Essays On Modern Portfolio Theory And Decision Making Under Risk

Ph.D. Thesis

George K. Samartzis

A thesis presented for the degree of

Doctor of Philosophy

Department of Banking and Financial Management

University of Piraeus

February, 2023

To my beloved parents and Lia who stood by me through this journey

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Nikitas Pittis, for his mentorship and guidance. I have been extremely fortunate to experience your unparalleled ability to explain complex ideas in a structured and intuitive way. Our discussions have been thought-provoking and fruitful. It was truly a pleasure and a great honour working with you.

Many thanks to my doctoral advisory committee, Professors Nikolaos Kourogenis and Nikolaos Englezos. I am deeply indebted to both of you, for your genuine interest, insightful comments and constructive criticism.

I am also very grateful to all my Professors at the Department of Banking and Financial Management who provided me support, motivation and knowledge. I must refer separately to Professor Michail Anthropelos, the Director of our Ph.D program, who devoted a lot of his time to organize the Doctoral and Academic Seminars that helped us so much to grow in experience. I would also like to extend my sincere thanks to our exceptionally well-organized Administrative Office.

Moreover, I would like to thank my fellow Ph.D students. I enjoyed sharing our research ideas and I am very grateful that I got to know all of you. A special thanks goes to my dear friends Dr. L. Zarpala and P. Zaverdinos with whom we shared this journey from the very beginning.

This endeavour would not have been possible without the support of my mother, father and brother. I am eternally grateful for the sacrifices you have made on my behalf. Your own accomplishments, taught me that the only way to achieve my goals is by being persistent and hard working.

Last but not least, words cannot express my gratitude to my girlfriend, Lia, for her support and patience all those years we have been together. You are my best and dearest friend and I am not sure how I could accomplish anything without you.

Contents

Aut inveniam viam aut faciam.

I shall either find a way, or make one.

Hannibas

Chapter 1

Introduction

In 1990, Harry Markowitz received the Nobel prize for his pioneering work in Modern Portfolio Theory. It was back in 1952 Markowitz [64] when he introduced the Mean Variance Criterion (hereafter; MVC) or as is else called the Mean Variance Optimization (hereafter; MVO) method. The idea was simple but rather intriguing. Under the appropriate assumptions, an investor could diversify his portfolio based only on the first two moments of his portfolio's returns. This perspective was so groundbreaking that led to some sort of skepticism, with the first reaction coming directly from Markowitz's doctoral supervisor, Nobel laureate Professor Milton Friedman. Professor Friedman remarked that although Markowitz's idea was interesting, it was not economics. The reason Friedman made such a statement was because at that point of time (and still to this day) the academic world had adopted the idea that all that an investor wants to do is to maximize his expected utility. This concept, known as the Representation Theorem, had been proven by Von Neumann and Morgenstern (hereafter; VN-M). Thus, any new idea had to be analyzed under this framework. In 1959, Markowitz [65] addressed this criticism by analyzing his idea under the notion of utility. More specifically, Markowitz made the case that for any investor with an approximately quadratic utility function, the MVO method is the optimal decision making rule. To this day, any academic textbook or research paper is referring to the MVO method under only two conditions. Namely, either (i) the investor has quadratic preferences, or (ii) the portfolio returns are Gaussian. Markowitz

(2010, 2014) [63], [66] insists that he never made the case that the utility functions have to be quadratic or that the portfolio returns have to be Gaussian. Chapters 2 and 3 of this thesis examine the MVO method with respect to its practicality as well as its connection to the well-known Stochastic Dominance Rules (hereafter; SDRs).

Another major notion is the degree of risk aversion of investors. Arrow and Pratt introduced this degree of risk aversion in absolute and relative terms. Namely, the Absolute and Relative Risk Aversion (hereafter; ARA and RRA). The meaning of this notion can be deduced directly by its name. It measures how much the investors "fear" risk. According to Arrow and Pratt, ARA considers the investor's level of risk-aversion with respect to a lottery for different levels of wealth, while RRA considers the investor's level of risk-aversion with respect to a lottery that is a function of the investor's wealth. The literature has put a great deal of effort to determine what is the slope of each coefficient as the level of wealth increases. In total, the majority of the empirical evidence showcases that investors exhibit decreasing ARA (hereafter; DARA). In terms of the RRA the literature has found mixed evidence showing decreasing, constant or increasing RRA (hereafter; DRRA, CRRA and IRRA). Most of the works approached this subject by doing cross-sectional analysis on a set of different portfolio allocations of investors. Another part has used panel data while others used options. In the fourth chapter, we will revisit this subject by introducing a new approach to extract information with respect to the investors' level of ARA and RRA.

Chapter 2, offers our own view on the practicality of the MVO method. A plethora of research works has declared the method to be extremely sensitive to the input parameters even under normality. In fact, Michaud (1989) [67] characterizes the MVO method as "*an estimation-error-maximizer*". To offer our view on that, we use Monte Carlo (hereafter; MC) Simulations. Specifically, following the rationale of DeMiguel et al. (2009) [24], we compare the MVO method to the so-called "naive" strategy (or equivalently, $1/N$) under multiple generations Gaussian returns. Contrary to DeMiguel et al., we conclude that the MVO method outperforms $1/N$ as soon as we control the differences between the "true" Sharpe ratios produced by each strategy and the distance between the "true" target portfolio return from the Global Minimum Variance (hereafter; GMV). Since the method works properly under normality, we argue that more effort has to be made in order to determine the effect of the Data Generating Process (hereafter; DGP) on the MVO method. In fact, we introduce a unique idea on this matter. Testing the method under a DGP with multiple structural breaks in mean. We find empirical evidence that justifies our idea and we proceed to test the method under multiple breaks in mean, using once again MC Simulations. Our results indicate that an investor that overlooks a large number of breaks will receive significantly worse results compared to a DGP with no breaks. On top of that, the MVO method delivers inferior results to the $1/N$ strategy. This is mainly due to the investor using the entire sample to estimate the variance-covariance matrix, which will lead to an inflated estimation and so to a higher portfolio risk.

Chapter 3, looks at the MVO method in terms of its theoretical value. In this chapter, we refer to an investor deciding between just two lotteries. The MVO term is replaced with the more common term known as the MVC. Our main goal is to clarify under which assumptions does the MVC coincide with the SDRs. A SDR offers a specific decision criterion, which constitutes a relation between the distributions of the two lotteries, for a specific class of investors. This rule leads this specific class of investors to maximize its expected utility. The reason we are interested into the relation between the SDRs and the MVC is because the SDRs preceded the MVC, which means that any new approach needs to be associated with them. To do that, we split our research in two parts. Namely, (i) we discuss under what type of distributions the MVC coincides with a SDR, and (ii) we discuss under what type of preferences the MVC coincides with a SDR. In terms of (i), we deduce that the MVC coincides with the Second-order of Stochastic Dominance Rule (hereafter; SSDR) under the Elliptical family of distributions for any risk-averse investor. The idea with respect to the Elliptical family of distributions originates from Chamberlain (1983) [17]. However, contrary to Schuhmacher et al. (2021) [80] this is not the case for the Skew-Elliptical family of distributions. Specifically, our MC Simulations showcase that in the case of a Skew-Normal distribution there are risk-averse investors who will not use the MVC. In terms of (ii), we find that the assumption of quadratic utility function can only make the MVC sufficient for the maximization of the expected utility of the investor. With regards to the premise of Markowitz (1959) [65] concerning the approximately quadratic utility functions, we apply MC Simulations. The simulations enable us to test

Markowitz's premise under several different types of distribution. Our results point that the premise seems to be valid for Elliptical and Skew-Elliptical distributions but as we deviate more from normality the MVC fails.

Chapter 4, which is the last part of this thesis deals with a somewhat different subject. In this chapter, we focus on the risk aversion coefficients defined by Arrow and Pratt, known as ARA and RRA. Doing a thorough analysis of the literature we deduce that historically the ARA of investors has been decreasing while the RRA has been decreasing, constant or increasing for different time periods. We offer a new way to track the level of ARA and RRA with respect to the investor's wealth. More specifically, we derive a closed-form expression for the degree of RRA of a lottery with nonzero mean. This formula serves as an alternative way to extract information with respect to the level of ARA and RRA of investors using only the returns and market capitalization of a market index and the 10-year Treasury bills of the respective market. We argue that the value of this formula stems from its simplicity and its robustness. In fact, we test it under recursive and rolling-window estimations. Both ways, confirm the empirical findings of DARA. In terms of the slope of the RRA we derive different conclusions for different time periods. Finally, we introduce a simple way to measure the differences in portfolio diversification among different utility functions.

Chapter 2

The Performance Of The MVO Method Under Structural Breaks In Mean

2.1 Introduction

The theory of Von Neumann-Morgenstern advocates that a rational investor's objective is to maximize his expected utility. The literature suggests that, if the utility function has a quadratic form, or equivalently, if the portfolio returns follow a Gaussian distribution, the investor's expected utility becomes a function only of the two first moments of the portfolio returns. This result implies that the investor is only interested in his portfolio's expected return and volatility. Based on this idea Markowitz (1952, 1959) [64], [65] gave birth to the MVO problem. The MVO method is even to this day very intriguing. In fact, Markowitz in two of his very recent works, [66] and [63] (2010, 2014), clarifies that a large part of the literature is still confused regarding the necessary and sufficient conditions for the practical use of mean–variance analysis. Markowitz's portfolio theory is also celebrated for being the basis of a price formation theory for financial assets called the Capital Asset Pricing Model (hereafter; CAPM), developed by Sharpe (1964) [82] and some other researchers.

The high interest on the MVO method stems from its intuitive nature. Namely, all that the investor really needs to do is to either maximize his expected utility for a specific level of portfolio risk, or minimize his portfolio risk for a specific level of portfolio return. This simple optimization problem can be enriched with other constraints, like no short-selling, a budget constraint, etc.. It comes as no surprise that the MVO method has become a benchmark among portfolio choice methods mainly due to its simplicity. However, this came along with multiple researchers seeking to understand how the model works along with how robust it is.

The earliest part of the literature, focused mainly on understanding how the model works. Frankfurter et al. (1971) [28] design an experiment which indicates that even under normally distributed stock returns, the efficient set of three-asset portfolios varies substantially among different sample sizes. So, they conclude that selecting portfolios using the MV approach does not ensure better performance than a random choice of portfolios. Barry (1974) [10] and Bawa and Klein [12] showcased that the expected portfolio return remains the same whether or not the two input parameters, the mean of the assets μ and the variance-covariance matrix Σ are unknown. On the other hand, the expected portfolio risk increases if one of the two parameters is unknown and it gets even higher if both of them are unknown. Dickinson (1974) [25] argues that in order to estimate the portfolio risk one needs a very large sample. He also finds evidence that the MVO strategy is unstable for portfolios of 2 assets. In another work, Lai et al. (2011) [48] propose a Bayesian approach assuming that μ and Σ are unknown. In particular, the authors assume a prior distribution for expected returns and covariances and reformulate the Markowitz problem as a stochastic optimization problem.

The majority of research on the issues observed on the MVO method has to do with studying the effect of the input parameters on the MVO method. Jobson and Korkie (1980, 1981) [40], [39] were among the first to study the plug-in estimates of the mean and variance-covariance matrix of stock returns and point to them as the main sources of sensitivity in the MVO method's results. In their work, the two authors conclude that, under the assumption of a Gaussian distribution for the stock returns, the plug-in estimates lead to extreme errors in the weights, the portfolio return and the portfolio variance. Based on this evidence, Michaud (1989) [67] characterized the MVO method as an "*estimation-error maximizer*". From that point on, a plethora of research papers have studied the sensitivity of the MVO method. Best and Grauer (1991) [13] developed analytical expressions describing the sensitivity of the optimal portfolio to changes in the mean returns of the assets. Chopra and Ziemba (1993) [19] found out that the relative effect of means accounts significantly more for the error on the investor's utility compared to the relative error in variances and covariances. In another work from Chopra et al. (1993) [20], the authors investigate alternative ways for improving historical estimates, including the use of Stein estimators. They decide that using Stein's estimation method, the MVO derives higher portfolio returns with less risk than the plug-in estimation method. Broadie (1993) [14] highlights that using a larger sample for the estimation of the input parameters can lead to them being nonstationary. So, he argues that there is a tradeoff between estimation error and stationarity. Moreover, Broadie approves the argument of Chopra and Ziemba. Namely, that the effect of estimated mean returns is far greater on the MV approach than the estimated variance-covariance matrix. So, in his words: "*One recommendation for practitioners is to use historical data to estimate standard deviations and correlations but use a model to estimate mean returns*".

A more recent part of the literature has done extensive research on finding more efficient ways to estimate the variance-covariance matrix. Ledoit and Wolf (2001, 2003) [51], [50], were among the first ones to publish works on how to improve the variance-covariance estimator. Their main argument rested on the fact that "*when the number of stocks under consideration is large, especially relative to the number of historical return observations available (which is the usual case) the sample covariance matrix is estimated with a lot of error*", because it becomes singular, i.e. non-invertible. To fix this, Ledoit and Wolf used Stein's shrinkage method [84] on the sample variance-covariance matrix. Jagannathan and Ma (2003) [38] show that constraining the portfolio weights to take only non-negative values has a shrinkage-like effect on the covariance matrix estimate. More specifically, the large covariances that would otherwise imply negative weights can be shrunk substantially just by imposing a non-negativity constraint or even an upper-bound constraint. Kourtis et al. (2012) [46] argue that shrinking the variance-covariance matrix leads to an underperforming portfolio. So, they propose a new framework in which they directly estimate the inverse of the covariance matrix.

