formulated as follows:

> *For what sample sizes, does the resampling rule lead to a statistically significant higher out-of-sample certainty equivalent of portfolio return than the plug-in rule?*

Due to the random nature of the simulated samples, the focus is laid on the expected values of r_{CE}. I therefore introduce a loss function ρ and then formulate the following null and alternative hypothesis:

$$H_0 : \mathbb{E}[\rho_{\text{resampled}}] - \mathbb{E}[\rho_{\text{plug-in}}] \geq 0$$
$$H_a : \mathbb{E}[\rho_{\text{resampled}}] - \mathbb{E}[\rho_{\text{plug-in}}] < 0$$

These two hypotheses will be used to compare the out-of-sample results of the above mentioned portfolio rules.

1.3 Outline

In the first chapter, I start by introducing the theoretical framework of MV optimisation. Beginning with the basic model and analytical solutions for given circumstances, I move on to more complicated settings with further restrictions imposed and solution methods for these problems. The second section focuses on utility functions. First, an introduction into the field of utility theory is given and then I derive how to include the concept of utility maximisation into the MV framework, because this utility maximisation based framework will be the basis of performance analysis within the simulation study.

The second chapter covers the application of MV optimisation in practice. Starting with in-sample optimisation for historic return data, I continue with the discussion of some problems arising in real world situations. Here, the focus is laid on the topic of estimation error and bad out-of-sample performance of the classical plug-in model.

In the third chapter, I introduce the resampling approach of Michaud and Michaud (2007) in order to make MV optimisation more robust for practical applications. The construction of the Resampled Efficiency Frontier (REF) is explained together with an example where the concept is applied to real historic return data.

The last chapter covers the actual simulation study of this thesis. I conduct a Monte Carlo simulation study to compare the classical plug-in rule to the resampling approach. Here, the focus is laid on the impact of sample size used to estimate the moments of the distribution on out-of-sample performance of the two portfolio rules. The results are then presented and discussed extensively and afterwards summarised in a short conclusion.

CHAPTER 2

Mean-Variance optimisation framework

This chapter provides an overview of the classical single period MPT, originally proposed by Harry M. Markowitz (1952) and later evolved by him in Markowitz (1959).

I start with a short but concise introduction to the optimisation framework. Then I provide a short digression into utility theory and show how to formulate an objective function in order to maximise an investors utility.

2.1 Classical Mean-Variance framework

The following sections are based on recent publications treating the MV framework, namely Brandt (2010), Chapados (2011) and Fabozzi et al. (2007).

Suppose, an investor wants to build a portfolio at time t for the upcoming time period $[t, t+1]$. Let us define $\boldsymbol{w} \in \mathbb{R}^N$ to be the weight vector of the portfolio, where element $\boldsymbol{w}_i$ represents the weight of asset i, i.e. the fraction of total capital invested in this asset. Let $\mathbf{r}_{t+1} \in \mathbb{R}^N$ be the vector of N random asset returns[1] between times t and $t+1$. Suppose, the mean vector $\boldsymbol{\mu} \in \mathbb{R}^N$ of the returns and the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ of the returns are given by:

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{r}_{t+1}] \tag{2.1}$$

$$\boldsymbol{\Sigma} = \mathrm{Cov}[\mathbf{r}_{t+1}] \tag{2.2}$$

[1] Returns are measured as Holding Period Returns (HPRs), also referred to as simple returns.

In general, the true moments of future returns are unknown and have to be estimated. This topic and problems arising from this will be treated in Section 3.2.

The portfolio's expected return and variance are calculated as follows:

$$\mu_P = \boldsymbol{w}'\boldsymbol{\mu} \tag{2.3}$$

$$\sigma_P^2 = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \tag{2.4}$$

We can now formulate the optimisation problem with the goal of minimising the portfolio's variance for a given desired return, μ_{des}. The optimisation problem then looks as follows:

$$\begin{aligned} \min_{w} \quad & \frac{1}{2}\sigma_P^2 = \frac{1}{2} \quad \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\ \text{subject to:} \quad & \boldsymbol{w}'\boldsymbol{\mu} = \mu_{\text{des}} \\ & \boldsymbol{w}'\mathbf{1} = 1 \end{aligned} \tag{2.5}$$

This problem can be solved analytically with the introduction of Lagrange multipliers, since both constraints are of equality type. For the derivation and simplification of the solution see for example Fabozzi et al. (2007, pp. 24–26), Focardi and Fabozzi (2004, pp. 204-206), Chapados (2011, pp. 9–10) or Merton (1972, pp. 1851–1857). One can find the optimal weight vector, $\boldsymbol{w}$, as being a linear function of the desired return μ_{des}:

$$\boldsymbol{w} = \mathbf{g} + \mathbf{h}\mu_{\text{des}} \tag{2.6}$$

where $\mathbf{g}$ and $\mathbf{h} \in \mathbb{R}^N$ are:

$$\mathbf{g} = \frac{1}{ac - b^2}\boldsymbol{\Sigma}^{-1}(c\mathbf{1} - b\boldsymbol{\mu}) \tag{2.7}$$

$$\mathbf{h} = \frac{1}{ac - b^2}\boldsymbol{\Sigma}^{-1}(a\boldsymbol{\mu} - b\mathbf{1}) \tag{2.8}$$

and:

$$a = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} \tag{2.9}$$

$$b = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \tag{2.10}$$

$$c = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \tag{2.11}$$

When looking at the optimisation problem without the return constraint, the problem can be solved analogically. The resulting portfolio is called the the global minimum variance (GMV) portfolio. Its weights and variance respectively, can be expressed as follows:

$$\boldsymbol{w}_{\text{GMV}} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} \tag{2.12}$$

$$\sigma_{\text{GMV}}^2 = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} \tag{2.13}$$

2.2 Long-only constraint

In order to make the framework more applicable to real world situations, further constraints can be added to the problem formulation. An important restriction is the so called *"long-only"* constraint, which means, that no short selling of assets is allowed, i.e. all portfolio weights must be non-negative. This constraint is imposed often, since short selling is associated with high costs for the individual investor, which is the main reason why many long-term investors avoid short positions. Furthermore, a large fraction of investments in firm equity is done via mutual funds. Regulation frequently prohibits mutual funds from taking short positions (hedge funds excepted). The optimisation problem can then be defined as follows:

$$\begin{aligned} \min_{w} \quad & \frac{1}{2}\sigma_P^2 = \frac{1}{2} \quad \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\ \text{subject to:} \quad & \boldsymbol{w}'\boldsymbol{\mu} = \mu_{\text{des}} \\ & \boldsymbol{w}'\mathbf{1} = 1 \\ & \boldsymbol{w}_i \geq \mathbf{0} \end{aligned} \tag{2.14}$$

This problem can in theory be solved analytically as it is is a classical Kuhn-Tucker formulation. An approach to an iterative Kuhn-Tucker method can be seen in Jagannathan and Ma (2003). The other approach is to solve this problem numerically using quadratic problem (QP) solvers, as we have to deal with a quadratic objective function with linear constraints.

2.3 Utility maximisation

The goal of this section is the positioning of the MV portfolio optimisation framework within the general theory of maximising expected utility. It is not immediately clear how an portfolio objective that is formulated in terms of mean and variance of returns fits into a theory that understands consumption as the source of utility. In this section, I will present how the objective function of MV portfolio optimisation can be derived from utility theory.

2.3.1 An introduction to utility functions

In economic theory, utility is the *"happiness"* or *"satisfaction"* gained from consuming a good or a service. We can define a utility function $U(x)$ representing the utility gained from consuming the amount of x. In the context of the one period portfolio optimisation setting that I treat in this thesis, we can think of x as the entirely consumed wealth at the end of period $[t, t+1]$.

Ingersoll (1987) defines a utility function as a twice-differentiable function of wealth $U(x)$ defined for $x > 0$, which fulfils the following two properties:

Non-satiation: $U'(x) > 0$ implies that the investor is never satisfied, i.e. more consumption is always preferred to less consumption.

Risk aversion: $U''(x) < 0$ implies the decreasing marginal utility (cause of a utility function's concavity).

In order to characterize utility functions, I introduce the Arrow–Pratt measure of absolute risk aversion (ARA) (Arrow, 1965; Pratt, 1964), which is defined as follows:

$$ARA(x) = -\frac{U''(x)}{U'(x)} \tag{2.15}$$

And the Arrow–Pratt measure of relative risk aversion (RRA):

$$RRA(x) = -x\frac{U''(x)}{U'(x)} = \gamma \tag{2.16}$$

which I will refer to as the coefficient of relative risk aversion, γ, for the rest of this thesis.

Utility functions can take on different forms. However, for this thesis it will be assumed, that the investor's utility follows a power utility function. An important peculiarity of this kind of function is that it has a constant relative risk aversion (CRRA). This means that as an investor's wealth increases, he will still invest the same percentage of his wealth in risky assets. I define the utility function as follows:

$$U(x) = \begin{cases} \frac{x^{1-\gamma}-1}{1-\gamma} & \gamma \neq 1 \\ \ln(x) & \gamma = 1 \end{cases} \tag{2.17}$$

In financial applications it is common to follow the principle of utility maximization. As Norstad (1999, p. 2) puts it: *"...a rational investor, when faced with a choice among a set of competing feasible investment alternatives, acts to select an investment which maximizes his expected utility of wealth."* Therefore we want to maximise $\mathbb{E}[U(x)]$ and not $\mathbb{E}[x]$ in our context. Note, that with the utility function's concavity, Jensen's inequality states the following:

$$\mathbb{E}[U(x)] \leq U(\mathbb{E}[x]) \tag{2.18}$$

An essential concept to introduce when working with utility functions is the certainty equivalent. Norstad (1999, p. 5) and Ashwin and Tikhon (2021, p. 204) define the certainty equivalent of an investment whose outcome is given by a random variable x as:

$$x_{CE} = U^{-1}(\mathbb{E}[U(x)] \tag{2.19}$$

$$U(x_{CE}) = \mathbb{E}[U(x)] \tag{2.20}$$

In order to be able to interpret the certainty equivalent, I provide two interpretations from introductory literature to utility theory. Ashwin and Tikhon (2021) interpret the certainty equivalent as the certain amount an investor would pay to consume an uncertain outcome.

Clemen (1996, p. 469) describes the certainty equivalent as the amount of money that is equivalent in an investors mind to a given situation that involves uncertainty. He also provides a short example which I want to quote at this point:

Suppose you face the following gamble:

- *Win 2000 with probability 0.50*
- *Lose 20 with probability 0.50*

Now imagine that one of your friends is interested in taking your place. "Sure," you reply, "I'll sell it to you." After thought and discussion, you conclude that the least you would sell your position for is 300. If your friend cannot pay that much, then you would rather keep the gamble. (Of course, if your friend were to offer more, you would take it!) Your certainty equivalent for the gamble is 300. This is a sure thing; no risk is involved. From this, the meaning of certainty equivalent becomes clear. If 300 is the least that you would accept for the gamble, then the gamble must be equivalent in your mind to a sure 300.