Another interesting part of the literature studies Markowitz under the assumption of non-normality of stock returns. Non-normality for daily stock returns has been generally accepted since the work of Mandelbrot (1963) [61]. Mandelbrot showed, empirically, that stock returns are leptokurtic and are part of the so-called Lévy stable (Paretian) family. Now, Jondeau and Rockinger (2006) [43] find evidence that for moderately non-normal distribution the MV approach works well, in that it approximates the maximization of expected utility under the framework of Von Neumann-Morgenstern. However, they also consider two more datasets which are increasingly non-normal and show that for such cases one needs higher moments, namely the skewness and kurtosis of stock returns, to get better approximations of the expected utility. Another interesting work is that of Xiong and Idzorek (2011) [89] who developed a different asset allocation approach based on the assumption of a truncated Lévy flight distribution. Once again, the importance of higher moments became evident though their work. Lastly, Jondeau and Rockinger (2012) [42] capture non-normality through GARCH-type models. These works with the exception of [43] deviate from the framework set by Markowitz.

An alternative approach for determining the effectiveness of the MV approach is to compare it against other portfolio building strategies. DeMiguel et al. (2009) [24] derived some theoretical results which they verified through MC Simulations, that showcased the inability of the MVO strategy to outperform $1/N$, out-of-sample. Pflug et al. (2012) [75] argued that one should prefer $1/N$ over Markowitz when the distribution of asset returns is ambiguous. The idea of comparing the MVO method directly to the simple "naive" strategy is very intuitive and for that reason we will follow it to make our conclusions with respect to the practicality of the MV approach.

Evidently, the literature has made a significant effort to address the effects originating from the plug-in estimates on Markowitz. However, it seems that we have reached to an impasse, which means that we need to revisit the building blocks of the MVO method and give careful answers to the following questions. Namely, (i) Is the MVO method truly impractical against $1/N$ under the assumption of normality? In other words, what is the necessary sample size T we need relative to the number of stocks N for the MVO method to outperform the naive rule? (ii) We will measure the rate of convergence of the estimated weights coming from the MVO method relative to T and the number of assets N. We expect Markowitz to work well under normality. We will see however that in order to make the right conclusions it is of paramount importance to compare $1/N$ with the MVO method under the same terms. Using MC Simulations, we will see that under normality Markowitz outperforms $1/N$ for relatively small samples. The results will lead naturally to the following question. (*iii*) If Markowitz's approach works well under normality then why does empirical evidence suggest that the method is extremely sensitive? We believe that the answer to this question lies in the DGP of stock returns. In other words, if an investor presumes that stock returns are normally distributed while in reality this is not the case it is more than expected that Markowitz's method will underperform.

In the literature, there is strong evidence that stock prices or returns exhibit regime shifts due to some drastic change in the economic environment, see for example [7], [31], [78] and [85]. The more significant the economic change the larger the shift in stock prices or returns. These papers identify only few large regime shifts. We will argue that daily stock returns are subject to multiple breaks in mean, mainly due to the fact that they are not i.i.d. in reality. We are going to provide empirical findings that justify that premise. These breaks happen abruptly at multiple points in time, due to various economic events or stock market reactions, pushing the DGP's means up or down. Moreover, these breaks might not be significantly large, but we will see that there is a large number of such breaks. In case the investor overlooks these breaks he will end up with misleading conclusions with regards to the usefulness of the MVO method. In fact, as the investor ignores more and more of those breaks he is going to estimate an inflated variance-covariance matrix, which will lead to higher estimated portfolio variances. Our aim is to study the performance of the MVO method against $1/N$ under multiple breaks in mean. Furthermore, we will also focus on how the "no break" case compares to the multiple break cases in terms of the MVO method's performance. The idea of testing the MVO method under a DGP with structural breaks in means has not yet been examined by the literature. Our results suggest that the presence of multiple breaks in means affects substantially the performance of Markowitz's method if they are ignored.

Our main findings split into two parts. (i) The MC Simulations showcase that the MVO strategy outperforms the $1/N$ strategy under the assumption of Gaussian DGPs, as long as we control the differences between the true MV Sharpe ratio and the true naive Sharpe ratio. (ii) Under DGPs with structural breaks in mean, the MC Simulations indicate that an investor that overlooks multiple breaks and proceeds with the assumption that the data is Gaussian, will estimate inflated covariances which will lead to the MVO method underperforming the $1/N$ case.

2.2 Theoretical Framework

Before anything it is crucial to analyze the statistical properties of the MVO method. We believe that both $1/N$ and MVO should be tested under the same terms, in order to draw safe conclusions. This can be achieved by setting the portfolio returns from each strategy to be equal. We will elaborate on that in the following sections. Now, the MVO problem is defined as shown below

$$\begin{aligned} &\min_{w} \; w'\Sigma w \\ &\text{subject to: } w'\mathbb{1} = 1 \\ &\qquad\qquad\quad w'\mu = \mu_r \end{aligned} \tag{2.1}$$

where μ_r represents the true portfolio return and

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_N^2 \end{bmatrix}.$$

represent the mean and variance-covariance matrix of N different stocks, respectively. Each stock $i = 1, \dots, N$ has a sample size of T stock returns. The solution of (2.1) is derived using Lagrange multipliers. So, the true weights w_r are expressed by

$$w_r = \left[\frac{\Gamma\mu_r - A}{\Delta}\right]\Sigma^{-1}\mu + \left[\frac{B - A\mu_r}{\Delta}\right]\Sigma^{-1}\mathbb{1} \tag{2.2}$$

where

$$A = \mathbb{1}'\Sigma^{-1}\mu, \quad B = \mu'\Sigma^{-1}\mu, \quad \Gamma = \mathbb{1}'\Sigma^{-1}\mathbb{1},$$
$$\Delta = B\Gamma - A^2$$

We observe that the weights produced are a non-linear function of μ, Σ and μ_r. These results are unknown to the investor but we need them to do the appropriate in-sample tests using MC Simulations in the following sections. So, in reality, he is obliged to estimate the input parameters μ and Σ in order to derive the estimated weights.

The most common way to estimate μ and Σ is to use the available historical data. Then, if we have N assets and T observations for each one of them our so called plug-in estimates $\hat{\mu}$ and $\hat{\Sigma}$ are expressed by

$$\widehat{\mu} = \begin{bmatrix} \widehat{\mu}_1 = \sum_{t=1}^{T} R_{1t} \\ \widehat{\mu}_2 = \sum_{t=1}^{T} R_{2t} \\ \vdots \\ \widehat{\mu}_N = \sum_{t=1}^{T} R_{Nt} \end{bmatrix}$$

and

$$\widehat{\Sigma} = \begin{bmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} & \cdots & \widehat{\sigma}_{1n} \\ \widehat{\sigma}_{21} & \widehat{\sigma}_2^2 & \cdots & \widehat{\sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\sigma}_{n1} & \widehat{\sigma}_{n2} & \cdots & \widehat{\sigma}_N^2 \end{bmatrix} =$$

$$\begin{bmatrix} \frac{1}{T-1}\sum_{t=1}^{T}(R_{1t}-\widehat{\mu}_1)^2 & \frac{1}{T-1}\sum_{t=1}^{T}(R_{1t}-\widehat{\mu}_1)(R_{2t}-\widehat{\mu}_2) & \cdots & \frac{1}{T-1}\sum_{t=1}^{T}(R_{1t}-\widehat{\mu}_1)(R_{Nt}-\widehat{\mu}_n) \\ \frac{1}{T-1}\sum_{t=1}^{T}(R_{1j}-\widehat{\mu}_1)(R_{2t}-\widehat{\mu}_2) & \frac{1}{T-1}\sum_{t=1}^{T}(R_{2t}-\widehat{\mu}_2)^2 & \cdots & \frac{1}{T-1}\sum_{t=1}^{T}(R_{2t}-\widehat{\mu}_2)(R_{Nt}-\widehat{\mu}_n) \\ \cdots & \cdots & & \cdots \\ \frac{1}{T-1}\sum_{t=1}^{T}(R_{1t}-\widehat{\mu}_1)(R_{Nt}-\widehat{\mu}_N) & \frac{1}{T-1}\sum_{t=1}^{T}(R_{2t}-\widehat{\mu}_2)(R_{Nt}-\widehat{\mu}_N) & \cdots & \frac{1}{T-1}\sum_{t=1}^{T}(R_{Nt}-\widehat{\mu}_{Nt})^2 \end{bmatrix}.$$

Now, if we make the additional assumption that the stock returns R_t follow an i.i.d. multivariate normal distribution.

$$\mathbf{R}_t = \begin{bmatrix} R_{1t} \\ R_{2t} \\ \vdots \\ R_{Nt} \end{bmatrix} \sim \mathbb{N}(\mu, \Sigma),$$

then, $\hat{\mu}$ and $\hat{\Sigma}$ are the best possible estimators of μ and Σ we can have. Meaning that they are both unbiased and consistent. So, in that case we have that

$$E[\widehat{\mu}] = \mu, \quad \widehat{\mu} \xrightarrow{P} \mu$$
$$E[\widehat{\Sigma}] = \Sigma, \quad \widehat{\Sigma} \xrightarrow{P} \Sigma$$

We recall from (2.2) that w is a function of μ, Σ and μ_r. So, the estimator of w_r is written as shown below.

$$\widehat{w} = \left[\frac{\widehat{\Gamma}\mu_r - \widehat{A}}{\widehat{\Delta}}\right]\widehat{\Sigma}^{-1}\widehat{\mu} + \left[\frac{\widehat{B} - \widehat{A}\mu_r}{\widehat{\Delta}}\right]\widehat{\Sigma}^{-1}\mathbb{1} \tag{2.3}$$

The next logical step is to ask ourselves if the properties of $\widehat{\mu}$ and $\widehat{\Sigma}$ apply also to $\widehat{w}$. For this we need to make use of Jensen's inequality.

Proposition 2.1. *(Jensen's inequality) Let X be an integrable random variable. Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a concave function such that it is also integrable. Then, the following inequality, called Jensen's inequality, holds:*

$$E[g(X)] \leq g(E[X]) \tag{2.4}$$

In our case, we have that $\widehat{w} = g(\widehat{\mu}, \widehat{\Sigma}, \mu_r)$, where g represents a non-linear function. This implies that under Proposition 2.1:

$$E[\widehat{w}] \neq w$$

So, although both $\widehat{\mu}$ and $\widehat{\Sigma}$ are unbiased estimators of the true μ and Σ this is not the case for $\widehat{w}$. Since this is the case one would wonder how big is the bias $E[\widehat{w}] - w$, as we increase the sample size T. To answer this question we have to examine whether or not $\widehat{w}$ is at least a consistent estimator. The answer comes from the Continuous Mapping Theorem.

Theorem 2.2. *(Continuous Mapping Theorem) Let X_n be a sequence of N-dimensional random vectors. Let $g : \mathbb{R}^N \mapsto \mathbb{R}^M$ be a continuous function, where N and M belong to the natural numbers. Then,*

$$X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X) \tag{2.5}$$

If we make use of Theorem 2.2 we see that

$$\left.\begin{array}{l}\widehat{\mu} \xrightarrow{P} \mu \\ \widehat{\Sigma} \xrightarrow{P} \Sigma\end{array}\right\} \Rightarrow \widehat{w} \xrightarrow{P} w$$

What this says is that for small samples T we expect a level of bias, which should decline as we increase T. The question is how significant will this level of bias be for small samples? That, we believe stems from (i) the level of bias in $\widehat{\mu}$ and $\widehat{\Sigma}$ and (ii) the level of non-linearity of $g(\widehat{\mu}, \widehat{\Sigma}, \mu_r)$. So, in small samples the level of bias in both $\widehat{\mu}$ and $\widehat{\Sigma}$ will be positive. However, it is of high importance to identify how much does the non-linear function g enhance this error and thus increases the bias in $\widehat{w}$. The best way to answer this question is to make use of the MC Simulations. We are going to evaluate the MVO method using the Sharpe ratio as a measure of performance. The "true" Sharpe ratio will be denoted by SR_r and it is defined as

$$SR_r = \frac{\mu_r}{\sigma_r},$$

where μ_r and σ_r represent the "true" target return and portfolio risk, respectively.

But first, we will review the work of DeMiguel et al. [24] in order to discuss some of their results.

2.2.1 An Overview on DeMiguel et al. 2009

In their seminal paper, DeMiguel et al. [24] showed that, out-of-sample, the Sharpe ratio of the $1/N$ strategy is higher than that of the sample-based mean-variance strategy. More specifically, the Sharpe ratio derived by maximizing the investor's utility,

$$\max_{w} w'\mu - \frac{\gamma}{2} w'\Sigma w$$

where γ denotes the investor's risk-aversion, is higher than that of the $1/N$ strategy, out-of-sample, only for very large window sizes. This evidence is highly interesting as it suggests that the simplest diversification strategy available is better to use. Their work is separated into the theoretical results and the ones that result from running MC Simulations.

The main attribute of the theoretical results is that they make use of the methodology developed from Kan and Zhou [44], which is based on maximizing the expected utility of the investor as shown above. Then, we need to define a variable that measures difference between the estimated Sharpe ratios and the ones that derive from the naive strategy. Namely, we set

$$L(w^*, \widehat{w}) = U(w^*) - E[U(\widehat{w})],$$

where w^* represents the optimal weight and $\widehat{w}$ represents the estimated weight.

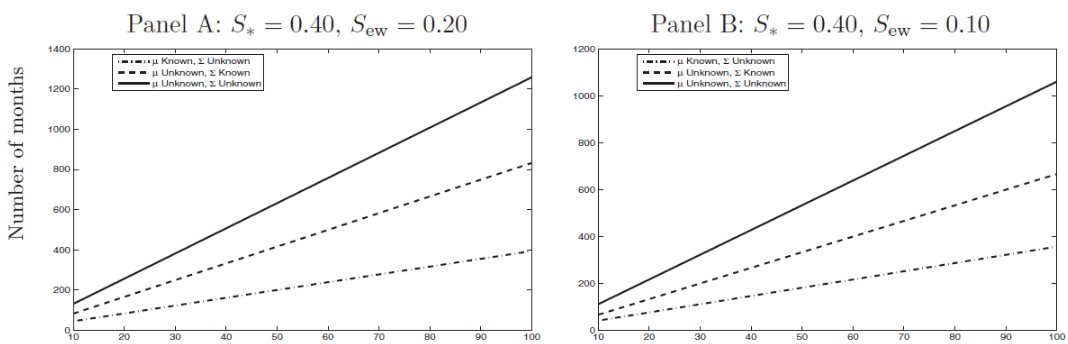
This measure is then used to derive the smallest window size M the MVO strategy needs in order to outperform $1/N$. The equation for this is shown below

$$\begin{aligned} M^* &= inf\{M : L(w^*, \hat{w}) < L(w^*, \frac{1}{N}\mathbb{1})\} \\ &= inf\{M : E[U(\hat{w})] > E[U(\frac{1}{N}\mathbb{1})]\} \\ &= inf\{M : E[U(\hat{w})] > U(\frac{1}{N}\mathbb{1})\} \end{aligned}$$

Now, the authors end up with a closed-form solution for determining when the MVO strategy outperforms the naive diversification[1].

$$\begin{aligned} & kSR_r^2 - SR_{1/N}^2 - h > 0 \qquad \text{with,} \\ & k = \frac{M}{M-N-2}\Big(2 - \frac{M(M-2)}{(M-N-1)(M-N-4)}\Big) < 1 \\ & h = \frac{MN(M-2)}{(M-N-1)(M-N-2)(M-N-4)} > 0 \end{aligned}$$

This criterion looks for the appropriate size of N and M the Markowitz method needs in order to outperform $1/N$. Using this criterion they form some examples. Two of them are shown below, taken from their paper. The figures indicate that even for very



large differences between the Sharpe ratios obtained by each strategy we need at least a

[1]For further details refer to DeMiguel et al. (2009) [24].

window size of $M \simeq 150$ months for $N = 10$ assets and $M \simeq 1050$ months for $N = 100$ assets. This theoretical evidence indicates that the MVO method looks very weak against $1/N$ even for very large differences between the Sharpe ratios. Moreover, their empirical analysis supported their theoretical results, by finding that $1/N$ outperforms MVO out-of-sample.

The authors then attempt to further support their theoretical findings by applying MC Simulations. In fact, they generate stock returns for $N = \{10, 25, 50\}$ stocks, $T = 24,000$ sample size and for window sizes $M = \{120, 360, 6000\}$, through

$$R_{i,t} = \beta_i R_t^{idx} + \epsilon_t,$$

where $\beta_i \sim \mathcal{U}(0.5, 1.5)$, $R_t^{idx} \sim \mathbb{N}(\mu_p, \sigma_p)$ and $\epsilon_t \sim \mathbb{N}(0, \sigma^2 \mathbb{1}_N)$. From this, we notice that the model that generates the stock returns is designed in a way that benefits the $1/N$ strategy. More specifically, the means derived from the generated returns as well as the variances will be very close to each other for each stock. So, applying these moments to the maximizing utility formula will lead to that are close to $1/N$. Evidently, this constitutes a very special case where the two strategies will obtain very similar Sharpe ratios. Numerically, they have $SR_r = 0.51$ and $SR_{1/N} = \{0.47, 0.50, 0.51\}$, according to $N = \{10, 25, 50\}$. So, it comes as no surprise that for just $N = 10$ assets we will need at least a window size of $M = 6000$ to outperform $1/N$.

So, although their empirical analysis found supporting evidence with respect to their theoretical findings it seems that the MC Simulations are designed in such a way that the strength of $1/N$ is overstated. Through the next pages, we will re-evaluate the two strategies, namely MVO and $1/N$, by applying MC Simulations. Our aim is to carry out our own calibration study so as to compare our findings with those of DeMiguel et al.. We will initiate our study under the assumption of stock returns following a multivariate normal distribution. Under this assumption the MVO method is expected to work properly, in that the only meaningful parameters are the mean and the variance-covariance matrix. So, if the evidence found in DeMiguel et al. is also confirmed by our study there is no point in using MVO, as there is a far more simple strategy to apply, that of spreading your wealth equally among the assets.

2.3 MVO Method Performance under Gaussian DGPs

2.3.1 Monte Carlo Simulations

The MC Simulations constitute an efficient way to do research on the statistical properties of the MVO method for various sample sizes. We have already proven that the sample weights $\widehat{w}$ are biased estimators of their real values, but they satisfy the property of consistency. Under the use of the MC Simulations we will specify the level of bias for both the input and output estimators as well as the rate at which it converges to zero as we increase the sample size. Concurrently, we will see how much the Markowitz portfolio is affected.

The MC Simulations will generate stock returns under an assumed distribution. Our assumption is that the stock returns follow a multivariate normal distribution with a mean vector μ and variance covariance matrix Σ, i.e. $\mathbb{N}(\mu, \Sigma)$. From now on, μ and Σ will be called the "true" parameters. Accordingly, the following results will represent the "true" MVO results.

$$\begin{aligned}
w &= \left[\frac{\Gamma\mu_r - A}{\Delta}\right]\Sigma^{-1}\mu + \left[\frac{B - A\mu_r}{\Delta}\right]\Sigma^{-1}\mathbb{1} \\
\mu_r &= w'\mu = \frac{1}{N}\mathbb{1}'\mu \\
\sigma_r^2 &= w'\Sigma w \\
SR_r &= \frac{\mu_r}{\sigma_r},
\end{aligned} \tag{2.6}$$

where N represents the number of assets. Observe that we have set the "true" target return to be exactly equal to the return obtained by the "naive" (or, equivalently $1/N$) strategy. The reason we set the same target return for both strategies is that we want to compare them on the same terms. The "true" naive results will be

$$\begin{aligned}
\mu_{1/N} &= \frac{1}{N}\mathbb{1}'\mu \\
\sigma_{1/N}^2 &= \frac{1}{N^2}\mathbb{1}'\Sigma\mathbb{1} \\
SR_{1/N} &= \frac{\mu_{1/N}}{\sigma_{1/N}},
\end{aligned} \tag{2.7}$$

Based on the "true" parameters we generate stock returns for each stock through a multivariate normal distribution $\mathbb{N}N(\mu, \Sigma)$ with a sample size of T. Thus, we obtain the sample mean and variance-covariance matrix $\widehat{\mu}$ and $\widehat{\Sigma}$. This procedure needs to be replicated multiple times so as to derive robust results. We denote the number of replications by S. So, for each asset $i = 1, 2, \dots, N$ we get

$$\begin{bmatrix} \mathbb{N}(\mu,\Sigma) & \Rightarrow & \{R_{i1}, R_{i2}, ..., R_{iT}\}_1 & \Rightarrow & \hat{\mu}_1, \hat{\Sigma}_1 \\ \mathbb{N}(\mu,\Sigma) & \Rightarrow & \{R_{i1}, R_{i2}, ..., R_{iT}\}_2 & \Rightarrow & \hat{\mu}_2, \hat{\Sigma}_2 \\ \dots & & & & \\ \mathbb{N}(\mu,\Sigma) & \Rightarrow & \{R_{i1}, R_{i2}, ..., R_{iT}\}_S & \Rightarrow & \hat{\mu}_S, \hat{\Sigma}_S \end{bmatrix}$$

Now, we turn to the next step, which is solving the MVO problem for each replication s, by applying the estimated parameters, $\hat{\mu}_s$ and $\hat{\Sigma}_s$. So, for each replication we solve the MVO problem, which derives the "fully estimated" results as shown below.

$$\begin{aligned} \hat{w}_s &= \left[\frac{\widehat{\Gamma}\mu_r - \widehat{A}}{\widehat{\Delta}}\right]\widehat{\Sigma}^{-1}\widehat{\mu} + \left[\frac{\widehat{B} - \widehat{A}\mu_r}{\widehat{\Delta}}\right]\widehat{\Sigma}^{-1}\mathbb{1} \\ \hat{\mu}_{e,s} &= \hat{w}_s'\widehat{\mu}_s = \mu_r \\ \hat{\sigma}^2_{e,s} &= \hat{w}_s'\widehat{\Sigma}_s\hat{w}_s \\ \widehat{SR}_{e,s} &= \frac{\hat{\mu}_{e,s}}{\hat{\sigma}_{e,s}}. \end{aligned} \tag{2.8}$$

The reader should observe that $\hat{\mu}_{e,s} = \mu_r$. This simply means that we solve MVO by setting the portfolio return to be always μ_r. This way we ensure that the estimated results are comparable to the "true" results. We will refer to (2.8) as the "fully" estimated solution. The "fully" estimated results will be the ones that are observed by the investor, since they are functions of his estimates. However, there is one additional interesting case as shown in Broadie (1993) [14]. There are also the "actual" results which represent the realized case, i.e. the case where the investor combines the estimated weights with the "true" parameters. More specifically,

$$\begin{aligned} \hat{w}_s &= \left[\frac{\widehat{\Gamma}\mu_r - \widehat{A}}{\widehat{\Delta}}\right]\widehat{\Sigma}^{-1}\widehat{\mu} + \left[\frac{\widehat{B} - \widehat{A}\mu_r}{\widehat{\Delta}}\right]\widehat{\Sigma}^{-1}\mathbb{1} \\ \hat{\mu}_{a,s} &= \hat{w}_s'\mu \\ \hat{\sigma}^2_{a,s} &= \hat{w}_s'\Sigma\hat{w}_s \\ \widehat{SR}_{a,s} &= \frac{\hat{\mu}_{a,s}}{\hat{\sigma}_{a,s}}. \end{aligned} \tag{2.9}$$

Evidently, the weights are estimated in the same way so they are denoted again by $\hat{w}_s$. However, the $\hat{\mu}_{a,s}$, $\hat{\sigma}_{a,s}$ and $\widehat{SR}_{a,s}$ will be different as we now use the "true" mean and variance-covariance matrix. Also, we should highlight that now the portfolio return is not equal to μ_r. Our focal point will be the "actual" case since it is more intuitive.

Now, the investor has to calculate the average of the "actual" and "fully" estimated results, over all replications. So, the overall results will be

$$\begin{aligned}
\hat{\mu}_e &= \mu_r & \bar{\hat{\mu}}_a &= \frac{1}{S}\sum_{s=1}^{S}\hat{\mu}_{a,s} \\
\bar{\hat{\sigma}}_e^2 &= \frac{1}{S}\sum_{s=1}^{S}\hat{\sigma}_{e,s}^2 & \bar{\hat{\sigma}}_a^2 &= \frac{1}{S}\sum_{s=1}^{S}\hat{\sigma}_{a,s}^2 \\
\overline{\widehat{SR}}_e &= \frac{1}{S}\sum_{s=1}^{S}\widehat{SR}_{e,s} & \overline{\widehat{SR}}_a &= \frac{1}{S}\sum_{s=1}^{S}\widehat{SR}_{a,s}
\end{aligned}$$

Now we get to the final steps. To simplify the notation we will denote the sample parameters by $\widehat{\theta}_s$ for each replication $s = 1, 2, \ldots, S$. For each replication we are going to calculate the mean, the variance, the bias and the Mean Squared Error (hereafter; MSE) of each $\widehat{\theta}_s$ over all replications. So, we have that

$$\begin{aligned}
E[\widehat{\theta}] &= \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_s \\
Var[\widehat{\theta}] &= \frac{1}{S}\sum_{s=1}^{S}\left[\widehat{\theta}_s - E[\widehat{\theta}]\right] \\
bias[\widehat{\theta}] &= E[\widehat{\theta}] - \theta \\
MSE[\widehat{\theta}] &= \frac{1}{S}\sum_{s=1}^{S}\left[\widehat{\theta}_s - \theta\right]^2
\end{aligned}$$

where $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_S)$. These measures will highlight the difference between the estimated input parameters and the estimated weights.

Now, as we have already said we will evaluate the quality of the MVO estimates following the rationale of DeMiguel et al. (2009) [24]. More specifically, we compare the MVO method with the simplest strategy, the so-called $1/N$ or "naive" strategy. Our aim, is to determine what size of sample is needed so as for the MVO method to outperform $1/N$. To do that, we will test for what sample size the following inequality holds.

$$SR_{1/N} < \overline{\widehat{SR}}_a$$

If the above inequality holds for any sample size T it basically signals that the MVO method is more efficient than the "naive". More importantly, the lower the necessary sample size T in order for the inequality to hold, the more useful will be the MVO method. Now, as $T \to \infty$, the above inequality holds, asymptotically. Namely, due to our analysis in the first section of this chapter we derive:

$$SR_{1/N} < \overline{\widehat{SR}_a} \overset{\text{CMT, } T\to\infty}{\to} SR_{1/N} < SR_r.$$

2.3.2 Rates of Convergence of $\hat{w}$

In this section, we want to focus on the estimated weights derived from the MVO method. More specifically, we would like to find a way to measure the speed at which the estimated weights, $\hat{w}$, converge to their true values, compared to that of the estimated parameters, $\hat{\mu}$ and $\hat{\sigma}^2$. Jensen's inequality showcased that the sample weights produced by the MVO strategy are biased but consistent estimators of the true weights. We expect that the non-linear relationship of the weights with the mean and variance-covariance matrix will play a major role in the efficacy of Markowitz's strategy.

In general, an optimization method's usefulness is characterized by each speed of convergence to the optimized value. For example, there are well-known methods like Newton's method, root-finding algorithms, etc., which try to distinguish the best candidate among a set of models. In our case, the idea behind the rate or speed of convergence is simple. If a method is fast it will need less iterations, or equivalently, smaller sample sizes to converge.