Another important concept to introduce at this point is the absolute risk premium, π_A. Ashwin and Tikhon (2021) define it as follows:

$$\pi_A = \mathbb{E}[x] - x_{CE} \tag{2.21}$$

Figure 2.1 shows Concepts (2.18), (2.19), (2.20) and (2.21) for a CRRA utility function (see Equation (2.17)) with a γ value of 3.

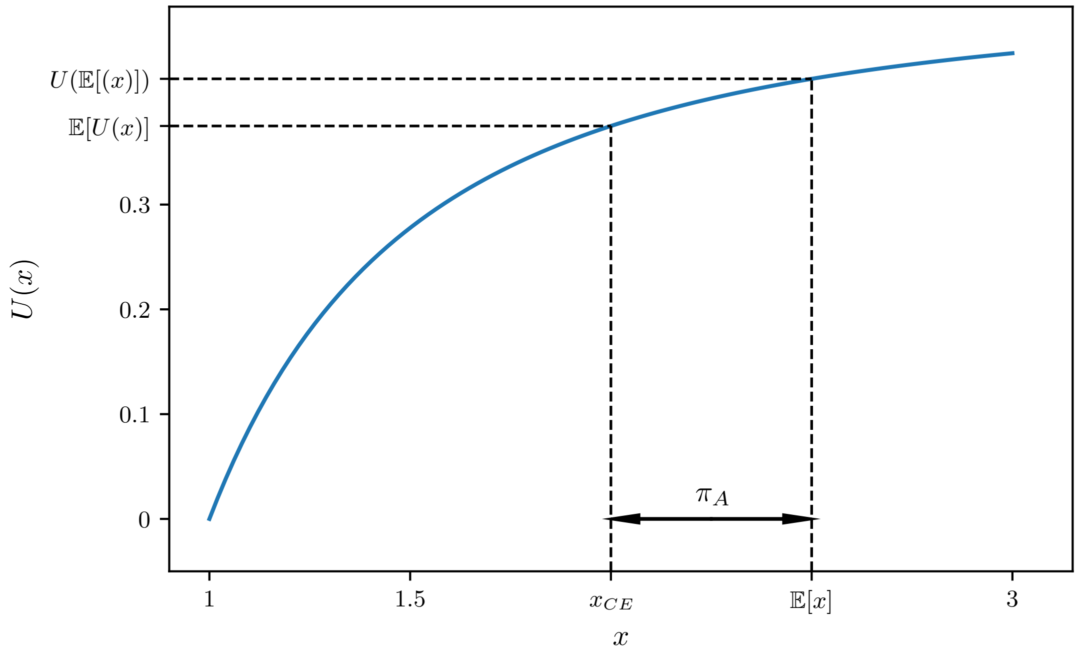

Figure 2.1: CRRA utility function $U(x)$ from consumption of end-of-period-wealth x with $\gamma = 3$. Jensen's inequality, the certainty equivalent x_{CE} and the absolute risk premium π_A are shown.

Arrow (1963) and Pratt (1964) independently developed an approximation for the absolute risk premium for a normally distributed random payoff x with small variance. They used a Taylor series approximation and Equation (2.15) to derive the following:[2]

$$\pi_A \cong \frac{1}{2}\sigma^2 ARA(x) \tag{2.22}$$

We have already established that we want to maximise $\mathbb{E}[U(x)]$ and not $\mathbb{E}[x]$ in our context. With Equation (2.20), it is obvious that, with the concave attributes of $U(x)$, one needs to maximise x_{CE} in order to maximise expected utility. With (2.21) and (2.22), we can write an approximation for the certainty equivalent as follows:

$$x_{CE} \cong \mathbb{E}[x] - \frac{1}{2}\sigma^2 ARA(x) \tag{2.23}$$

So far, I defined utility in terms of wealth x which is entirely consumed at the end of the considered time period. In the context of portfolio optimisation, we therefore need to rewrite the problem in terms of return r:

[2] For a detailed derivation, see e.g. Eeckhoudt et al. (2011, pp. 9–12) and Ashwin and Tikhon (2021, pp. 204–205).

$$x = x_0(1 + r) \tag{2.24}$$

$$\mathbb{E}[x] = x_0(1 + \mathbb{E}[r]) = x_0(1 + \mu) \tag{2.25}$$

$$Var(x) = \sigma_x^2 = x_0^2\sigma_r^2 \tag{2.26}$$

When plugging the equations above in (2.23), we get:

$$x_0(1 + r_{CE}) \cong x_0(1 + \mathbb{E}[r]) - \frac{1}{2}ARA(x)x_0^2\sigma_r^2 \tag{2.27}$$

and after some simplification:

$$r_{CE} \cong \mathbb{E}[r] - \frac{1}{2}\underbrace{ARA(x)x_0}_{\cong ARA(x)x = RRA(x) = \gamma}\sigma_r^2 \tag{2.28}$$

and therefore finally our function to optimise:

$$r_{CE} \cong \mu - \frac{1}{2}\gamma\sigma_r^2 \tag{2.29}$$

Rewriting this in terms of portfolio return, yields:

$$r_{CE} \cong \mu_p - \frac{1}{2}\gamma\sigma_p^2 = \boldsymbol{w}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \tag{2.30}$$

Note that this is only an approximation of the certainty equivalent for a power utility function (CRRA function) and returns with a small variance. However there are many studies that show for small returns it is a very good approximation. Levy and Markowtiz (1979) concluded, that MV formulations approximate power utility functions well for returns from -30% to 60%.

2.3.2 Applied utility maximisation in MPT

With the objective function given in (2.30) and the long-only constraint introduced Chapter 2.2, we can formulate a slightly different version of Problem (2.14) and write the following:

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & r_{CE} \cong \boldsymbol{w}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\ \text{subject to:} \quad & \boldsymbol{w}'\mathbf{1} = 1 \\ & \boldsymbol{w}_i \geq \mathbf{0} \end{aligned} \tag{2.31}$$

I now derived two versions of the optimisation problem. The original Markowitz formulation (2.14) as a function of the target expected return on one hand, and Problem (2.31) as a function of the relative risk aversion coefficient γ on the other. Chapados (2011, pp. 12–13) shows very neatly, that both problem definitions produce the same efficiency frontier when solved.

Without the long-only constraint, Problem (2.31) can again be solved analogically to Problem (2.5) by introducing Lagrange multipliers.

Again, when taking into account the long-only constraint, the problem is not as trivial to solve as the one ignoring the long-only constraint. For this thesis, I solve the optimisation problem using the CVXPY package for python, provided and documented by Diamond and Boyd (2016) and Agrawal et al. (2018). Within this framework I use OQSP, an open-source C library for solving convex quadratic problems, documented by Stellato et al. (2020). The actual implementation of calculating MV efficient portfolios can be found in the function efficiency_frontier() in Listing B.1 in Appendix B.

CHAPTER 3

Mean-Variance optimisation in practice

This chapter will discuss the application of the MV framework in practice. I start with the in-sample optimisation with an example using real historic return data. Then I introduce the debate of out-of-sample performance of MV optimal portfolios, which is heavily discussed in financial literature.

3.1 In sample Mean-Variance application

For this thesis, I use weekly return data of companies in the S&P500 grouped by sector, starting from October 1987 until November 2021.[1] At this point, I want to introduce the following notation: I will refer to the entire time series return data as $\mathbf{\Phi}_{FS}$ (FS standing for Full Sample), with $\mathbf{\Phi}_{FS} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{1673}\}$, where $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{1673} \in \mathbb{R}^N$ are the return vectors each week.

Figure 3.1 shows the return data's cumulative evolution and Table 3.1 summarises the data by showing the mean and standard deviation of weekly returns per sector. The full description of the data set with mean vector $\boldsymbol{\mu}_{FS}$ and covariance matrix $\mathbf{\Sigma}_{FS}$ of weekly returns can be found in Appendix A.

[1] The data is provided by Refinitiv Datastream (previously Thomson Reuters DataStream)

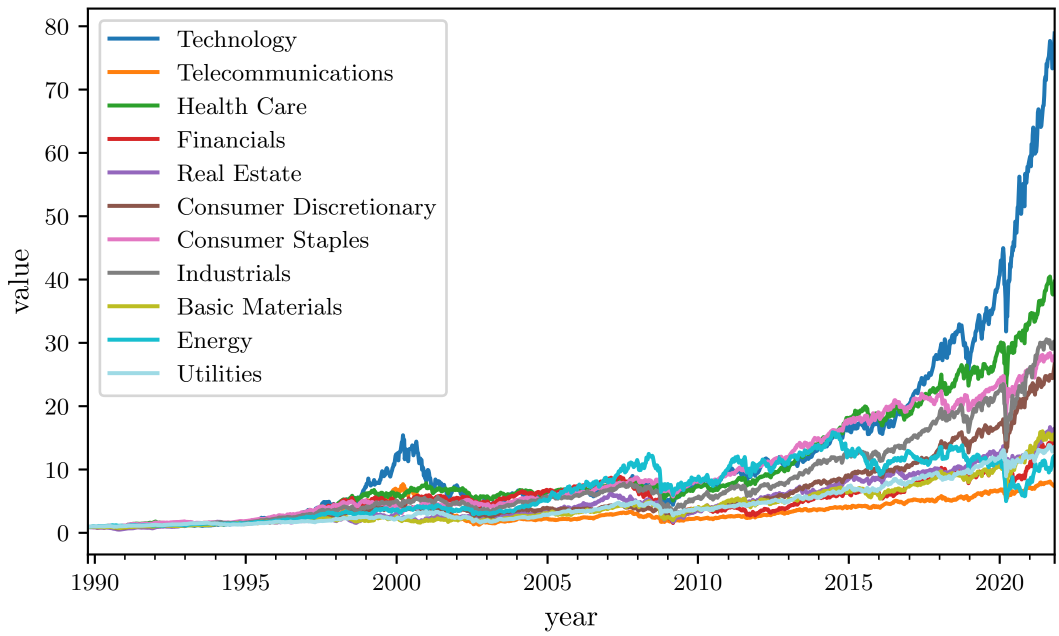

Figure 3.1: Cumulative evolution of the 11 sectors in the S&P500 from Oct 1989 to Nov 2021

	mean	sd
Technology	0.003174	0.033339
Telecommunications	0.001578	0.027619
Health Care	0.002499	0.024267
Financials	0.002250	0.035750
Real Estate	0.002307	0.035471
Consumer Discretionary	0.002305	0.026059
Consumer Staples	0.002201	0.020475
Industrials	0.002400	0.027256
Basic Materials	0.002092	0.029906
Energy	0.002023	0.032260
Utilities	0.001826	0.023046

Table 3.1: Summary statistics of the weekly returns of the 11 sectors in the S&P500 from Oct 1989 to Nov 2021

When solving Problem (2.31) for different values of relative risk aversion, γ, the efficient portfolio frontiers shown in Figure 3.2 can be obtained for the data shown above. This figure shows the expected weekly return over the weekly standard deviation for various portfolios and single assets. The red curve corresponds to the MV efficient portfolios calculated with the long-only constraint and the green curve represents the efficient portfolios, when short selling is allowed. I highlighted the optimal portfolios for γ values of 0.5, 1, 3 and 5 on the long-only frontier as these are also used in the simulation study in Chapter 5. For aestethic reasons, the green curve stops at $\gamma = 3$ as the standard deviation and expected return of portfolios with lower γ become very large when short selling is allowed. Besides the two curves representing the portfolios on the efficient frontiers, the single sectors within the S&P500 are marked with black crosses in order to show the potential benefit of portfolio diversification and optimisation.