How should one define the rate or speed of convergence? Our approach to this subject starts by introducing a simple example. Consider an investor that is given two alternative strategies to choose from. The aim is to find the one that converges faster to 0. Let's say that these strategies are $1/1000n$ and $1/2^n$. We know that both of them converge to 0 as $n \to \infty$. The investor is not aware of the strategies. All that he knows is the result of the first iteration $n = 1$ of each strategy. Namely, the first strategy derives 0.001 while the second one derives 0.5. Let's set as a level of precision 10^{-5}. So, the investor could mistakenly choose the first strategy. The truth however is that $1/1000n$ needs $n = 100$ to satisfy the 10^{-5} level of precision while $1/2^n$ needs only $n = 17$. Evidently, $1/2^n$ is much faster than $1/1000n$. This should highlight the importance of the rate of convergence.

So, we need a specific way of identifying the rate of convergence as well as the type of convergence of a sequence. For this we are led to the following definition.

Definition 2.3. *Assume that a sequence x_n converges to x, with $x \in \mathbb{R}$. Let also,*

$$\lim_{n\to\infty} \frac{x_{n+1} - x}{(x_n - x)^p} = r. \tag{2.10}$$

Here, p defines the order of convergence and r the speed or rate of convergence.

We also need the following proposition to compare orders and rates of convergence.

Definition 2.4.

(i)If $p = 1$ and $r = 1$ then the convergence is called sub-linear

(ii)If $p = 1$ and $0 < r < 1$ then the convergence is called linear

(iii)If $p = 1$ and $r = 0$ then the convergence is called super-linear

(iv)If $p > 2$ and $r > 0$ then the convergence is of p^{th}-order

Therefore, the order of convergence specifies in what way a sequence converges while the rate determines how quickly the same sequence converges to its limit. With respect to the initial example, under Definition 2.4, we can easily observe that $1/1000n$ converges to 0 sub-linearly while $1/2^n$ converges to 0 linearly, and thus, as shown above, will need a lower number of iterations to converge.

In our case, instead of focusing directly on the estimated parameters, $\hat{w}, \hat{\mu}$ and $\hat{\sigma}^2$, we focus on their variances over the number of replications. The idea behind this is to determine how quickly does the variance of each estimator converge to zero. We start by identifying the rates of convergence of the variance of the sample mean $Var(\widehat{\mu}_i)$ and sample variances $Var(\widehat{\sigma}_i^2)$, for $i = 1, \ldots, N$. Since both estimators are unbiased and consistent we can derive the closed-form expressions. So, the variance of the sample mean over all replications will be

$$Var(\widehat{\mu}_i) = \frac{1}{S}\sum_{s=1}^{S} Var(\widehat{\mu}_{is}) = \frac{S}{TS}\sigma_i^2 = \frac{1}{T}\sigma_i^2.$$

Now, in order to derive the formula for the variance of the sample variance we use the fact that the variance-covariance matrix $\widehat{\Sigma}$ follows a Wishart distribution with $T-1$ degrees of freedom. So, $Var(\widehat{\Sigma}) = \frac{1}{T-1}(\sigma_{ij} + \sigma_i^2\sigma_j^2)$, for $i, j = 1, 2, \ldots, N$. More specifically, we deduce that over all replications we get

$$\begin{aligned} Var(\widehat{\sigma}_i^2) &= \frac{2\sigma_i^4}{(T-1)} \\ Var(\widehat{\sigma}_{ij}) &= \frac{1}{(T-1)}(\sigma_{ij} + \sigma_i^2\sigma_j^2) \end{aligned}$$

So, we can now derive the rates of convergence. Under (2.10) we get

$$\begin{aligned} \frac{Var(\widehat{\mu}_{iT+100})}{Var(\widehat{\mu}_{iT})} &= \frac{\frac{\sigma_i^2}{(T+100)}}{\frac{\sigma_i^2}{T}} = \frac{T}{T+100} \overset{T\to\infty}{\to} 1 \\ \frac{Var(\widehat{\sigma}_{iT+100}^2)}{Var(\widehat{\sigma}_{iT}^2))} &= \frac{Var(\widehat{\sigma}_{ijT+100})}{Var(\widehat{\sigma}_{ijT}))} = \frac{T+99}{T} \overset{T\to\infty}{\to} 1 \end{aligned}$$

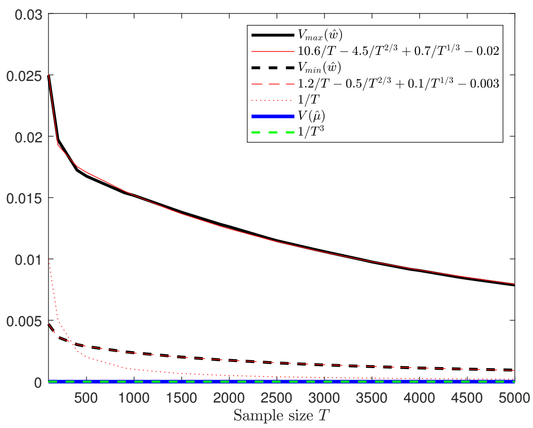
Thus, according to Definition 2.4 the variances of the input sample estimates converge sub-linearly to zero. The next step is to investigate the rate of convergence of the output estimates, i.e. the estimated weights. However, in this case we cannot derive a closed-form expression as we did for the input parameters. Referring to (2.3) we see that the parameters are dependent with each other. So, we choose to approach this case numerically. We start by a sample size $T = 100$ and reach up to $T = 5000$ with a 100 size of step. This means that we have a sequence of 200 variances of each estimated weight and for each replication. Using this method we conclude that

$$\frac{Var(\widehat{w}_{T+100})}{Var(\widehat{w}_T)} \overset{T\to\infty}{\to} 1.$$

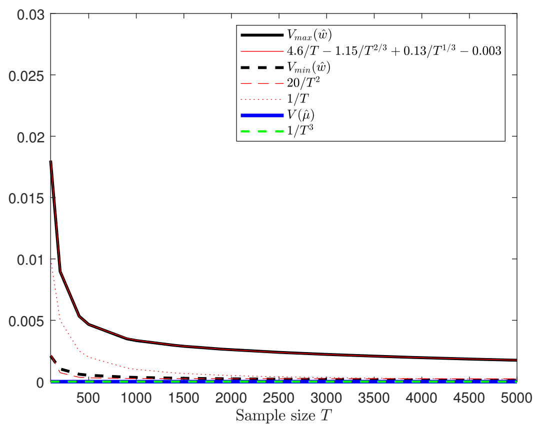
This result holds for any number of assets N. This was expected since we know that the estimated weights are consistent estimators. So, we deduce that the rate of convergence of the variance of the sample weights over all replications is also sub-linear. Therefore, we need to find a way of identifying which rate of convergence is the faster among the input and output sample estimates. We know for example that $1/n$ and $1/n^2$ both converge sub-linearly to zero but with a different rate. As a result, the simplest thing to do is to plot the variances of the estimated weights against the variances of the estimated means as well as against two known sequences, $1/T$ and $1/T^3$. In addition to that we can fit the variances of the estimated weights by polynomials to have a more meaningful comparison. We will plot assets with the maximum and minimum variance in the weight.

The data we shall use is 300 daily stocks from S&P 500 spanning from January 2010 to December 2019. Through several tests we have observed that we can simply set some random values for the Sharpe ratio SR_r, the target return μ_r and the difference between the target return and the return we get from the GMV, μ_r/μ_{GMV}. Controlling the distance between μ_r and μGMV proves to be crucial as it determines the level of impact the estimated means will have on the estimated weights. Namely, the further away we get from the GMV, the more impact the estimated means will have on the MVO method. Our main focus is how the rate of convergence of the variances of the estimated weights behaves after increasing the number of assets N.

Now, we first start by finding 100 different combinations of $N = \{10, 25\}$ assets that have the same μ_r, SR_r and μ_r/μ_{GMV}. For each case we will run $S = 10,000$ MC Simulations and keep track of the variances of the estimated parameters, with respect to the sample size T. Below, we present the plotted variances of the sample weights against the variances of the sample means, the $1/T$ and the polynomials fitted on the variances of the sample weights.



(a) $N = 10$, $SR_r = 1.2$, $\mu_r = 15\%$, $\frac{\mu_r}{\mu_{GMV}} = 1.5$



(b) $N = 25$, $SR_r = 1.2$, $\mu_r = 15\%$, $\frac{\mu_r}{\mu_{GMV}} = 1.5$

Looking at the graphs we observe that the variances of the estimated weights for $N = 25$ are drastically better compared to the $N = 10$ case. In fact, we see that for $N = 25$ assets, the maximum variance, $V_{max}(\hat{w})$, moves along with $1/T$ while the minimum variance, $V_{min}(\hat{w})$, becomes almost equal to $1/T^2$. Initially, this may look like a paradox, since we expect that the model will deliver more sensitive results as we increase the number of assets. However, we should not confuse the sensitivity of the overall result of the model with the size of error in the estimated weights. As we increase the number of assets to $N = 25$ the mean as well as the variance of the real weights drops significantly compared to the $N = 10$ case. Now, as we raise the number of assets we have to deal with the estimation of more weights which means that from that perspective the problem of optimization becomes more complicated, hence more sensitive overall.

Overall, we conclude that as we raise the number of assets N the variance of the estimated weights improves. Also, when we use daily data the variance of the estimated means is almost fitted by $1/T^3$. These findings lead us to the central and more practical question. What T do we need according to N in order for MVO to outperform $1/N$, i.e.

$$\overline{\widehat{SR}}_a > SR_{1/N}.$$

2.3.3 Calibration Results

Based on the work from DeMiguel et al. [24], we argued that it is crucial for the investor to know exactly the difference between the Sharpe ratios of the MVO and $1/N$ strategies together with the difference between the target return and the return he would get from the GMV,

$$\frac{SR_r}{SR_{1/N}} \text{ and } \frac{\mu_r}{\mu_{GMV}}.$$

So, we need to adjust the values for the above criteria, so as to be able to deduce how much does each one affect our final decision. This is a rather complex and demanding computational problem. For that reason we will do the tests for $N = \{10, 25\}$ assets. Below, we present the table consisting of the different cases we will examine. The interested reader can refer to Appendix A 2.6, at the end of this chapter, in which we cover 12 more Criteria for $N = \{10, 15, 20, 25, 30, 35, 40\}$ assets.

Table 2.1: Criteria values

$\frac{SR_r}{SR_{1/N}} = 1.2, \frac{\mu_r}{\mu_{GMV}} = 1.5$	$\frac{SR_r}{SR_{1/N}} = 1.2, \frac{\mu_r}{\mu_{GMV}} = 1.75$
$\frac{SR_r}{SR_{1/N}} = 1.3, \frac{\mu_r}{\mu_{GMV}} = 1.5$	$\frac{SR_r}{SR_{1/N}} = 1.3, \frac{\mu_r}{\mu_{GMV}} = 1.75$

For each criterion, we collect 2000 different combinations of $N = \{10, 25\}$ assets that satisfy it. Then, using $S = 10,000$ MC Simulations we determine the minimum sample size T for which the "actual" MVO Sharpe ratios outperform the $1/N$ Sharpe ratios, for each case of Table 2.1. Below, we can see the tables with the results for each case.

Table 2.2: $\frac{SR_r}{SR_{1/N}} = 1.2$ and $\frac{\mu_r}{\mu_{GMV}} = 1.5$

	Sample size			
	min	max	**mean**	std
$N = 10$	1000	3200	**2012**	408
$N = 25$	1500	4400	**2405**	437

Table 2.3: $\frac{SR_r}{SR_{1/N}} = 1.2$ and $\frac{\mu_r}{\mu_{GMV}} = 1.75$

	Sample size			
	min	max	**mean**	std
$N = 10$	1500	3700	**2478**	427
$N = 25$	2100	6600	**3435**	617

The first two Tables 2.2, 2.3 showcase the importance with regards to the number of assets together with the difference between the portfolio returns. More specifically, in Table 2.2 we see an increase of 20% in the average sample size needed to trust MVO over $1/N$. Accordingly, increasing the portfolio returns difference to $\mu_r/\mu_{GMV} = 1.75$ translates into a 39% raise in the mean sample size. As a result, we can deduce that controlling for the portfolio returns difference will surely impact the final decision of the investor in terms of the model he should use. This result stems from our choice of model, that accounts for the target return. The next two cases are covered in Tables 2.4, 2.5 shown below.

Evidently, the magnitude of difference between the Sharpe ratios impacts the final results drastically. Namely, from Table 2.2 to 2.4 we see a drop of 55% and 48% for $N = \{10, 25\}$, respectively. With respect to Tables 2.3 and 2.5 the drop is around 55%

Table 2.4: $\frac{SR_r}{SR_{1/N}} = 1.3$ and $\frac{\mu_r}{\mu_{GMV}} = 1.5$

	Sample size			
	min	max	**mean**	std
$N = 10$	500	1300	**907**	143
$N = 25$	800	1900	**1146**	173

Table 2.5: $\frac{SR_r}{SR_{1/N}} = 1.3$ and $\frac{\mu_r}{\mu_{GMV}} = 1.75$

	Sample size			
	min	max	**mean**	std
$N = 10$	1000	1900	**1338**	208
$N = 25$	1200	3200	**1886**	312

for both $N = \{10, 25\}$. This showcases the importance of controlling for the difference between the Sharpe ratios. Moreover, our findings contradict the theoretical results of DeMiguel et al. [24]. In fact, since the mean sample size T needed to outperform $1/N$ plummets due to such a small increase in the difference between the Sharpe ratios, we may infer that in more extreme cases the MVO strategy will outperform $1/N$ for significantly smaller samples. So, overall, the results from this section indicate that Markowitz's method works well under normality.

2.4 MVO under Structural Breaks in Mean

As the results indicate in the previous section, Markowitz's method functions well under the ideal assumption of normally distributed data. In fact, we conclude that the non-linearity of weights as well as the estimators of the input parameters should not lead the investor to reject the MVO method. More importantly, contrary to DeMiguel et al., we argue that the MVO method outperforms the simple naive diversification strategy for large sample sizes, even in cases of including a higher number of assets in our portfolio and the data is not i.i.d.. This however, depends on the investor's awareness of the differences between the Sharpe ratios, the portfolio returns and the number of assets.

Now, consider the following case. Two assets exhibit a number of b structural breaks in their means at the same point in time and in the same direction.

$$\begin{aligned}
&R_t \sim \mathbb{N}(\mu_1, \Sigma) \quad \text{for } 1 \leq t < \tau_1 \\
&R_t \sim \mathbb{N}(\mu_2, \Sigma) \quad \text{for } \tau_1 \leq t < \tau_2 \\
&\vdots \\
&R_t \sim \mathbb{N}(\mu_{b+1}, \Sigma) \quad \text{for } \tau_{b+1} \leq t \leq T
\end{aligned}$$

with $\tau_1, \tau_2, \ldots, \tau_{b+1}$ representing the points at which the breaks in mean take place. In our framework, the variance-covariance matrix, Σ, remains the same. Assume also that the investor ignores the breaks and proceeds with estimating the means and variance-covariance matrix of the two assets over the entire sample size. Thus, based on his estimation he believes that

$$R_t \sim \mathbb{N}(\hat{\mu}, \widehat{\Sigma}) \quad \text{for } 1 \leq t \leq T.$$

Subsequently, he uses the estimated parameters, $\hat{\mu}, \widehat{\Sigma}$, to build a portfolio by applying the MVO method. How would the results of the MVO method be affected? Intuitively, ignoring one or more breakpoints should result in an erroneous positive correlation between the two assets. Moreover, as the number of breaks gets larger the correlation should increase accordingly. Some numerical examples will help us see that.

Assume that the starting parameters of the two assets are the following.

$$\mu = (0.05, 0.07), \quad \Sigma = \begin{pmatrix} 0.7 & 0 \\ 0 & 1 \end{pmatrix}.$$

Initially, we generate returns with $B = 50$ breaks in mean which take place at the same point in time for both assets and in the same direction. The magnitude of each break lies somewhere inside $[-50\%, +50\%]$. Then, we calculate the correlation between the two assets and plot the returns in order to see if the positive correlation is apparent. We then do the same for $B = 200$ breaks. The choice for the number of breaks is not random. We will see later that these numbers of breaks are justified through real data. Both cases, can be found in the following Figures.

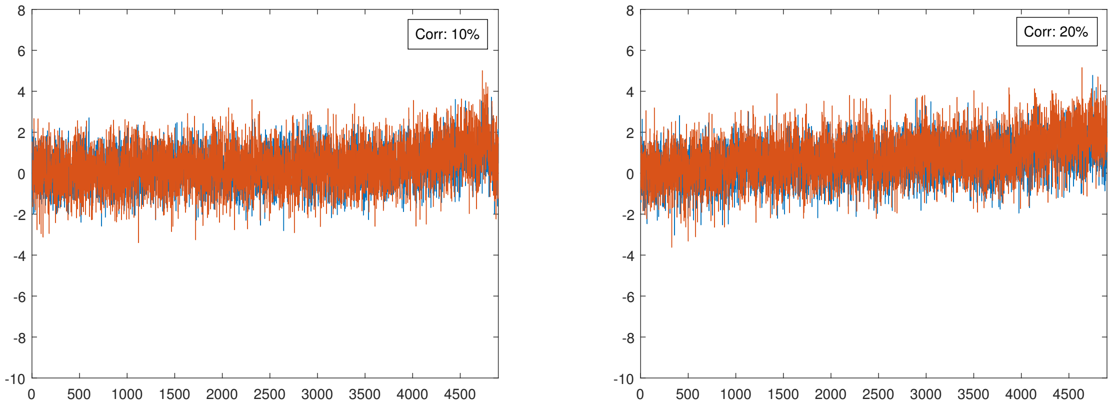


Figure 2.2: $N = 2$ assets with 50 breaks

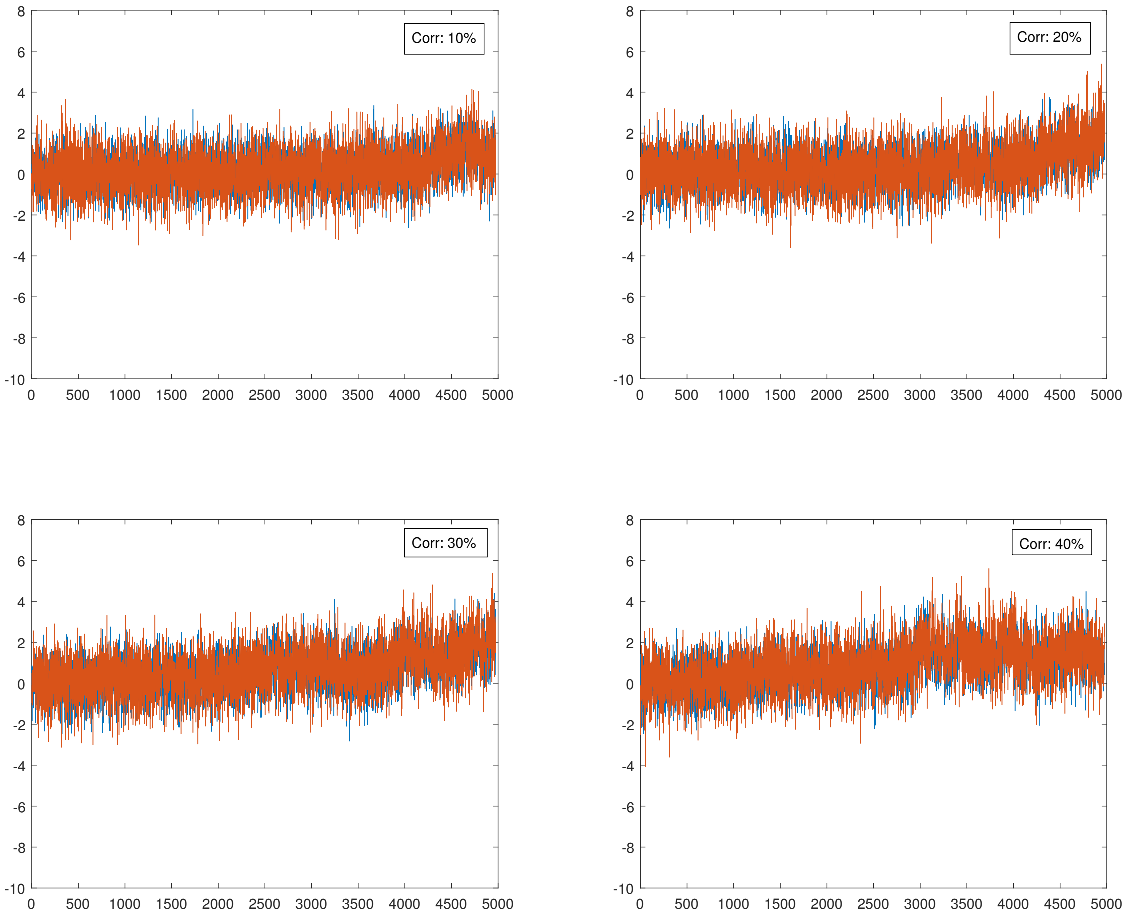


Figure 2.3: $N = 2$ assets with 200 breaks

Figure 2.2 indicates that $B = 50$ breaks in mean can result in estimating either a 10% or a 20% positive correlation if ignored. Furthermore, there is no apparent evidence of breaks in the two graphs. As Figure 2.3 shows, when the number of breaks increases to 200 we may estimate even higher positive correlations, namely, 30% and 40%, by mistakenly estimating μ and Σ using the entire sample size. So, as we overlook more and more breaks in mean our estimations with respect to Σ will get even worse. Obviously, supplying the MVO method with wrong estimations will lead to wrong decisions.

The previous examples, give as an idea of how important it is to determine whether or not the Data Generating Process (DGP) under assumption exhibits structural breaks. There is a large part of the literature on DGPs with structural breaks (equivalently, regime shifts) in their parameters. Chu et al. (1996) [22] find evidence of six regimes with significantly different volatility, using a Markov switching model to market returns. Pesaran and Timmermann (2004) [73] argue that large breaks in means or variances can lead to biased and inconsistent forecasts. Rapach and Wohar (2006) [78] find strong evidence of structural breaks in five bivariate predictive regression models of S&P 500 returns. We will argue that stock returns, and especially daily stock returns, exhibit multiple structural breaks in mean. This happens mainly because daily stock returns are not i.i.d. in reality. Ignoring those breaks will lead us using the entire sample to do our estimates. Thus, we will end up misestimating μ and Σ. This way, we might deduce that the $1/N$ strategy outperforms the MVO method, which will be a wrong inference. The current literature has not made yet any effort to study the link between structural breaks and their effects on the MVO method.

2.4.1 Methods for Identifying Structural Breaks in Mean

In this section, we are going to review the various methods for identifying shifts in means. Chow (1960) [21], developed the well-known Chow test for determining whether the true coefficients in two linear regressions on different data sets are equal. Assuming that we have evidence that a break point in mean takes place at a known period, τ, we can use the Chow test on two intervals, namely, $[1, \ldots, \tau]$ and $[\tau + 1, \ldots, T]$. A similar approach for detecting a single change-point in the mean of independent identically distributed random variables was introduced by Sen and Srivastava (1975) [81]. Brown, Durbin and Evans (1975) [15] developed Cusum and Cusum Squared tests to handle cases in which the time of the break is unknown. Later on, Andrews (1993) [3] tested for structural changes in parametric models with unknown change point, considering Wald, Lagrange multiplier, and likelihood ratio-like tests. An important part of the literature extended these tests to allow for multiple breaks. Bai and Perron (1998, 2002) [5] and [6] were among the first ones who considered cases related to multiple structural changes with unknown dates. Still, the number of breaks expected to be found has to be predetermined. Lavielle and Moulines (2000) [49] use penalized least-squares to estimate the number of known change-points in the presence of long memory in the error process. Pesaran et al. (2006) [72] introduce a hidden Markov chain approach to model the meta distribution of the parameters of the stochastic process underlying structural breaks. A similar approach is found in [74].

In our work, we will use the approach introduced by Killick et al. (2012) [45] which is similar to Lavielle and Moulines (2000) [49]. The test developed by Killick et al., detects change points by minimising a cost function over all possible numbers and locations of change points. The test can identify multiple structural breaks in mean or variance at unknown periods. The main reasons we use Killick's et al. approach are: (i) it is the only method that does not require to predetermine the number of breaks, and (ii) we do not need to specify the dates at which the breaks happen.

So, we define $R_{i,1:T} = (R_{i1}, R_{iT}, \dots, R_{iT})$ for $i = 1, 2, \dots, N$. Also, the number of changepoints is represented by B and their positions by $\tau_{1:B} = (\tau_1, \tau_2, \dots, \tau_B)$. We further assume that the possible breakpoints are between 1 and $T-1$ inclusive and are ordered, meaning that $\tau_i < \tau_j$ if and only if $i < j$. Consequently, the B different breakpoints split the data into $B+1$ segments, with the ith segment containing $R_{\tau_{i-1}+1:\tau_i}$.

Now, in order to identify multiple breaks in mean Killick et al. argue that we need to minimize a function of the following form

$$\sum_{b=1}^{B+1} [\mathcal{C}(R_{\tau_{j-1}+1:\tau_j})] + \beta f(B),$$

where $\mathcal{C}$ represents the cost function of a segment and $\beta f(B)$ is a penalty term. In our case, $\beta = 0$. This means that we are not looking to identify only the most significant large shifts in mean that happen due to high-impact events like the pandemic or the financial crisis of 2008. In reality, we look for multiple breaks in mean that might be small in value and happen due to low-impact events. The reason we expect to find multiple breaks in mean is because stock returns are not i.i.d. in reality. Following that, we expect to find out that if we ignore a large number of common breaks in mean for a set of N assets we will end up with large estimates of the correlations between the stocks. This idea will become more clear as we carry on.

Since we are interested in breaks in mean, the cost function $\mathcal{C}$ is actually the sum of squared residuals (SSR) from the "best" horizontal level for each segment. So, if we ask for a unique break in mean the method will search for the point τ where the total SSR is minimized. In mathematical terms, we need to minimize

$$\begin{aligned}\mathcal{C} &= \sum_{t=1}^{\tau-1}(R_{1t} - \frac{1}{\tau-1}\sum_{t=1}^{\tau-1} R_{1t})^2 + \sum_{t=\tau}^{T}(R_{1t} - \frac{1}{T-\tau+1}\sum_{t=\tau}^{T} R_{1t})^2 \\ &= (\tau-1)Var(R_1, \dots, R_{\tau-1}) + (T-\tau+1)Var(R_{1\tau}, \dots, R_T),\end{aligned}$$

for each asset and each breakpoint at time τ. This kind of approach for detecting multiple shifts in mean is very intuitive and easy to replicate.

As we have already highlighted, we are only interested in identifying the common breaks of a portfolio of N assets with the same direction. So, we first use the method to identify the breaks in mean for each asset. Then, we keep only the common breaks of a specific set of assets N with the same direction. Doing that, we want to showcase the fact that if we ignore the multiple common breaks of the assets, we will end up misestimating the correlations. More specifically, for sets of N assets with a higher number of common breaks in mean, we expect to derive higher estimated values of the correlations between the assets over the entire sample. So, the main idea is that we ignore the multiple common breaks in mean and we proceed with estimating the correlations between the stocks using the entire sample as if it was i.i.d.. Thus, we expect to derive a more "inflated" correlation matrix as we increase the number of common breaks. In the following section, we will see that this is truly the case.