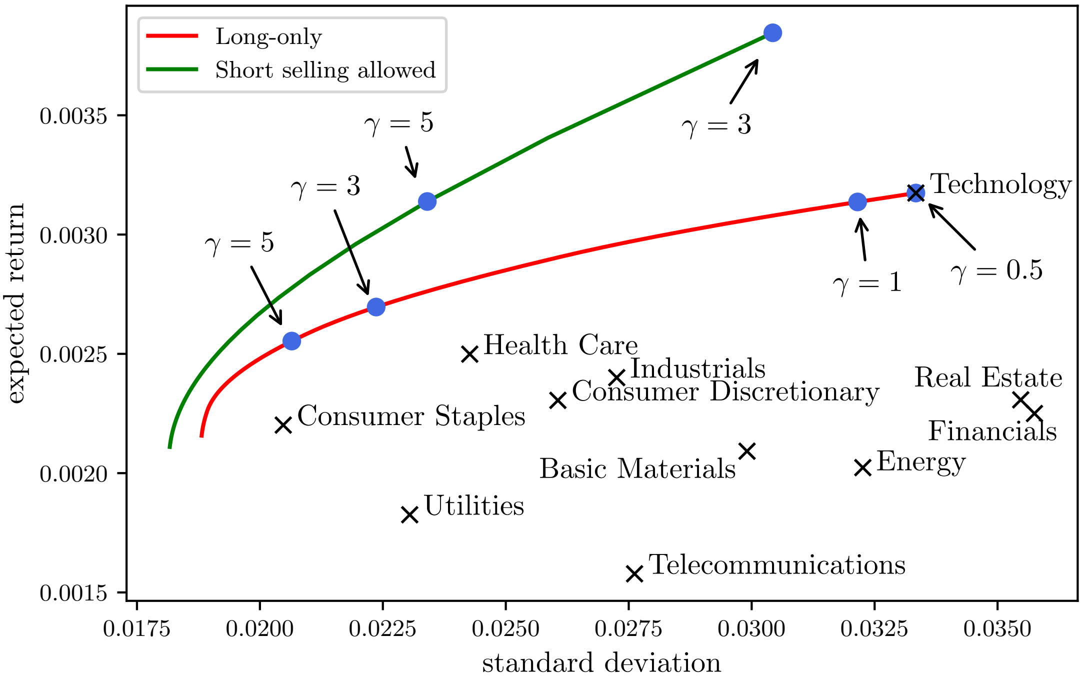

Figure 3.2: MV efficient frontier with and without short selling

3.2 Problem of classical Mean-Variance optimisation: estimation error

In practice, portfolio optimisation is not as straightforward as in the previous section, because at time t the true parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) of the distribution of future asset returns for period $[t, t+1]$ are unknown. Therefore the optimal weight vector, $\boldsymbol{w}^*$, is not computable at time t.

The classical (often also called traditional) approach is to estimate the distribution's moments based on observed historical return data and treating them as if they were the true values. These estimates are then plugged into Problem (2.31) in order to obtain optimal portfolio weights for period $[t, t+1]$. The corresponding portfolio is called the "plug-in portfolio" (Kan and Zhou, 2007) and I will refer to the whole procedure as the "plug-in rule" for the remainder of this thesis. The problem with this portfolio rule is, that it is shown to have very bad out-of-sample performance (Frankfurter et al., 1971; Chopra et al., 1993) and and is even often outperformed by the "equal-weights" portfolio rule[2] (Demiguel et al., 2009).

In his original paper, Markowitz (1952) already recognised this issue and separates between two stages of portfolio optimisation. The first one being about observing securities and gaining beliefs about their future performances. The second one is about the choice of an optimal portfolio given the inputs from the first stage. Concerning the first stage, he states: *"To use the E-V rule in the selection of securities we must have procedures for finding reasonable μ_i and σ_{ij}. These procedures, I believe, should combine statistical techniques and the judgment of practical men."* (Markowitz, 1952, p. 91)

Many studies have shown, that the MV optimisation framework is very sensible to these two input parameters and estimation error in them can have costly effects on the out-of-sample portfolio return (Chopra and Ziemba, 1993; Dickinson, 1974; Frankfurter et al., 1971; Jobson and Korkie, 1980). Usually, estimates of the mean are much noisier than estimates of the variance. Chopra and Ziemba (1993) therefore conclude that estimation errors in the mean of returns have a more fatal impact than errors in the covariance matrix. Kan and Zhou (2007) showed, using their analytically derived formula, that this is true as long as time series are long and the cross-section of assets is small. However for a large ratio of number of assets N to periods of historical return data T, the estimation error in the covariance matrix can become significant and very costly. Another analytical argument for large estimation errors in the covariance matrix in settings where $T < N$ can be found in Dangl and Kashofer (2013). In this thesis, in order to reduce estimation errors in the covariance matrix and to save computational power, I grouped the assets in the S&P500 into eleven sectors.

Considering the debate about parameter estimation error, one might question the applicability of the MV framework in practice. Michaud (1989) or Best and Grauer (1991) discuss the problems of practically implementing MV optimal portfolios. Therefore, the focus of a lot of research in portfolio theory has been laid on finding methods to better handle estimation error. Multiple approaches using the Bayesian approach have been implemented, beginning with statistical approaches relying on diffuse-priors (Barry, 1974; Bawa et al., 1979) onto shrinkage estimators (Jobson and Korkie, 1980; Jorion, 1986) to more recent approaches that rely on an asset pricing model for establishing a prior (Pástor, 2000; Ľuboš Pástor and Stambaugh, 2000), or the model introduced by Black and Litterman (1990) and further developed in Black and Litterman (1991, 1992) that combines two sets of priors - one based on an equilibrium asset-pricing model and the other on the investor's subjective views. Next to those methods exist the non-Bayesian

[2]As the name already suggests, for this portfolio rule all possible assets are weighted equally in the portfolio.

approaches that contain factor-based moment restrictions (MacKinlay and Ľuboš Pástor, 2000), methods that try to reduce estimation error in the covariance matrix (Best and Grauer, 1991; Ledoit and Wolf, 2004) or portfolio rules that impose shortselling constraints (Frost and Savarino, 1986; Chopra and Ziemba, 1993; Jagannathan and Ma, 2003).

Another non-Bayesian but rather heuristic approach in order to obtain more robust parameter estimates was introduced by Michaud (1989) and further discussed in Michaud and Michaud (2007). This portfolio rule uses a resampling approach that will be discussed in detail in the upcoming chapter.

CHAPTER 4

Resampled efficiency framework

Many complicated methods in order to handle the issue of estimation error have been developed in the past. Kan and Zhou (2007) for example, assume joint normally distributed returns and analyze the small sample properties of the distributions of estimates of mean and covariance, from which they derive analytical formulas for the estimation error.

In this section, I will concentrate on a more intuitive approach in order to reflect the uncertainty in the mean and covariance estimators for the MV optimization, first introduced by Richard O. Michaud (1989). His method involves a resampling technique that uses Monte-Carlo simulation studies in order to create many possible scenarios and averaging them in order to receive more robust estimates for the parameters needed for MV optimisation.

4.1 Calculatig the Resampld Efficiency Frontier

This section is based on a number of papers about the concept of the Resampled Efficiency Frontier (REF) by its inventor (Michaud, 1989; Michaud and Michaud, 2004, 2007, 2008).

In their paper, Michaud and Michaud (2007, p. 5) define the exact procedure to create a REF as follows:

Step 1 Sample a mean vector and covariance matrix of returns from a distribution of both centered at the original (point estimate) values normally used in MV optimization.

Step 2 Calculate a MV efficient frontier based on these sampled risk and return estimates.

Step 3 Repeat Steps 1 and 2 K times (until enough observations are available for convergence in Step 4).

Step 4 Average the portfolio weights from Step 2 to form the RE optimal portfolio, $\boldsymbol{w}^{\text{resampled}}$.

Step 5 (optional) Apply investability constraints to 4.

The first step can be realised with resampling or bootstrapping. By resampling, I mean simulating a scenario, i.e. simulating return time series of length R. The i-th sample of returns within the resampling approach is denoted by $\boldsymbol{\Phi}_i = \{\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_R\}$, with $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_R \in \mathbb{R}^N$ being the weekly return vectors of N assets. The resampling is carried out using a multivariate normal distribution centred at the original historic point estimates of the two moments $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$. For the remainder of this chapter, the two moments of the Full Sample $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$ serve as the historic point estimates.

Another possible approach to generating the scenario is by bootstrapping the original historic data and calculating the mean and covariance for this bootstrapped data set. For this thesis the first approach is used, however, the bootstrapping approach is a very interesting area for further research.

For clarification purposes, Figure 4.1 provides a detailed graphic summary of the above described resampling portfolio rule.

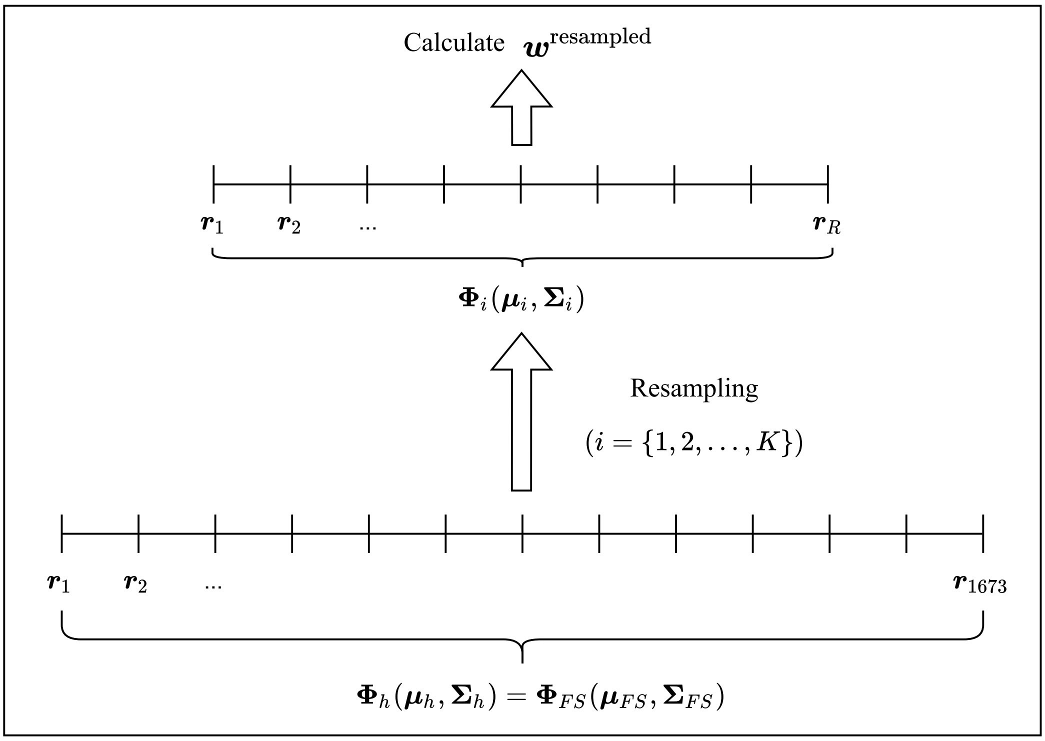

Figure 4.1: Graphical overview of the resampling portfolio rule

Since R is a free parameter in this model, it is important to introduce the conept of Forecast Confidence (FC). Logically, the more periods are simulated, the more we experience a convergence of the sample moments to the moments of the underlying distribution. This can be explained with the law of "large numbers" (DeGroot and Schervish, 2012, pp. 233–235). Therefore, with increasing sample size, the REF becomes more and more similar to the classical frontier computed using the plug-in rule. Michaud and Michaud (2004) provide a detailed overview of the concept of FC. For the remainder of this thesis, I now introduce three exemplary confidence levels that will be used in this chapter aswell as in the simulation study. Beginning with low FC which corresponds to a sample size of $R = 350$ weeks onto medium FC representing a sample size of $R = 1000$ weeks and finally high FC corresponding to a sample size of $R = 3000$ weeks. Since the interpretation of these numerical values is not straightforward, the next section focuses on an example where the REF for these exemplary FC levels is constructed using real return data. There, Figure 4.4 shows the efficient portfolio frontiers for the three exemplary FC levels next to the original efficient frontier.

Concerning Step 4, one has to take a closer look on how to average the portfolio weights. As not every simulated efficiency frontier has an equal range of return or risk (see Figure 4.2), Michaud and Michaud (2007) suggest to average the portfolio weights is by averaging all the portfolios for each risk aversion parameter γ. Mathematically speaking, Michaud and Michaud (2007, p. 7) describe the REF as *"...an integral in portfolio space, over all possible return distributions consistent with the forecast, of the expected value of the MV optimal portfolio weights conditional on the constraints"*, with the resampling/bootstrapping method being a Monte Carlo method for estimating this integral.

4.2 Application

In this section, an example of the REF's construction process, as explained in the section above, will be shown with the help of detailing graphs. The return data used is again the Full Sample, $\mathbf{\Phi}_{FS}$.[1] The corresponding python implementation of calculating the REF can be found in the function re_efficient_frontier() in Listing B.1 in Appendix B.

For Step 1, I will use a multivariate normal distribution to sample a vector of weekly returns for the N sectors over R weeks. Then the MV efficient frontier for this scenario is calculated (Step 2). The first two steps are then repeated K times.[2]

Figure 4.2 shows the resulting efficient frontiers for each scenario (with a sample size of $R = 1000$ weeks which corresponds to medium FC) together with the original efficient frontier that is obtained with estimates $\boldsymbol{\mu}_{FS}$ and $\mathbf{\Sigma}_{FS}$ from the entire history of return data. It is astonishing but also predictable, that the risks as well as the expected returns

[1] For the summary statistics, see Appendix A

[2] I chose $K = 100$ as this has shown to be robust in order to reach convergence in Step 4.

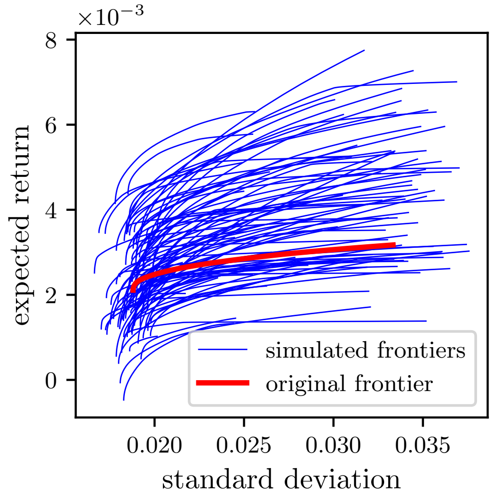

Figure 4.2: Simulated frontiers and original frontier

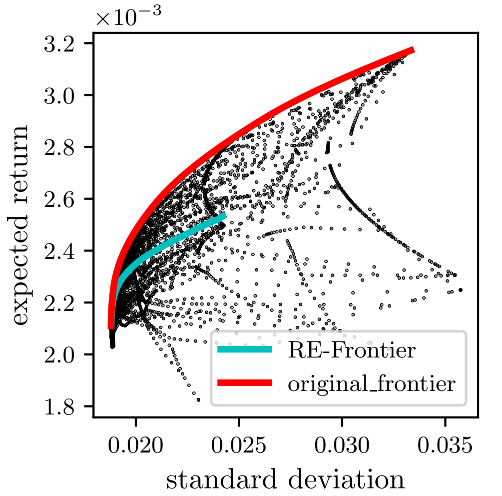

Figure 4.3: Simulated portfolios, REF and original frontier

vary wildly between the simulated scenarios. Some scenarios only have a fraction of the risk and/or return range of the original frontier, whereas others have more than twice the range.

To continue documenting the construction process of the REF, let us take a closer look at Figure 4.3. There, all portfolios of Figure 4.2 are evaluated with respect to the original point estimates, $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$.[3] By averaging all the portfolio weights of the simulated scenarios, i.e. the portfolios represented by the black dots in Figure 4.3, for each value of γ, the REF can be obtained, which is represented by the cyan line (Step 4). The red line corresponds to the original MV efficient frontier that is constructed using $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$. Since all portfolios are evaluated with respect to these estimates, the red curve therefore yields the maximum utility for given values of γ.

I carried out this procedure not only for the case of medium FC (with $R = 1000$ weeks) which I demonstrated with Figure 4.2 and Figure 4.3 but also for low FC (with $R = 350$ weeks) and high FC (with $R = 3000$ weeks). Figure 4.4 shows the corresponding REFs together with the original efficient frontier.

It can easily be deducted, that the resampled frontiers are not optimal anymore, as the original frontier contains portfolios that offer more expected return for the same amount of risk. However it is important to focus on the fact that we are comparing the in-sample performance here and the original efficiency frontier is by its definition the optimal solution (see Problem (2.31)). The real strengths of the resampling approach

[3]At this point it is important to clarify the following: Here, "With respect to" means that portfolio risks and portfolio returns for all the portfolios represented by the black dots in Figure 4.3 are computed using Equation (2.3) and (2.4) with distribution moments $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$.

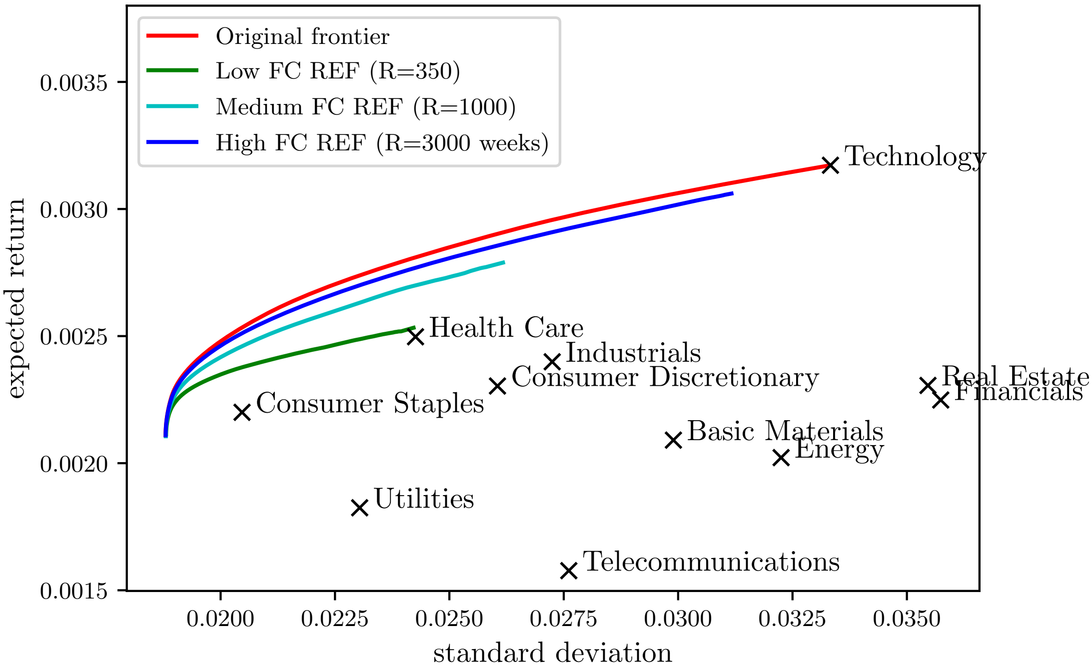

Figure 4.4: Original frontier vs. resampled frontiers

only show up when comparing the out-of-sample performance of the resulting portfolios, a subject I will focus on in Chapter 5.

However, a really intuitive advantage of the resampling approach is the gain in diversification of the efficient portfolios. Figure 4.5 shows the asset weights in portfolios on the original frontier (red curve in Figure 4.4) compared to the asset weights in portfolios on the medium FC REF (cyan curve in Figure 4.4). It is visible that the portfolios on the REF are substantially more diversified and therefore also more robust compared to the portfolios on the original efficient frontier. Just to state a numerical example: For a standard deviation of 0.024, the efficient portfolio on the original frontier contains three assets. The efficient portfolio on the medium FC REF, on the other hand, contains all of the eleven possible assets.