Before applying the method of Killick et al., we remove the GARCH effect from our data. According to Mandelbrot (1963) [61], an important empirical regularity of stock returns is known as "volatility clustering", which argues that large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. This might result in periods of extreme levels of volatility which should not be considered as breakpoint segments. Thus, we need to adjust the stock returns accordingly. A common way to capture this empirical regularity, is by assuming that the stock returns are produced by a GARCH model. In practice, we use a GARCH(1,1) model to estimate the conditional variances as shown below.

$$\begin{aligned}
&R_t = \epsilon_t \\
&\epsilon_t = \sigma_t z_t, \quad z_t \sim \mathbb{N}(0,1) \\
&\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2 + \beta\epsilon_{t-1}^2
\end{aligned}$$

Then, we divide the stock returns, R_t, by the estimated conditional variances σ_t. This way, we remove the GARCH effect from our data.

$$R_t^{adj} = \frac{R_t}{\sigma_t}.$$

So, we apply Killick's et al. test on the adjusted stock returns. Following that, we search for the common breaks in mean for a portfolio of N assets which move in the same direction.

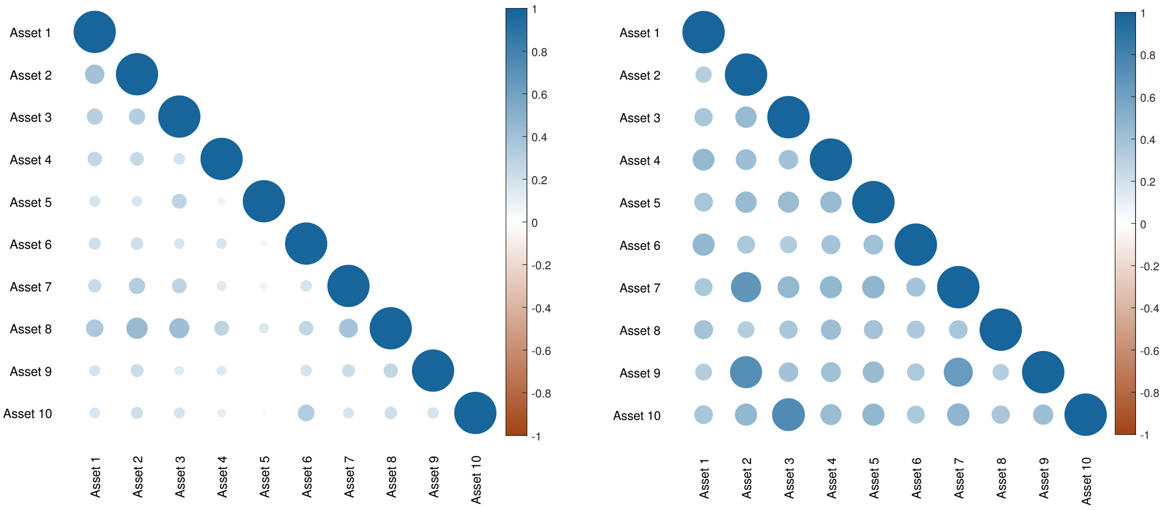
2.4.2 Empirical Motivation

We collect daily stock returns for the constituents of the following indices, FTSE 100, STOXX 600 and S&P 500. Our data is spanning from 2000 to 2021. The reason we collect data from different markets is to determine whether or not our conclusions for S&P 500 are consistent for the other markets too. Our main goal, is to examine Markowitz's MVO method when the underlying process exhibits structural breaks in mean and the investor ignores them. So, first, we collect those assets that exhibit breaks in the same direction. In other words, the assets display either a concurrent negative shift in their means or a concurrent positive shift at the same point in time. Then, we build portfolios consisting of $N = 10$ assets which exhibit $B = \{50, 100, 150, 200\}$ common breaks.

For each portfolio, we set the annualized target return to be $\mu_r = 15\%$ (the conclusions made are the same when checking for different levels of target return). We then solve the classic form of the MVO optimization problem, as shown below.

$$\min_{w} w'\Sigma w$$
$$\text{s.t. } w'e = 1$$
$$w'\mu = \mu_r$$

The variables μ and Σ are estimated using the entire sample. This implies that the investors overlook the multiple breaks that took place through the years. We do this for approximately 100 different portfolios of $N = 10$ assets, for each break case. Thus, we are certain that our findings are robust. Evidently, the only factor that varies among the different numbers of breaks, is the portfolio risk σ_r. Below, we present the correlation matrix of two representative portfolios of $N = 10$ assets, for 50 and 200 common breaks in mean, respectively.



(a) 50 common breaks (b) 200 common breaks

Figure 2.4: Correlation matrices of $N = 10$ assets from S&P 500

The remaining portfolios of $N = 10$ assets, for each break case, derive very similar results to those illustrated in Figure 2.4. Apparently, the correlations considering the 50 breaks case are mostly around 20% or lower, while in the 200 breaks case most of the correlations are around 40% or higher. This indicates that the higher the number of breaks that the investor ignores, the higher will be the values of the estimated correlations. Since this is the case for each portfolio of $N = 10$ assets and for each break case, we have sufficient evidence that there exist multiple breaks in mean in real stock returns and, more importantly, they affect the correlations between the stocks. We argue that this will affect both the estimated portfolio risk and the estimated Sharpe ratio from the MVO method and, by extension, the investor's decision making. This is showcased in the following graphs.

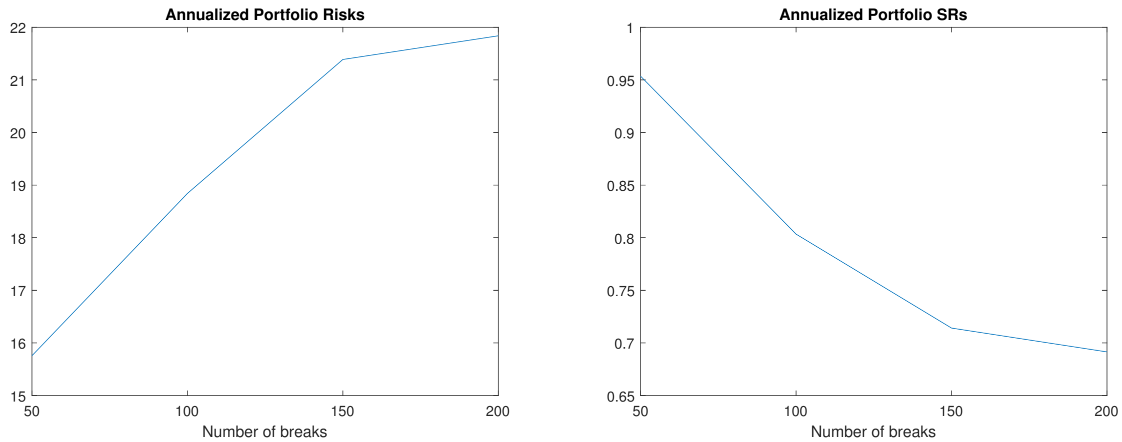
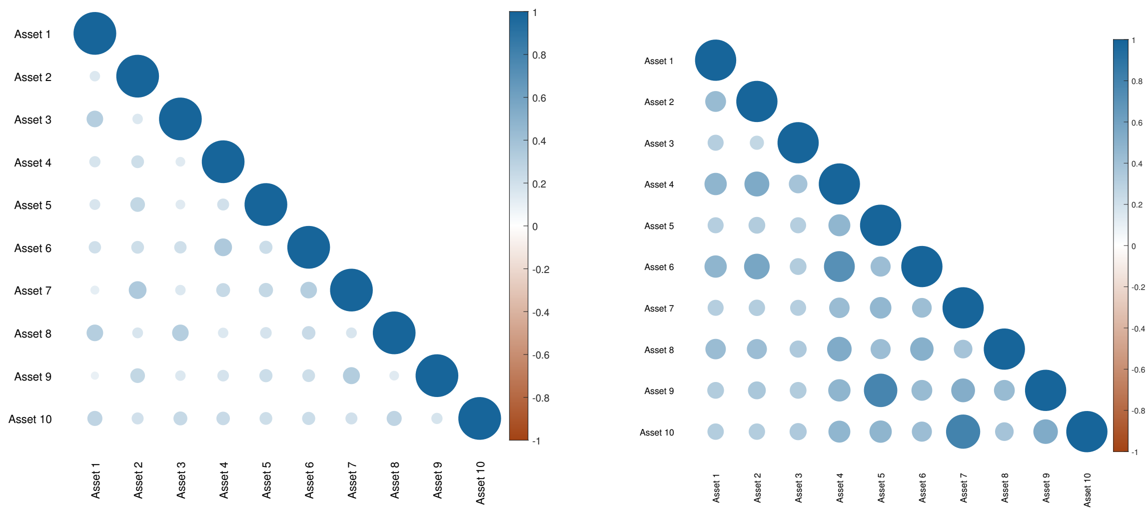


Figure 2.5: Portfolio risks and Sharpe ratios of $N = 10$ assets from S&P 500

Figure 2.5 illustrates the impact of structural breaks on the estimated portfolio risk and by extension on the estimated Sharpe ratios. In fact, we see that the annualized portfolio risk is approximately 38% higher in the 200 breaks case compared to the 50 breaks case. Likewise, the annualized Sharpe ratio drops by almost 26%. From that, we deduce that a DGP exhibiting multiple common breaks in mean can alter our portfolio allocation decisions by a considerable amount.

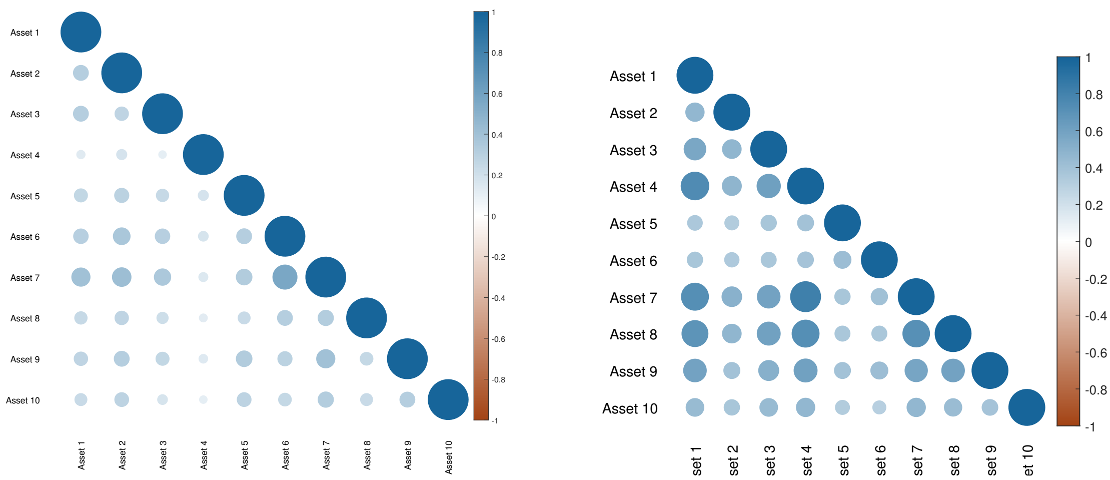
What is interesting, is that the same conclusions can be made also for the European market. Namely, the 200 breaks case for both FTSE 100 and STOXX 600 contains significantly correlated assets compared to the 50 breaks case for 50 different portfolios of $N = 10$ assets. This is shown in the following figures.



(a) 50 common breaks

(b) 200 common breaks

Figure 2.6: Correlation matrices of $N = 10$ assets from FTSE 100



(a) 50 common breaks

(b) 200 common breaks

Figure 2.7: Correlation matrices of $N = 10$ assets from STOXX 600

Our empirical findings, justify the use of MC Simulations on the MVO method for DGPs with multiple common breaks in mean. The main difference is that we are going to replace the assumption of a multivariate normal DGP with that of a DGP exhibiting multiple common breaks in mean. The simulations will help us make a one-to-one comparison between the MVO method and $1/N$ under the no break case and the multiple breaks cases.

2.4.3 MC Simulations

In this framework, we will consider the "no break" case, i.e. $B = 0$, in which the DGP is a multivariate NIID, as well as the "break" cases with $B = \{50, 100, 150, 200\}$. In the "no break" case, the investor sill simply need to use the whole sample to estimate the parameters and solve the MVO problem. As in the Gaussian case the "true" parameters will be denoted by μ and Σ. The only extra assumption in this case is that we assume that Σ has also zero covariances. The main idea behind this premise is that the non-zero covariances we see in a group of assets are mainly due to the effect of multiple common breaks in mean. We illustrated some interesting examples showing that in the previous section. We further assume that the investor sets the same target return for the MVO strategy to be exactly the same as the return obtained from the $1/N$ strategy. Thus, the "true" MVO and $1/N$ results with no breaks, i.e. $B = 0$, will be

$$\begin{aligned} \mu^0_{1/N} &= \mu^0_r = \frac{1}{N}\mathbb{1}'\mu \\ \sigma^0_{1/N} &= \frac{1}{N}\sqrt{\mathbb{1}'\Sigma\mathbb{1}} \\ SR^0_{1/N} &= \frac{\mu^0_{1/N}}{\sigma^0_{1/N}} \end{aligned} \qquad \begin{aligned} & w(\mu, \Sigma, \mu^0_r) \\ \mu^0_r &= \mu^0_{1/N} = w'\mu \\ \sigma^0_r &= \sqrt{w'\Sigma w} \\ SR^0_r &= \frac{\mu^0_r}{\sigma^0_r} \end{aligned}$$

We proceed with the estimation of the "actual" results using MC Simulations. The "actual" results for the "no break" case are estimated replicating the MC methodology for the Gaussian case. Thus, we derive the same results as in (2.9). Finally, we take their averages over all replications and divide by the respective "true" $1/N$ results. Thus, we end up with

$$\frac{\overline{\hat{\mu}}^0_a}{\mu^0_{1/N}} \qquad \frac{\overline{\hat{\sigma}}^0_a}{\sigma^0_{1/N}} \qquad \frac{\overline{\widehat{SR}}^0_a}{SR^0_{1/N}}$$

Obviously, as we increase the sample size T the "actual" values of the MVO strategy will converge to their "true" values. From now on, the "no break" case, namely $B = 0$, will act as the benchmark for our analysis in the structural "break" framework.