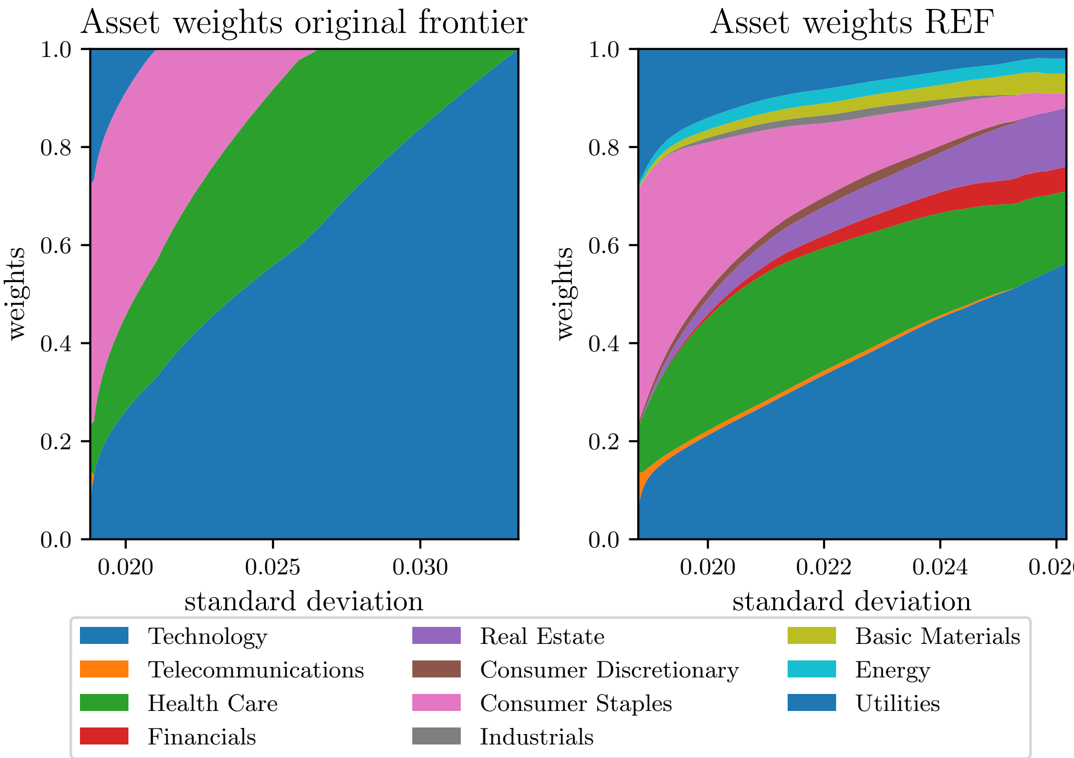

Figure 4.5: Asset weights of portfolios on the original frontier vs. asset weights of portfolios on the medium FC resampled frontier

CHAPTER 5

Simulation study

To check the robustness and the out-of-sample performance of the previously introduced resampling approach, I will compare it to the classical plug-in rule through a Monte-Carlo simulation study, which is common practice in portfolio research. There is a lot of literature on comparisons of different portfolio rules available. There are also papers on the comparison of the traditional approach versus the resampling approach (Michaud and Michaud, 2007; Scherer, 2002; Becker et al., 2009). However, in this simulation study I concentrate on impact of sample size on the results of out-of-sample performance, as this has not really been pointed out in the past.

5.1 Approach

For the simulation study, the estimates of $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$ from the entire sample of return data, $\boldsymbol{\Phi}_{FS}$, from October 1989 until November 2021[1] are taken as the true mean and covariance of the joint return distribution. They are needed in order to generate histories of returns and finally also to score the resulting portfolios.

A detailed explanation of the simulation study's steps can be formulated as follows:

Step 1 Sample a history of returns, $\boldsymbol{\Phi}_j$, with length T centred around $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$.[2] I will refer to the j-th sample of simulated historic returns as $\Phi_j = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_T\}$, with $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_T \in \mathbb{R}^N$ being the simulated return vectors for periods $[0, 1], [1, 2], ..., [T - 1, T]$ respectively.

[1] For the summary statistics, see Appendix A.

[2] An implementation of a multivariate normal simulation can be seen in the function multivariate_normal_simulation() in Appendix B. However, for the actual simulation study of this thesis I used the NumPy implementation as it is implemented more efficiently and showed substantial performance advantages.

Step 2 Compute the optimal portfolio weights for the simulated scenario based on $\boldsymbol{\Phi}_j$ using the plug-in as well as the resampling rule. The optimal weight vector for each portfolio rule will be referred to as $\hat{\boldsymbol{w}}_j^{\text{plug-in}}$ and $\hat{\boldsymbol{w}}_j^{\text{resampled}}$, respectively.[3] Thereafter, calculate the out-of-sample performance for each portfolio rule with the portfolio weights $\hat{\boldsymbol{w}}_{\boldsymbol{j}}$ with respect to the true moments $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$.

Step 3 Repeat Step 1 and 2 M times and calculate the expected out-of-sample performance for each portfolio rule (i.e. $j = \{1, 2, .., M\}$).[4]

Step 4 Execute Steps 1-3 for different lengths of simulated histories, T, different values of γ (0.5, 1, 3 and 5) and the three exemplary levels of FC introduced in Chapter 4 ($R = 350$, $R = 1000$ and $R = 3000$ weeks).

Figure 5.1 displays a graphical overview of the simulation study's structure. It shows, that we deal with a simulation nested in another simulation. This is due to the fact, that I simulate histories of returns, $\boldsymbol{\Phi}_j$, and then calculate the optimal portfolios via the plug-in rule and the resampling rule. To calculate the optimal resampled portfolio weight vector, $\hat{\boldsymbol{w}}_j^{\text{resampled}}$, the resampling rule also needs a simulation of return data, $\boldsymbol{\Phi}_i$. When comparing Figure 4.1 and Figure 5.1 it is important to point out that in this simulation study we do not have a single set of historic point estimates $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$ like in Chapter 4, but rather multiple point estimates from simulated histories $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ centred around the two moments of the Full Sample $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$. This structure is necessary to test the robustness of the resampling approach and compare the out-of-sample performance of the two portfolio rules.

[3] For simplification reasons, the superscripts will be dropped from now on, as I always refer to both portfolio rules in the following sentences and equations.

[4] To determine an appropriate number of simulated histories M in order to get a robust estimate of the expected loss $\mathbb{E}[\rho(\boldsymbol{w}^*, \hat{\boldsymbol{w}})]$, the following procedure was carried out: Under assumption of the M resulting out-of-sample loss values $\{\rho_1, \rho_2, ...\rho_M\}$ being independent and identically distributed (i.i.d.), the mean of the sample $\overline{\rho}$ is approximately normally distributed around the real expected value $\mathbb{E}[\rho(\boldsymbol{w}^*, \hat{\boldsymbol{w}})]$ with an expected deviation of 0 (efficient estimator) and standard error as the standard deviation. This follows from the central limit theorem (CLT). The standard error is computed as follows: $SE = \sqrt{\frac{1}{M-1}\sum_{k=1}^{M}(\rho_k - \overline{\rho})^2}$, and respectively, the relative standard error $RSE = \frac{SE}{\overline{\rho}}$. The goal was to obtain a relative standard error value smaller than 2% for all possible simulation scenarios (all possible γ values, T-values and FC-levels). The limit was deliberately chosen this small due to the fact that it is not an exact representation of the actual relative error, as we do not know the true value of $\mathbb{E}[\rho(\boldsymbol{w}^*, \hat{\boldsymbol{w}})]$ and use $\overline{\rho}$ in the formula. I have found, that with increasing γ, RSE increases and with T and FC increasing, RSE decreases and 5000 simulations are needed in order to obtain a RSE smaller than 2% in every setting needed for this simulation study.

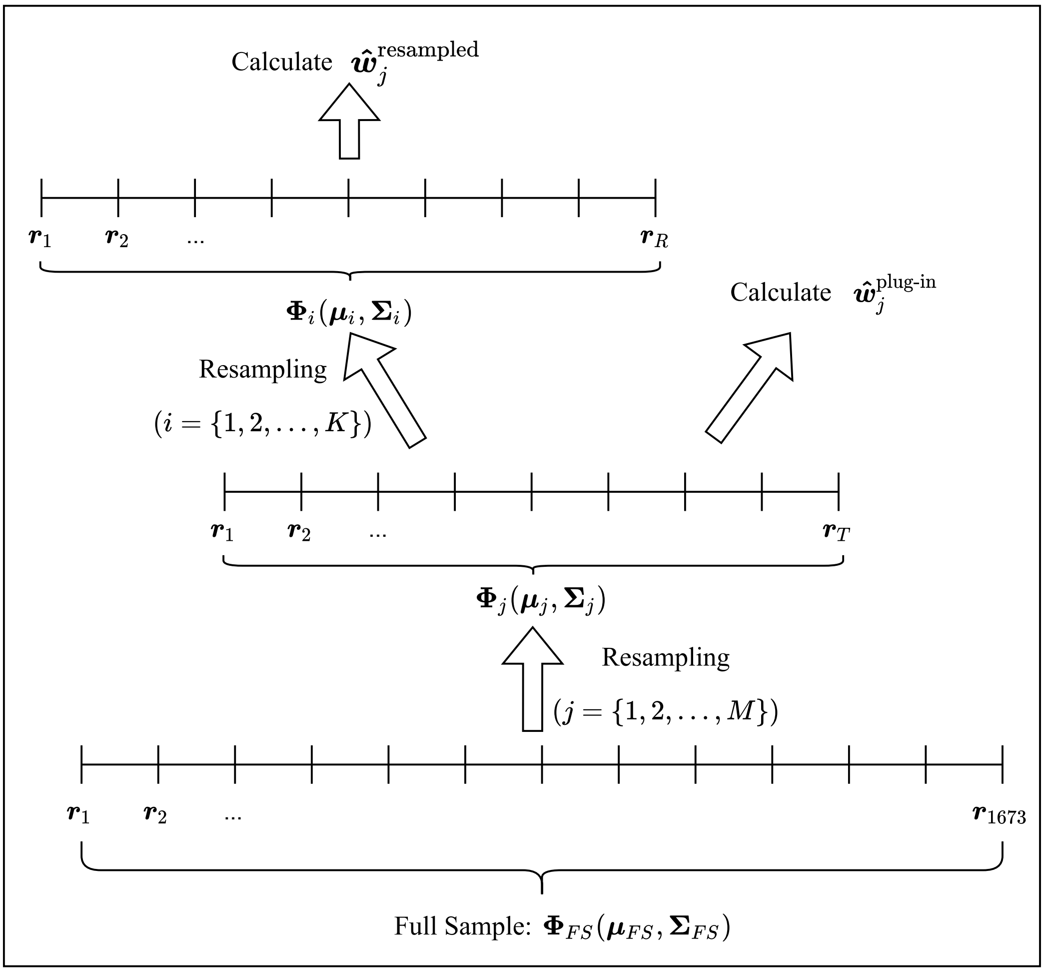

Figure 5.1: Graphical overview of the simulation study

After discussing the structure of the simulation study in detail, I want to focus on the following: An important question to pose concerning Step 2 is how to measure the out-of-sample performance of a portfolio rule. In this thesis, I compare the two portfolio rules based on their expected out-of-sample certainty equivalent of the portfolio return, $\widetilde{r}_{CE}$.[5] I chose this performance measure as it is consistent with the objective function of the optimisation problem defined in Equation (2.31), where I maximise the expected utility by maximising the certainty equivalent of portfolio return. For the remainder of this thesis, I will be using $\widetilde{CE}$ when referring to $\widetilde{r}_{CE}$ for aesthetic reasons in order to improve reading fluency. For each scenario j, the out-of-sample certainty equivalent of portfolio return, $\widetilde{CE}_j$, as a function of the portfolio weights $\hat{\boldsymbol{w}}_j$ can be written as follows:

[5]The tilde over the variable indicates, that it is an out-of-sample property. This notation is valid for the remainder of this thesis.

$$\widetilde{CE}_j(\hat{\boldsymbol{w}}_j) = \tilde{\mu}_{pj} - \frac{\gamma}{2}\tilde{\sigma}^2_{pj} = \hat{\boldsymbol{w}}'_j \boldsymbol{\mu}_{FS} - \frac{\gamma}{2}\hat{\boldsymbol{w}}'_j \boldsymbol{\Sigma}_{FS} \hat{\boldsymbol{w}}_j \tag{5.1}$$

In line with what other authors do when comparing performance (Stambaugh, 1997; Jorion, 1986; Frost and Savarino, 1986; Horst et al., 2000; Kan and Zhou, 2007), I introduce a relative loss function $\rho_j(\boldsymbol{w}^*, \hat{\boldsymbol{w}}_j)$, with $\hat{\boldsymbol{w}}_j$ being the weight vector calculated in scenario j in Step 2 and $\boldsymbol{w}^*$ being the optimal weight vector computed via (2.31) with $\boldsymbol{\mu}_{FS}$ and $\boldsymbol{\Sigma}_{FS}$. The loss function can then be written as follows:

$$\rho_j(\boldsymbol{w}^*, \hat{\boldsymbol{w}}_j) = \frac{L_j(\boldsymbol{w}^*, \hat{\boldsymbol{w}}_j)}{CE(\boldsymbol{w}^*)} = \frac{CE(\boldsymbol{w}^*) - \widetilde{CE}_j(\hat{\boldsymbol{w}}_j)}{CE(\boldsymbol{w}^*)} \tag{5.2}$$

It is important to reinforce here, that since $\hat{\boldsymbol{w}}_j$ is a random variable depending on $\boldsymbol{\Phi}_j$ and the porfolio rule used and different in each scenario j, $\widetilde{CE}_j(\hat{\boldsymbol{w}}_j)$ and ρ_j also take on different values in each scenario and therefore one has to evaluate a portfolio rule based on its expected out-of-sample loss $\mathbb{E}[\rho(\boldsymbol{w}^*, \hat{\boldsymbol{w}})]$. Kan and Zhou (2007, p. 625) put it into the following very well describing words: *"Therefore $\mathbb{E}[\rho]$ represents the expected loss over all possible realizations of $\boldsymbol{\Phi}_j$ that are incurred in using the portfolio rule $\hat{\boldsymbol{w}}$."*

Having addressed the structure and performance measure of the simulation study, I hereby provide a short insight into the actual implementation of it. For this study, sample sizes from 50 weeks up to 1500 weeks with a step size of 50 were considered ($T = 50, 100, ..., 1500$). For each sample size 5000 simulations had to be carried out as explained above. Considering that I compare the plug-in strategy to 3 different resampling strategies (3 different FC levels), the function efficiency_frontier() in Listing B.1 has to be called 45.150.000 times in total. For each function call, the Problem (2.31) had to be solved for the four different γ values that were compared. Therefore the optimisation problem had to be solved 180.600.000 times in total. Due to this considerable computational load, I implemented a parallel computing routine in order to increase efficiency. Listing B.2 shows the implemented python code. This code was then executed on a virtual machine on the Google Cloud Platform (GCP) with 16CPUs. The results were afterwards interpreted on a local machine.

5.2 Results

In this section I will present and discuss the outcome of the simulation study.

Figure 5.2 shows a visual representation of the results. For each value of γ, I show the expected loss function $\mathbb{E}[\rho(\boldsymbol{w}^*, \hat{\boldsymbol{w}})]$ over different simulated sample sizes T for different levels of FC. An issue that needs addressing at this point is the relation between the number of weeks used for generating histories (T) and the number of weeks used for resampling within the resampling portfolio rule (R). It might seem questionable to resample for $R = 3000$ weeks (high FC) when dealing with historic return data of $T = 50$ weeks at first. Since R is a measure of how confident an investor is in the historic point estimates $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$ (or $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ in the simulation study), it is questionable whether it is useful to fully trust the estimates of the distribution's moments with 50 weeks of return data ($T = 50$). However, when looking at the graphs in Figure 5.2, it is visible that the difference in expected out-of-sample loss between the plug-in rule and the different levels of FC is smaller for shorter histories, i.e. low values of T, than for larger ones. This fact relativizes the above mentioned concern, however, this would be an interesting topic for further research.

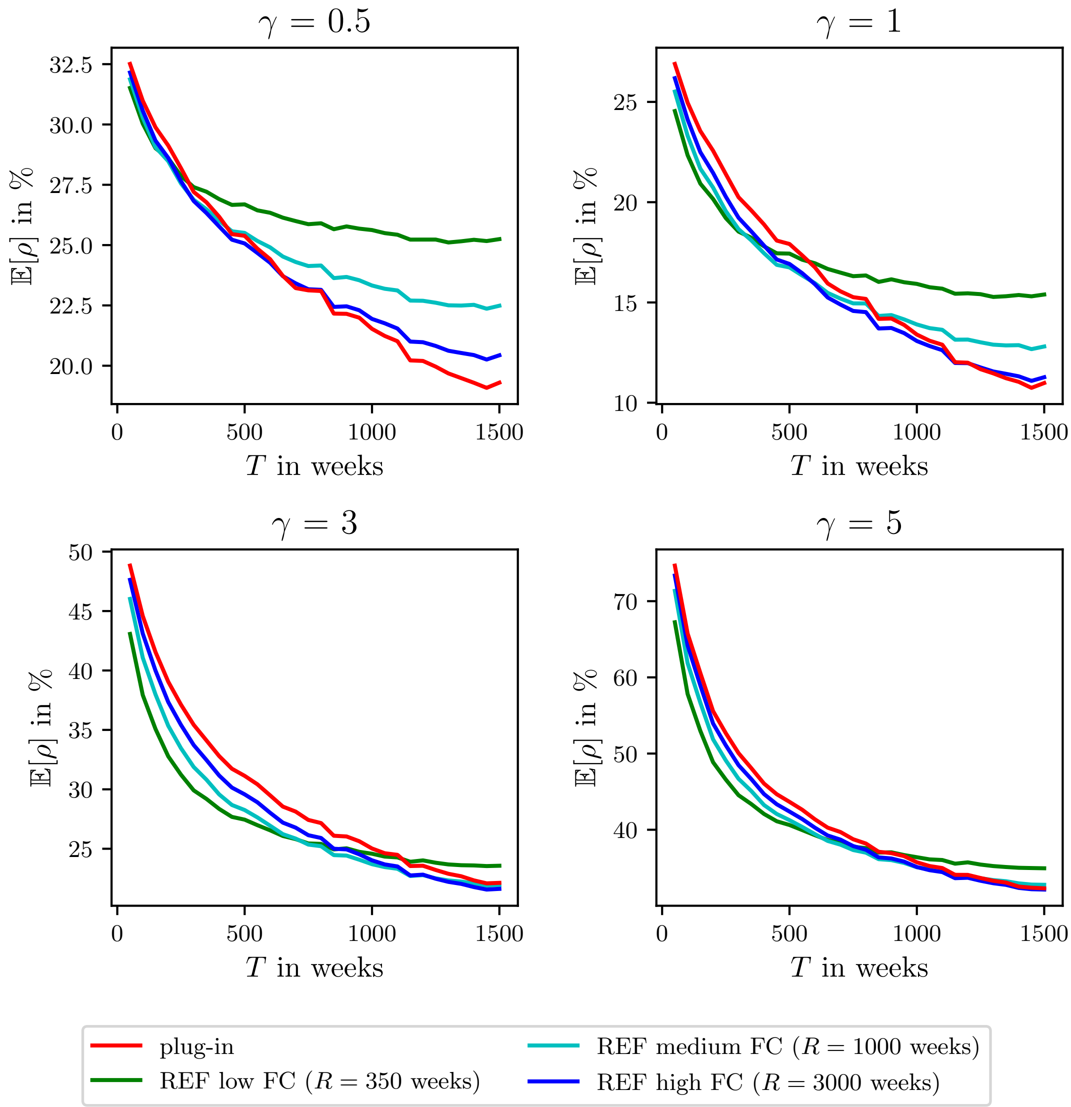

Figure 5.2: Expected loss $\mathbb{E}[\rho]$ of the different portfolio rules over sample size T

In order to show the exact values of the data, Table 5.1 provides a numerical summary of the results.

γ	T	plug-in	resampled low FC	resampled med FC	resampled high FC
0.5	50	32.5	31.5	31.9	32.2
	250	28.2	27.9	27.6	27.7
	500	25.4	26.7	25.5	25.1
	750	23.1	25.9	24.1	23.2
	1000	21.5	25.6	23.3	21.9
1.0	50	26.9	24.5	25.5	26.2
	250	21.4	19.2	19.6	20.3
	500	17.9	17.4	16.7	16.9
	750	15.3	16.3	15.0	14.6
	1000	13.4	15.9	13.9	13.1
3.0	50	48.8	43.1	46.0	47.6
	250	37.1	31.2	33.5	35.4
	500	31.1	27.4	28.3	29.6
	750	27.4	25.4	25.4	26.1
	1000	25.0	24.6	23.7	24.0
5.0	50	74.7	67.2	71.4	73.4
	250	52.7	46.6	49.1	51.1
	500	43.7	40.6	41.3	42.4
	750	38.8	37.8	37.3	37.9
	1000	35.7	36.4	35.1	35.1

Table 5.1: Expected loss $\mathbb{E}[\rho]$ in %

As Figure 5.2 and Table 5.1 show the expected loss for different parametrizations, i.e. the arithmetic mean values over the 5000 simulations, I conducted a hypothesis test in order to add some statistical relevance to the results. The following null and alternative hypotheses were defined:

$$H_0 : \mathbb{E}[\rho_{\text{resampled}}] - \mathbb{E}[\rho_{\text{plug-in}}] \geq 0 \tag{5.3}$$

$$H_a : \mathbb{E}[\rho_{\text{resampled}}] - \mathbb{E}[\rho_{\text{plug-in}}] < 0 \tag{5.4}$$

Due to the nature of the data, a paired student's t-test was implemented. With a significance level of $\alpha = 0.05$, the sample sizes T for which the null-hypothesis could be rejected are summarised in Table 5.2

In general, one can observe from Figure 5.2, Table 5.1 and Table 5.2 that the performance of the resampling approach is good, i.e. the loss is low, if the level of confidence (R) corresponds to the length of the history (T). If T is small, the sample moments of a

γ	resampled low FC	resampled med FC	resampled high FC
0.5	50-250	50-400	50-600
1.0	50-550	50-800	50-1100
3.0	50-1100	50-1500	50-1500
5.0	50-800	50-1200	50-1500

Table 5.2: Sample sizes T for which the null hypothesis can be rejected with $\alpha = 0.05$

simulated history vary around the full sample moments by a large extent. The results confirm that resampling with low R (accounting for the high sampling error) leads to a robust portfolio choice. Larger R (putting more confidence to a short history) result in larger losses. The worst performing portfolio rule with short histories is the plug-in rule, which implicitly assumes that the sample of length T reflects the full sample moments, which is unlikely to be true. With increasing history length T, a portfolio rule that is overly cautious, i.e. the resampling rule with small R, overestimates the sampling error more and more and lags behind resampling rules with higher R, that assign reasonable confidence to the values sample moments. With T further increasing, larger and larger R become optimal. Finally, very long simulated histories resemble the full sample moments well and, consequently, the plug-in rule starts to become the rule with the smallest loss as it is by definition the optimal portfolio strategy when knowing the return distribution's true moments.

Another result to be deducted from the data is that the higher an investor's risk aversion is, the more attractive the resampling portfolio rule becomes. When looking at Table 5.1 and comparing the expected loss values of the resampling approach with low FC ($R = 350$ weeks) and the plug-in rule for the smallest sample size of $T = 50$ weeks and $\gamma = 5$, it can be seen that the difference in expected loss is 7.5 percentage points. With decreasing γ, the difference in expected loss decreases as well and when comparing the values for $\gamma = 0.5$, the difference in expected loss between the plug-in rule and the resampling approach with low FC is only 1 percentage point. This behaviour of decreasing difference of expected loss with decreasing γ is also present when comparing the plug-in rule to the other two resampling rules with medium and high FC. Therefore one can state, that the more risk averse an investor is, the more attractive the concept of diversification becomes and therefore the more preferable the resampled approach becomes.

Another result to discuss concerning the relationship of relative risk risk aversion with the performance of the resampling rule is the following: With low relative risk aversion, portfolio optimisation puts more weight on maximising expected return rather than reducing portfolio variance. I.e., these investors have less desire for portfolio robustness and, consequently, suffer a loss from being overly cautious (resampling with small R). Investors with high relative risk aversion put a higher value on portfolio robustness and, hence, suffer less losses from cautious portfolio strategies, meaning that they prefer rather low R values over larger ranges of T. This fact can easily be seen in Figure 5.2 when comparing the four sub-graphs. In general, one can deduct that the larger the relative risk aversion, the larger the range of sample size T where the resampling rules with low R outperform the ones with higher R and the plug-in rule becomes. Table 5.2 also shows this result very clearly.

One might argue that this is not the case when comparing the results for $\gamma = 3$ and $\gamma = 5$. Merton (1980) showed that in general the estimates of covariances we obtain from return data are more accurate than the estimates of means. When looking at our objective function in Problem (2.31), it becomes obvious that estimated means do not play such a significant role if $\gamma = 5$ compared to $\gamma = 3$. This might be a possible explanation for this deviation in the results. However I have to emphasise that there are likely to be other effects that are not straightforward, but rather complex and nonlinear that might lead to this deviation. Nonetheless they do not destroy the general image discussed in the paragraph above.

CHAPTER 6

Conclusion

This thesis studied the impact of sample size on the out-of-sample performance of optimal portfolios. I started with a short introduction to the classical MV framework, then provided an insight into the field of utility theory and derived an optimisation problem for the MV framework using a CRRA utility function. After applying this framework to real historic return data of 11 sectors of the S&P500, the problems that arise frequently when using this framework in the real world were pointed out and discussed extensively, namely parameter uncertainty and estimation error. Many approaches have been developed to deal with these problems, but the one I focused on in this thesis is a resampling approach, that tries to build more robust portfolios when facing parameter uncertainty.

After a detailed description and an application example of the resampling portfolio rule, the classical plug-in rule and the resampling approach were examined and compared against each other using a Monte Carlo simulation study. I simulated a large number of histories of different lengths and compared the expected out-of-sample certainty equivalent of portfolio return for portfolios constructed using both portfolio rules. The results showed that the resampling rule has a significantly higher expected out-of-sample certainty equivalent compared to the plug-in rule, when the samples used for parameter estimation were small enough. With increasing sample size, both portfolio rules started to show a similar out-of-sample performance and eventually when samples became large enough, the classical plug-in rule outperformed the resampling approach. This behaviour can be explained by the fact that as sample size increases, we experience a convergence of the sample moments to the moments of the underlying distribution. Since these moments of the underlying distribution are also used to score the portfolios, it becomes clear that the plug-in rules performs better, as it is by definition the method that maximises expected return for a given risk, when knowing the true parameters of the distribution of returns. The other main implication of the results is, that the resampling approach becomes more attractive for an investor as his coefficient of relative risk aversion, γ, increases. When comparing the expected out-of-sample certainty equivalent for both portfolio rules, the gain in performance when using the resampling rule instead of the the plug-in rule increases with increasing risk aversion.

List of Figures

List of Tables

Bibliography

Agrawal, A., Verschueren, R., Diamond, S., and Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60.

Arrow, K. (1965). *Aspects of the theory of risk-bearing.* Yrjö Jahnssonin Säätiö.

Arrow, K. J. (1963). Liquidity preference. lecture vi in lecture notes for economics 285, the economics of uncertainty. *Stanford University*, pages 33–53.

Ashwin, R. and Tikhon, J. (2021). *Foundations of Reinforcement Learning with Applications in Finance.*

Barry, C. B. (1974). Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, 29(2):515–522.

Bawa, V. S., Brown, S. J., and Klein, R. W. (1979). *Estimation Risk and Optimal Portfolio Choice.* North Holland.

Becker, F., Gürtler, M., and Hibbeln, M. (2009). Markowitz versus michaud: Portfolio optimization strategies reconsidered. Working Paper Series IF30V3, Braunschweig.

Best, M. and Grauer, R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, 4:315–42.

Black, F. and Litterman, R. (1990). Asset allocation: Combining investors views with market equilibrium. *Fixed Income Research.*

Black, F. and Litterman, R. (1991). Global asset allocation with equities, bonds, and currencies. *Fixed Income Research.*

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

Brandt, M. W. (2010). Chapter 5 - portfolio choice problems. In AÏT-SAHALIA, Y. and HANSEN, L. P., editors, *Handbook of Financial Econometrics: Tools and Techniques*, volume 1 of *Handbooks in Finance*, pages 269–336. North-Holland, San Diego.

Chapados, N. (2011). *Portfolio Choice Problems: An Introductory Survey of Single and Multiperiod Models*. SpringerBriefs in Electrical and Computer Engineering. Springer New York.

Chopra, V. K., Hensel, C. R., and Turner, A. L. (1993). Massaging mean-variance inputs: Returns from alternative global investment strategies in the 1980s. *Management Science*, 39(7):845–855.

Chopra, V. K. and 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.

Clemen, R. T. (1996). *Making hard decisions: an introduction to decision analysis*. Brooks/Cole Publishing Company.

Dangl, T. and Kashofer, M. (2013). Minimum-variance stock picking - a shift in preferences for minimum-variance portfolio constituents. *SSRN Electronic Journal*.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and statistics*. Addison Wesley.

Demiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22.

Diamond, S. and Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5.

Dickinson, J. P. (1974). The reliability of estimation procedures in portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 9(3):447–462.

Eeckhoudt, L., Gollier, C., and Schlesinger, H. (2011). *Economic and Financial Decisions under Risk*. Princeton University Press.

Fabozzi, F., Focardi, S., Kolm, P., and Pachamanova, D. (2007). *Robust Portfolio Optimization and Management*. Frank J. Fabozzi series. Wiley.

Focardi, S. M. and Fabozzi, F. J. (2004). *The mathematics of financial modeling and investment management*, volume 138. John Wiley & Sons.

Frankfurter, G. M., Phillips, H. E., and Seagle, J. P. (1971). Portfolio selection: The effects of uncertain means, variances, and covariances. *The Journal of Financial and Quantitative Analysis*, 6(5):1251–1262.

Frost, P. A. and Savarino, J. E. (1986). An empirical bayes approach to efficient portfolio selection. *The Journal of Financial and Quantitative Analysis*, 21(3):293–305.

Horst, J., de Roon, F., and Werker, B. (2000). Incorporating estimation risk in portfolio choice. *Tilburg University, Center for Economic Research, Discussion Paper*.

Ingersoll, J. E. (1987). *Theory of financial decision making / Jonathan E. Ingersoll, Jr.* Rowman Littlefield studies in financial economics. Rowman Littlefield, Totowa, N.J.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Jobson, J. D. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *The Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

Levy, H. and Markowtiz, H. (1979). Approximating expected utility by a function of mean and variance. *American Economic Review*, 69:308–17.

MacKinlay, A. C. and Ľuboš Pástor (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies*, 13(4):883–916.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. A Cowles Foundation Monograph. Yale University Press.

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis*, 7(4):1851–1872.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Michaud, R. and Michaud, R. (2004). Forecast confidence level and portfolio optimization. *Newsletter: New Frontier Advisors, Boston, July.*

Michaud, R. and Michaud, R. (2007). Estimation error and portfolio optimization: A resampling solution. *Journal of Investment Management*, Vol. 6:pp. 8 – 28.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42.

Michaud, R. O. and Michaud, R. O. (2008). *Efficient asset management: a practical guide to stock portfolio optimization and asset allocation*. Oxford University Press.

Norstad, J. (1999). An introduction to utility theory. *Unpublished manuscript at http://homepage. mac. com/j. norstad*.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136.

Pástor, (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223.

Scherer, B. (2002). Portfolio resampling: Review and critique. *Financial Analysts Journal*, 58(6):98–109.

Stambaugh, R. F. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics*, 45(3):285–331.

Stellato, B., Banjac, G., Goulart, P., Bemporad, A., and Boyd, S. (2020). OSQP: an operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 12(4):637–672.

Ľuboš Pástor and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381.