Now, let us turn to the focal point of this section, which is the structural breaks in mean. We consider $B = \{50, 100, 150, 200\}$ common breaks in mean. All breaks in means take place at the same point in time for N assets and in the same direction. The most intuitive way to compare the "no break" case to the multiple "breaks" cases is to simply test the differences between the MVO "actual" results and the $1/N$ results under each "break" case. We expect that as we increase the number of breaks in mean the MVO "actual" Sharpe ratio will get closer to the $1/N$ Sharpe ratio, compared to the "no break" case, which means that the MVO strategy will become weaker. We will explain further this idea in the following lines.

We assume that in each "break" interval the stock returns follow an multivariate NIID with a mean vector μ_b and a variance-covariance matrix Σ, with $b = 1, 2, \ldots, B+1$. The position of each break is denoted by $\tau_{1:B} = (\tau_1, \tau_2, \ldots, \tau_B)$. We also set the distance between each break to be equal to $\frac{T}{B+1}$. Subsequently, we generate data for each interval $[1, \tau_1], [\tau_1, \tau_2], \ldots, [\tau_B, T]$, through $\mathbb{N}(\mu_b, \Sigma)$, for $b = 0, \ldots, B$. Schematically,

$$
\begin{bmatrix}
\mathbb{N}(\mu_0,\Sigma) & \overset{[1,\tau_1]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\} \\
\mathbb{N}(\mu_1,\Sigma) & \overset{[\tau_1,\tau_2]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\} \\
\vdots & & \vdots \\
\mathbb{N}(\mu_B,\Sigma) & \overset{[\tau_B,T]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\}
\end{bmatrix}_1 \qquad \Rightarrow \hat{\mu}_1, \widehat{\Sigma}_1
$$

$$
\begin{bmatrix}
\mathbb{N}(\mu_0,\Sigma) & \overset{[1,\tau_1]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\} \\
\mathbb{N}(\mu_1,\Sigma) & \overset{[\tau_1,\tau_2]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\} \\
\vdots & & \vdots \\
\mathbb{N}(\mu_B,\Sigma) & \overset{[\tau_B,T]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\}
\end{bmatrix}_2 \qquad \Rightarrow \hat{\mu}_2, \widehat{\Sigma}_2
$$

$$
\vdots \qquad\qquad\qquad\qquad \vdots
$$

$$
\begin{bmatrix}
\mathbb{N}(\mu_0,\Sigma) & \overset{[1,\tau_1]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\} \\
\mathbb{N}(\mu_1,\Sigma) & \overset{[\tau_1,\tau_2]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\} \\
\vdots & & \vdots \\
\mathbb{N}(\mu_B,\Sigma) & \overset{[\tau_B,T]}{\Rightarrow} & \{R_{i1}, R_{i2}, ..., R_{i\frac{T}{B+1}}\}
\end{bmatrix}_S \qquad \Rightarrow \hat{\mu}_S, \widehat{\Sigma}_S
$$

where S represents the number of replications.

Now, let us turn to the "true" MVO and $1/N$ results for the "break" case. The investor could take the average of all means, namely $\mu = \frac{\mu_0 + \ldots \mu_B}{B+1}$, together with the "true" Σ to derive the "true" results for both strategies. This way, the investor will derive the "true" MVO and $1/N$ results of each "break" case as follows

$$\mu^B_{1/N} = \mu^B_r = \frac{1}{N}\mathbb{1}'\mu \qquad\qquad w(\mu, \Sigma, \mu^B_r)$$

$$\sigma^B_{1/N} = \frac{1}{N}\sqrt{\mathbb{1}'\Sigma\mathbb{1}} \qquad\qquad \mu^B_r = \mu^B_{1/N} = w'\mu$$

$$SR^B_{1/N} = \frac{\mu^B_{1/N}}{\sigma^B_{1/N}} \qquad\qquad \sigma^B_r = \sqrt{w'\Sigma w}$$

$$SR^B_r = \frac{\mu^B_r}{\sigma^B_r}$$

Now, we have already argued that the multiple structural breaks will impact the MVO strategy if the investor ignores those breaks. In fact, as the investor overlooks more and more breaks in mean he will end up using the entire sample size to estimate the variance-covariance matrix Σ. This will lead to an inflated estimator $\widehat{\Sigma}$. We will see that this inflated $\widehat{\Sigma}$ impacts the MVO strategy hugely. The MC Simulations will help us measure this impact accurately.

The "actual" results for the "break" case will follow the same rationale as in the Gaussian case. The reason being that the investor ignores the breaks in mean and regards the data as being NIID. Thus, the "actual" results for the "break" case will be

$$\hat{\mu}^B_{a,s} = \hat{w}'_s\mu \qquad \hat{\sigma}^B_{a,s} = \sqrt{\hat{w}'_s\Sigma\hat{w}_s}, \qquad \widehat{SR}^B_{a,s} = \frac{\hat{\mu}^B_{a,s}}{\hat{\sigma}^B_{a,s}}$$

for $s = 1, \ldots, S$. Obviously, as the number of breaks B decreases the "actual" results will converge to their "true" values.

Taking the average of all the over all replications we derive

$$\overline{\hat{\mu}}^B_a = \frac{1}{S}\sum_{s=1}^{S}\hat{\mu}^B_{a,s} \quad \overline{\hat{\sigma}}^B_a = \frac{1}{S}\sum_{s=1}^{S}\hat{\sigma}^B_{a,s} \quad \overline{\widehat{SR}}^B_a = \frac{1}{S}\sum_{s=1}^{S}\widehat{SR}^B_{a,s},$$

$B = \{50, 100, 150, 200\}$

Finally, we focus on the differences between the "actual" MVO results and the results of the "true" $1/N$ case, for each "break" case. These are shown below.

$$\frac{\overline{\hat{\mu}}^B_a}{\mu^B_{1/N}} \qquad \frac{\overline{\hat{\sigma}}^B_a}{\sigma^B_{1/N}} \qquad \frac{\overline{\widehat{SR}}^B_a}{SR^B_{1/N}}$$

These differences will prove to be very different than the respective ones of the "no break" case. In the following section we will discuss our empirical findings using the MC methodology that we analyzed in the current section.

2.4.4 Calibration Results

For this experiment, we get the "no break" parameters μ and Σ for $N = \{10, 40\}$ assets of 100 random groups of assets from S&P 500. In terms of the "break" cases we vary the means between $[-2\%, 2\%]$. The choice of $[-2\%, 2\%]$ is based on the empirical findings that we derived in the previous section. More specifically, the various means that hover around $[-2\%, 2\%]$ will lead to the estimated correlations being approximately equal to 20% for the $B = 50$ break case and 40% for the $B = 200$ break case, which is what we showed to be the case in real data. The sample sizes will be $T = \{1000, 5000, 10000\}$ while the number of replications is set to $S = 10,000$. Finally, the different cases of breaks are $B = \{0, 50, 100, 150, 200\}$. For each replication, we follow the steps described in the MC Simulations section.

The following Tables summarize the results for the $N = 10$ assets case.

Table 2.6: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 10$ assets and $T = 1,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	0.97	1.20	**1.05**	0.07	0.82	0.98	**0.93**	0.04	0.93	1.02	**0.97**	0.03
$B = 50$	0.63	0.89	**0.74**	0.08	0.97	1.26	**1.10**	0.09	0.78	0.93	**0.82**	0.04
$B = 100$	0.48	0.83	**0.60**	0.10	1.06	1.39	**1.19**	0.12	0.62	0.90	**0.70**	0.08
$B = 150$	0.44	0.82	**0.56**	0.03	1.08	1.42	**1.21**	0.12	0.57	0.90	**0.67**	0.09
$B = 200$	0.43	0.82	**0.55**	0.08	1.08	1.41	**1.21**	0.12	0.55	0.90	**0.66**	0.10

Table 2.7: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 10$ assets and $T = 5,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.02	1.24	**1.05**	0.06	0.80	0.96	**0.91**	0.05	0.96	1.01	**0.99**	0.01
$B = 50$	0.85	1.08	**0.95**	0.06	0.86	1.09	**0.99**	0.06	0.92	0.96	**0.93**	0.01
$B = 100$	0.63	0.88	**0.74**	0.08	0.98	1.28	**1.11**	0.10	0.79	0.92	**0.82**	0.04
$B = 150$	0.52	0.85	**0.64**	0.09	1.04	1.36	**1.17**	0.11	0.69	0.91	**0.74**	0.06
$B = 200$	0.49	0.84	**0.61**	0.10	1.05	1.38	**1.18**	0.11	0.64	0.90	**0.71**	0.08

Table 2.8: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 10$ assets and $T = 10,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.04	1.25	**1.10**	0.06	0.80	0.95	**0.91**	0.05	0.98	1.00	**0.99**	0.01
$B = 50$	0.93	1.16	**1.01**	0.06	0.83	1.03	**0.95**	0.06	0.96	0.97	**0.96**	0.01
$B = 100$	0.73	0.94	**0.83**	0.07	0.93	1.20	**1.06**	0.08	0.87	0.93	**0.88**	0.02
$B = 150$	0.60	0.87	**0.72**	0.08	0.99	1.31	**1.13**	0.10	0.78	0.92	**0.81**	0.04
$B = 200$	0.55	0.86	**0.66**	0.09	1.02	1.34	**1.15**	0.11	0.71	0.91	**0.76**	0.06

Tables 2.6, 2.7 and 2.8, showcase that if an investor overlooks more and more breaks in mean he will end up with lower portfolio returns and Sharpe ratios and a higher portfolio risk. More specifically, on Table 2.6 we see that contrary to the “no break” case for which the “actual” Sharpe ratio is 5% than that of the naive strategy, when the number of breaks B is 50, the “actual” Sharpe ratio is approximately 26% lower than that of the $1/N$ strategy. In fact, as the number of breaks increases and gets closer to $B = 200$, the “actual” Sharpe ratio becomes 45% worse than the $1/N$ case. We will find out that the major reason for this drop comes from the inflated $\widehat{\Sigma}$. Observe that in the “no break” case the “actual” portfolio risk is approximately 7% lower than the $1/N$ portfolio risk, while for the “break” cases it ends up being 21% higher. While the “actual” portfolio

risk of the "break" cases increases, the "actual" portfolio returns drop. So, although the investor will receive a lower "actual" portfolio return he will also get an even worse portfolio risk, compared to the respective results of the naive strategy. Tables 2.7 and 2.8 illustrate the effect of increasing the sample size. Although there is a slight improvement in the "actual" "break" results, the MVO is far from dominating the $1/N$ method.

Tables 2.9, 2.10 and 2.11 summarize the results for the $N = 40$ assets case. Increasing the number of assets seems to neutralize the MVO strategy completely. The "actual" Sharpe ratios of the "break" cases are almost 80% lower than the naive Sharpe ratios. More importantly, increasing the sample size has a only a very mild positive effect on the MVO strategy. The only case which benefits the most from increasing the sample size is the $B = 50$ case for which the "actual" Sharpe ratio goes from -70% with respect to the naive Sharpe ratio, to -30%. Still, the $1/N$ strategy dominates the MVO strategy by a huge gap.

Table 2.9: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 1,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.02	1.14	**1.07**	0.03	0.91	1.14	**1.09**	0.01	0.93	1.03	**0.99**	0.03
$B = 50$	0.25	0.31	**0.27**	0.02	1.53	1.94	**1.80**	0.11	0.46	0.50	**0.48**	0.02
$B = 100$	0.16	0.20	**0.18**	0.01	1.51	1.94	**1.79**	0.11	0.30	0.36	**0.33**	0.02
$B = 150$	0.15	0.18	**0.17**	0.01	1.49	1.92	**1.76**	0.11	0.26	0.32	**0.29**	0.02
$B = 200$	0.14	0.18	**0.17**	0.01	1.47	1.89	**1.74**	0.11	0.25	0.32	**0.29**	0.02

Table 2.10: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 5,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.07	1.14	**1.09**	0.02	0.89	1.14	**1.09**	0.01	0.97	1.01	**1.00**	0.01
$B = 50$	0.44	0.55	**0.47**	0.03	1.37	1.70	**1.60**	0.10	0.75	0.76	**0.75**	0.01
$B = 100$	0.24	0.31	**0.27**	0.02	1.55	1.98	**1.84**	0.12	0.47	0.50	**0.49**	0.01
$B = 150$	0.19	0.23	**0.21**	0.02	1.53	1.96	**1.81**	0.12	0.36	0.40	**0.38**	0.02
$B = 200$	0.17	0.21	**0.19**	0.01	1.49	1.91	**1.77**	0.11	0.31	0.37	**0.34**	0.02

Table 2.11: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 10,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\overline{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\overline{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.07	1.13	**1.10**	0.02	0.89	0.93	**0.91**	0.01	0.99	1.00	**1.00**	0.01
$B = 50$	0.65	0.77	**0.68**	0.04	1.15	1.38	**1.31**	0.07	0.89	0.90	**0.89**	0.01
$B = 100$	0.36	0.46	**0.39**	0.03	1.47	1.87	**1.74**	0.12	0.67	0.69	**0.68**	0.01
$B = 150$	0.26	0.33	**0.29**	0.02	1.56	1.99	**1.85**	0.12	0.51	0.54	**0.53**	0.01
$B = 200$	0.22	0.28	**0.24**	0.02	1.54	1.96	**1.82**	0.12	0.42	0.47	**0.45**	0.01

Clearly, the above results indicate that Markowitz's method under structural breaks in mean is completely outperformed by the naive strategy. The main reason being, that the naive strategy is not affected at all by the investor's wrong assumption with respect to the DGP of the data. Specifically, even if the investor ignores multiple breaks in mean, assuming that the data is NIID, the naive strategy will deliver the same results. At the end of this chapter, in Appendix B 2.6, we also include the same Tables with means taking values inside $[-1\%, 1\%]$. As one can see, we have the same conclusions but, as expected, the impact of the breaks on the "actual" results is more subtle, especially for the $N = 10$ assets case. However, even under this milder environment our conclusions still hold.