APPENDIX A

S&P 500 summary statistics

	mean
Technology	0.003174
Telecommunications	0.001578
Health Care	0.002499
Financials	0.002250
Real Estate	0.002307
Consumer Discretionary	0.002305
Consumer Staples	0.002201
Industrials	0.002400
Basic Materials	0.002092
Energy	0.002023
Utilities	0.001826

Table A.1: Mean weekly returns of 11 sectors in the S&P500 from October 1989 until November 2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	0.1112	0.0642	0.0385	0.0660	0.0543	0.0624	0.0268	0.0610	0.0526	0.0432	0.0255
(2)	0.0642	0.0763	0.0339	0.0600	0.0442	0.0493	0.0277	0.0501	0.0409	0.0399	0.0297
(3)	0.0385	0.0339	0.0589	0.0522	0.0371	0.0400	0.0359	0.0433	0.0383	0.0367	0.0277
(4)	0.0660	0.0600	0.0522	0.1278	0.0770	0.0709	0.0430	0.0800	0.0709	0.0646	0.0396
(5)	0.0543	0.0442	0.0371	0.0770	0.1258	0.0580	0.0359	0.0665	0.0726	0.0496	0.0390
(6)	0.0624	0.0493	0.0400	0.0709	0.0580	0.0679	0.0336	0.0600	0.0541	0.0430	0.0281
(7)	0.0268	0.0277	0.0359	0.0430	0.0359	0.0336	0.0419	0.0370	0.0343	0.0320	0.0283
(8)	0.0610	0.0501	0.0433	0.0800	0.0665	0.0600	0.0370	0.0743	0.0668	0.0569	0.0345
(9)	0.0526	0.0409	0.0383	0.0709	0.0726	0.0541	0.0343	0.0668	0.0894	0.0659	0.0332
(10)	0.0432	0.0399	0.0367	0.0646	0.0496	0.0430	0.0320	0.0569	0.0659	0.1041	0.0383
(11)	0.0255	0.0297	0.0277	0.0396	0.0390	0.0281	0.0283	0.0345	0.0332	0.0383	0.0531

Table A.2: Covariance matrix, scaled with 10^2, with indices as follows: (1)...Technology, (2)...Telecommunications, (3)...Health Care, (4)...Financials, (5)...Real Estate, (6)...Consumer Discretionary, (7)...Consumer Staples, (8)...Industrials, (9)...Basic Materials, (10)...Energy, (11)...Utilities

APPENDIX B

Python implementation

Listing B.1: Implemented functions for the simulation study

```python
import numpy as np
import pandas as pd
import cvxpy as cp

def efficiency_frontier(returns, gamma_vals=[]):
    """
    Calculates the efficient portfolios for a given matrix of returns
        and gamma values. If no gamma values are passed, it
        calculates efficient portfolios for gammas from 0.1 to 100.

    :param returns: returns [TxN] with T being the number of periods
        and N being the number of stocks
    :param gamma_vals: gamma values for which the efficient portfolio
        should be calculated
    :type returns: np.ndarray
    :type gamma_vals: list
    :return: dataframe with efficiency frontier
    :rtype: pandas.DataFrame
    """
    cov_matrix = np.cov(returns.T)
    mu = np.mean(returns, axis=0)
    n = mu.size
    if not bool(gamma_vals):
        points = 100
        gamma_values = np.geomspace(0.1, 100, num=points)
    else:
        points = len(gamma_vals)
        gamma_values = gamma_vals

```

```
    sd_data = np.zeros(points)
    ret_data = np.zeros(points)
    w_data = []

    # problem definition
    w = cp.Variable(n)
    gamma = cp.Parameter(nonneg=True)
    ret = mu.T @ w
    variance = cp.quad_form(w, cov_matrix)
    constraints = [cp.sum(w) == 1, w >= 0]
    prob = cp.Problem(cp.Maximize(ret - (1/2) * gamma * variance),
        constraints)

    # problem solving
    for i in range(points):
        gamma.value = gamma_values[i]
        prob.solve(solver=cp.OSQP)
        sd_data[i] = cp.sqrt(variance).value
        ret_data[i] = ret.value
        w_data.append(np.array(w.value))
    frontier = pd.DataFrame({'gamma': gamma_values, 'w': w_data, 'sd'
        : sd_data, 'ret': ret_data})
    return frontier


def re_efficient_frontier(returns, weeks_resampling, gamma_vals=[]):
    """
    Calculates the resampled efficiency frontier for a given matrix
        of returns and gamma values. If no gamma values are passed, it
         calculates efficient portfolios for gammas from 0.1 to 100.

    :param returns: returns [TxN] with T being the number of periods
        and N being the number of stocks
    :param gamma_vals: gamma values for which the resampled efficient
         portfolio should be calculated
    :param weeks_resampling: number of weeks used for sample
        generation in the resampling process, i.e. ForecastConfidence
        (FC)
    :type returns: np.ndarray
    :type gamma_vals: list
    :type weeks_resampling: int
    :return: dataframe with resampled efficiency frontier
    :rtype: pandas.DataFrame
    """
    # simulation of scenarios and calculation of the efficient
        frontiers
    simulations = 100
    efficiency_frontiers = []
    mean = np.mean(returns,axis=0)
```

```
    cov = np.cov(returns.T)
    for i in range(simulations):
        resampled_returns = np.random.multivariate_normal(mean,cov,
            weeks_resampling)
        efficient_frontier = efficiency_frontier(resampled_returns,
            gamma_vals)
        efficiency_frontiers.append(efficient_frontier)

    # calculation of the resampled efficiency frontier
    re_efficient_portfolios = []
    for gamma in gamma_vals:
        asset_weights = np.zeros(mean.size)
        for efficient_frontier in efficiency_frontiers:
            w = efficient_frontier[efficient_frontier['gamma'] ==
                gamma]['w'].iat[0]
            asset_weights = asset_weights + w
        asset_weights = asset_weights / asset_weights.sum()
        ret_portfolio = portfolio_return(asset_weights, mean)
        sd_portfolio = portfolio_sd(asset_weights, cov)
        re_efficient_portfolios.append([gamma, asset_weights,
            sd_portfolio, ret_portfolio])
    re_efficient_portfolios_df = pd.DataFrame(re_efficient_portfolios
        , columns=['gamma', 'w', 'sd', 'ret'])
    return re_efficient_portfolios_df


def multivariate_normal_simulation(data, periods):
    n = len(data.columns)
    mu = data.mean()
    cov_matrix = data.cov()
    sd = np.sqrt(data.var()).values
    volatility_matrix = np.zeros((n, n))
    np.fill_diagonal(volatility_matrix, sd)
    volatility_matrix_inv = np.linalg.inv(volatility_matrix)
    corr_matrix = np.dot(np.dot(volatility_matrix_inv, cov_matrix),
        volatility_matrix_inv)
    eig_values, eig_vectors = np.linalg.eig(corr_matrix)
    eig_values_matrix = np.zeros((n, n))
    np.fill_diagonal(eig_values_matrix, eig_values)
    normal_matrix = np.random.normal(0, 1, (periods, n))
    return_matrix = np.dot(np.dot(np.dot(normal_matrix, np.sqrt(
        eig_values_matrix)), eig_vectors.T), volatility_matrix)
    for i in range(n):
        return_matrix[:, i] += mu.values[i]
    return return_matrix


def portfolio_return(weights, mu):
    ret = weights.T.dot(mu)
```

```
    return ret


def portfolio_sd(weights, cov):
    return np.sqrt(weights.T.dot(cov.dot(weights)))
```

Listing B.2: Simulation study with multiprocessing

```
import numpy as np
import pandas as pd
import functions
import multiprocessing
import time

if __name__ == '__main__':
    # load the financial data
    SP500_data = pd.read_csv(
        'SPCOMP_IndustryRIAbsSimpR_W_USD.csv',
        usecols=['date', "data.Technology", "data.Telecommunications", "data.Health.Care", "data.Financials",
                 "data.Real.Estate", "data.Consumer.Discretionary", "data.Consumer.Staples",
                 "data.Industrials", "data.Basic.Materials", "data.Energy", "data.Utilities"])

    SP500_data['date'] = pd.to_datetime(SP500_data['date'])
    SP500_data.set_index('date', inplace=True)
    SP500_data = SP500_data.dropna(how='all')
    SP500_data = SP500_data.divide(100)
    SP500_data.rename(columns=lambda x: x[5:], inplace=True)
    SP500_data.rename(columns=lambda x: x.replace(".", " "), inplace=True)

    # Parameters to config the script
    number_of_histories = 5000
    FC_1 = 350
    FC_2 = 1000
    FC_3 = 3000
    gamma_values = [0.5, 1, 3, 5]
    print('number of available CPUs:', multiprocessing.cpu_count())

    # Simulate a large number of histories
    true_returns = SP500_data
    mu_true = true_returns.mean().to_numpy()
    cov_true = true_returns.cov().to_numpy()
    start = int(input('At what T to start?'))
    for length_simulated_history in range(start, 1550, 50):
        print('@T=' + str(length_simulated_history))
        histories = []
        for i in range(number_of_histories):
            histories.append(np.random.multivariate_normal(mu_true, cov_true, length_simulated_history))

        # Calculate original and re_efficient frontiers with multiprocessing for speedup
        print('computing plug-in rule...')
```

```
        pool1 = multiprocessing.Pool(processes=multiprocessing.cpu_count())
        input_original_frontier = []
        for i in histories:
            input_original_frontier.append([i, gamma_values])
        start = time.time()
        original_frontiers = pool1.starmap(functions.efficiency_frontier, input_original_frontier)
        pool1.close()
        pool1.join()
        pool2 = multiprocessing.Pool(processes=multiprocessing.cpu_count())
        input_re_frontier_1 = []
        input_re_frontier_2 = []
        input_re_frontier_3 = []
        for i in histories:
            input_re_frontier_1.append([i, FC_1, gamma_values])
            input_re_frontier_2.append([i, FC_2, gamma_values])
            input_re_frontier_3.append([i, FC_3, gamma_values])
        print('computing RE-frontier low FC...')
        re_frontiers_1 = pool2.starmap(functions.re_efficient_frontier, input_re_frontier_1)
        pool2.close()
        pool2.join()
        print('computing RE-frontier medium FC...')
        pool3 = multiprocessing.Pool(processes=multiprocessing.cpu_count())
        re_frontiers_2 = pool3.starmap(functions.re_efficient_frontier, input_re_frontier_2)
        pool3.close()
        pool3.join()
        print('computing RE-frontier high FC...')
        pool4 = multiprocessing.Pool(processes=multiprocessing.cpu_count())
        re_frontiers_3 = pool4.starmap(functions.re_efficient_frontier, input_re_frontier_3)
        pool4.close()
        pool4.join()
        end = time.time()
        print('time for computation parallel: ', end - start)

        # group list of dataframes to single dataframe
        original_frontiers_df = pd.concat(original_frontiers)
        original_frontiers_df['method'] = 'classical mv'
        re_frontiers_df_1 = pd.concat(re_frontiers_1)
        re_frontiers_df_1['method'] = 'resampling approach FC=350 weeks'
        re_frontiers_df_2 = pd.concat(re_frontiers_2)
        re_frontiers_df_2['method'] = 'resampling approach FC=1000
```

```
weeks'
        re_frontiers_df_3 = pd.concat(re_frontiers_3)
        re_frontiers_df_3['method'] = 'resampling approach FC=3000 weeks'
        df_out = pd.concat([original_frontiers_df, re_frontiers_df_1, re_frontiers_df_2, re_frontiers_df_3]).reset_index(drop=True)

        df_out.to_pickle('results_simulation_study/df_T='+str(length_simulated_history)+'.pkl')
```