Partially estimated weights

One natural question is how can we be certain that the main impact on the MVO strategy comes from the inflated $\widehat{\Sigma}$ and not from $\hat{\mu}$? Let us assume that the investor who overlooks the multiple breaks in mean observes the "true" variance-covariance matrix, namely Σ. This would mean that the investor would only use the entire sample to estimate just $\hat{\mu}$. Thus, the estimated weights would be $\hat{w}(\hat{\mu}, \Sigma, \mu_r^B)$. Then we should compare the "partially" estimated "actual" results of both the "no break" and "break" cases with the respective "true" naive results. The information we need is summarized in the following Tables.

Table 2.12: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 10$ assets and $T = 1,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	0.97	1.21	**1.05**	0.05	0.82	0.97	**0.93**	0.04	0.93	1.02	**0.97**	0.03
$B = 50$	1.00	1.23	**1.07**	0.07	0.81	0.96	**0.92**	0.05	0.95	1.02	**0.98**	0.02
$B = 100$	0.96	1.19	**1.05**	0.07	0.82	0.97	**0.93**	0.04	0.92	1.04	**0.97**	0.04
$B = 150$	0.95	1.18	**1.04**	0.07	0.83	0.98	**0.94**	0.04	0.92	1.05	**0.97**	0.04
$B = 200$	0.95	1.17	**1.03**	0.07	0.84	0.98	**0.94**	0.04	0.91	1.06	**0.97**	0.04

Table 2.13: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 1,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.04	1.16	**1.09**	0.04	0.89	0.93	**0.91**	0.01	0.93	1.03	**0.99**	0.03
$B = 50$	1.06	1.15	**1.09**	0.03	0.89	0.93	**0.91**	0.01	0.95	1.02	**0.99**	0.02
$B = 100$	1.03	1.17	**1.08**	0.04	0.89	0.93	**0.91**	0.01	0.92	1.04	**0.99**	0.03
$B = 150$	1.02	1.18	**1.08**	0.04	0.89	0.94	**0.91**	0.01	0.91	1.05	**0.98**	0.04
$B = 200$	1.02	1.18	**1.08**	0.05	0.89	0.94	**0.91**	0.01	0.90	1.05	**0.98**	0.04

The results clearly indicate that in our framework the main issue the investor faces is that if he misses the multiple breaks in mean the problems arising with respect to the effectiveness of the MVO strategy against $1/N$, will only be due to the misestimation of variance-covariance matrix. This showcases how much impactful are the positive correlations on the MVO strategy. It also gives substance to our rationale that ignoring multiple structural breaks in mean constitutes a significant reason for the underperformance of the MVO method with respect to $1/N$.

2.5 Conclusions

We started by analyzing a part of the work of DeMiguel et al. [24] which motivated us to revisit Markowitz's portfolio theory. Our main objections regarding their work stemmed from the way they denounce the MVO method for being useless when compared to the naive diversification method. We argued that the two models should be tested under the same terms. This is why we chose to set the target return of the MVO model equal to the return coming from the $1/N$ strategy. In our view, this constitutes a more appropriate way to study the statistical properties of the MVO method but also to be able to do a more meaningful analysis.

Since the weights produced by the Markowitz method are a non-linear function of the true means, variance-covariance matrix and target return, Jensen's inequality indicates that the estimated weights are biased estimators. However, under the Continuous Mapping Theorem they are consistent. One of the most well-known ways to use in order to evaluate how these findings translate into reality is by applying MC Simulations. We started by studying the speed at which the variances of the estimated weights converge to zero. We found out that as we increase either (i) the difference between the MVO Sharpe ratio and the $1/N$ Sharpe ratio or (ii) the number of assets N, the convergence rate speeds up. From there, we continued with the comparison between the MVO method and $1/N$ for $N = \{10, 25\}$ assets. We highlighted the importance of controlling $SR_r/SR_{1/N}$ and μ_r/μ_{GMV}. Our results indicated that as we increase $SR_r/SR_{1/N}$ the "actual" results of the MVO method outperform those of $1/N$ for relatively small sample sizes. In fact, we found no evidence of the MVO method being outperformed by $1/N$ under Gaussian DGPs, other than cases where $SR_r/SR_{1/N}$ was very close to 1. Overall, we concluded that under Gaussian DGPs, we can safely argue that the MVO method is preferable to the naive diversification method.

Subsequently, we examined the functionality of the MVO method under the assumption that the DGP of stock returns presents multiple breaks in mean. This part of our research has been overlooked by the literature. Collecting daily stock returns spanning from 2000 to 2021 from the constituents of S&P 500, FTSE 100 and STOXX 600, we showcased that there are groups of $N = 10$ assets that exhibit $B = \{50, 100, 150, 200\}$ common breaks in mean. For these groups the estimated correlations considering the $B = 50$ breaks case hover around 20%, while in the $B = 200$ breaks case most of the estimated correlations are approximately 40%. In order to evaluate these findings, we reapplied MC Simulations, to find out how impactful multiple breaks are with respect to the MVO method. Specifically, we focused on the comparison between the MVO method and the naive strategy, as in the Gaussian case. Both strategies were tested in terms of their "actual" results under $B = \{50, 100, 150, 200\}$ breaks as well as under a "no break" case. In our framework, we considered that the investor ignores the multiple breaks that take place and proceeds with estimating the DGP's parameters using the entire sample size. Our results indicated that overlooking more and more breaks in mean leads to increasingly worse "actual" Sharpe ratios derived by the MVO strategy compared to the Sharpe ratios of $1/N$. More importantly, as we increase the number of assets the MVO method collapses. We argued that the main reason for that is the estimation of the variance-covariance matrix using the entire sample size. Specifically, the higher the number of breaks the more inflated $\widehat{\Sigma}$ will be. Overall, our results clearly indicate that Markowitz's performance is highly dependent on the nature of the DGP. Putting Markowitz's method under a DGP with multiple breaks proves to be an interesting area for further research.

2.6 Appendix

Appendix A: MVO vs $1/N$ under Gaussian DGPs

Table 2.14: $\frac{SR_r}{SR_{1/N}} = 1.1$ and $\frac{\mu_r}{\mu_{GMV}} = 1.25$

	Sample size			
	min	max	**mean**	std
$N = 10$	1800	3900	**3140**	380
$N = 15$	2000	4750	**3475**	429
$N = 20$	2500	5500	**3777**	437
$N = 25$	2700	6000	**4284**	432
$N = 30$	2800	6300	**4700**	441
$N = 35$	3100	6700	**5228**	476
$N = 40$	3300	7400	**5568**	478

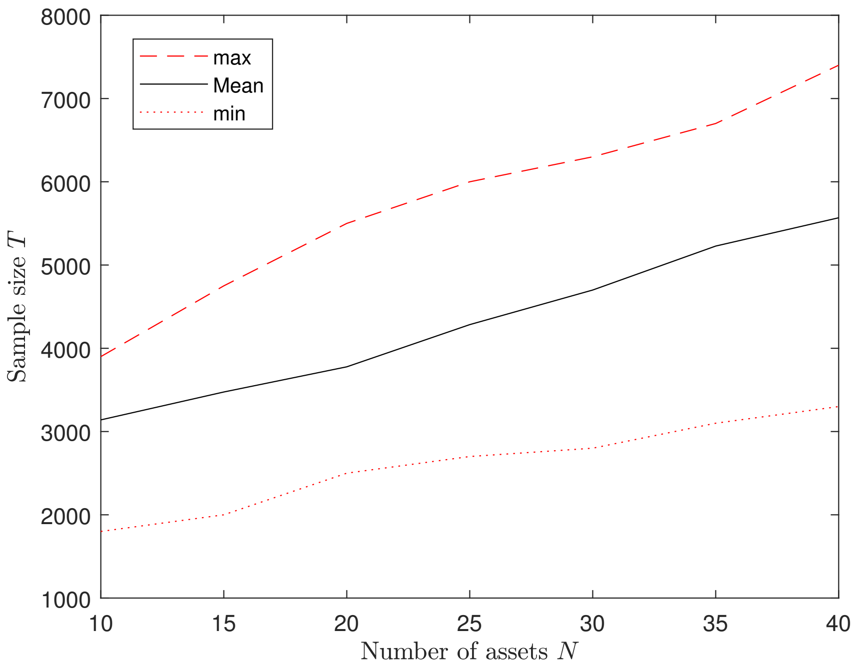


Table 2.15: $\frac{SR_r}{SR_{1/N}} = 1.1$ and $\frac{\mu_r}{\mu_{GMV}} = 1.5$

	Sample size			
	min	max	**mean**	std
$N = 10$	2700	7100	**4563**	621
$N = 15$	3400	7300	**4903**	568
$N = 20$	3700	7500	**5175**	564
$N = 25$	4100	7800	**5354**	543
$N = 30$	4200	8100	**5752**	526
$N = 35$	4300	8200	**6087**	521
$N = 40$	4500	8700	**6554**	536

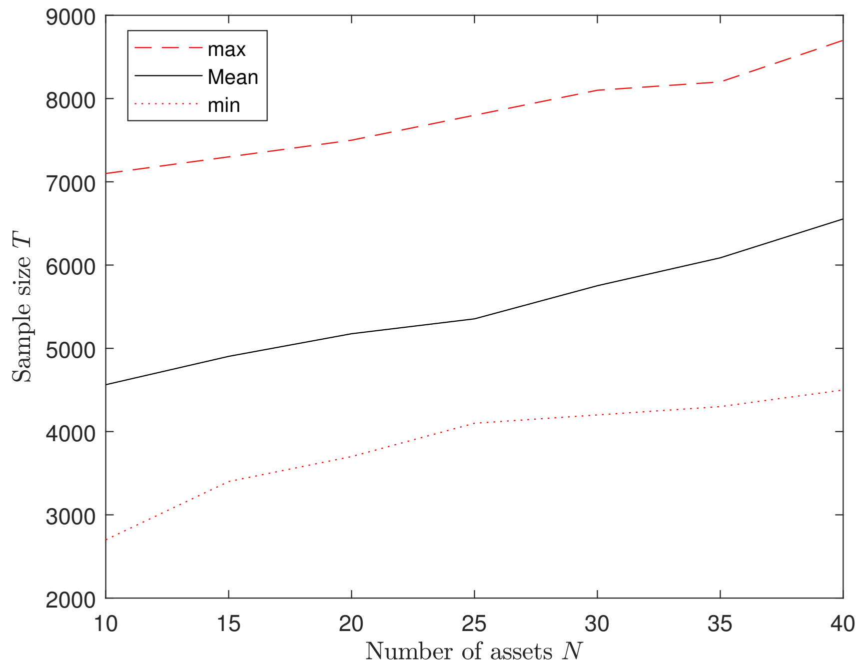


Table 2.16: $\frac{SR_r}{SR_{1/N}} = 1.1$ and $\frac{\mu_r}{\mu_{GMV}} = 1.75$

	Sample size			
	min	max	**mean**	std
$N = 10$	3400	7300	**5084**	572
$N = 15$	4000	7500	**5513**	554
$N = 20$	4300	7800	**5980**	548
$N = 25$	4300	8000	**6363**	556
$N = 30$	4500	8200	**6795**	554
$N = 35$	4900	8500	**7194**	532
$N = 40$	5100	8900	**7726**	549

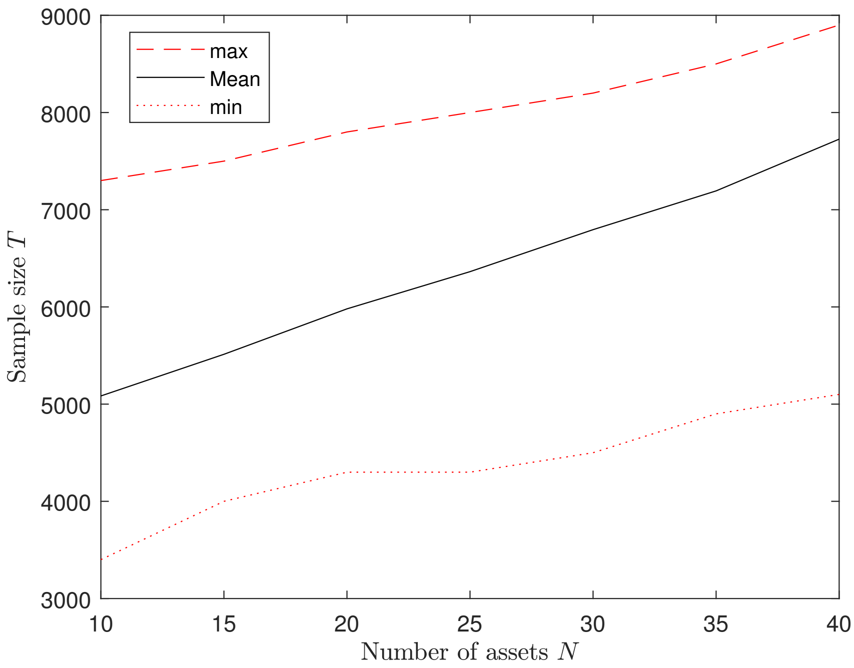


Table 2.17: $\frac{SR_r}{SR_{1/N}} = 1.1$ and $\frac{\mu_r}{\mu_{GMV}} = 2$

	Sample size			
	min	max	**mean**	std
$N = 10$	4200	7400	**5467**	642
$N = 15$	4800	9000	**6172**	693
$N = 20$	5100	9300	**6571**	708
$N = 25$	5400	9400	**7083**	719
$N = 30$	5300	9700	**7312**	734
$N = 35$	5500	10000	**8017**	765
$N = 40$	5700	10200	**8750**	784

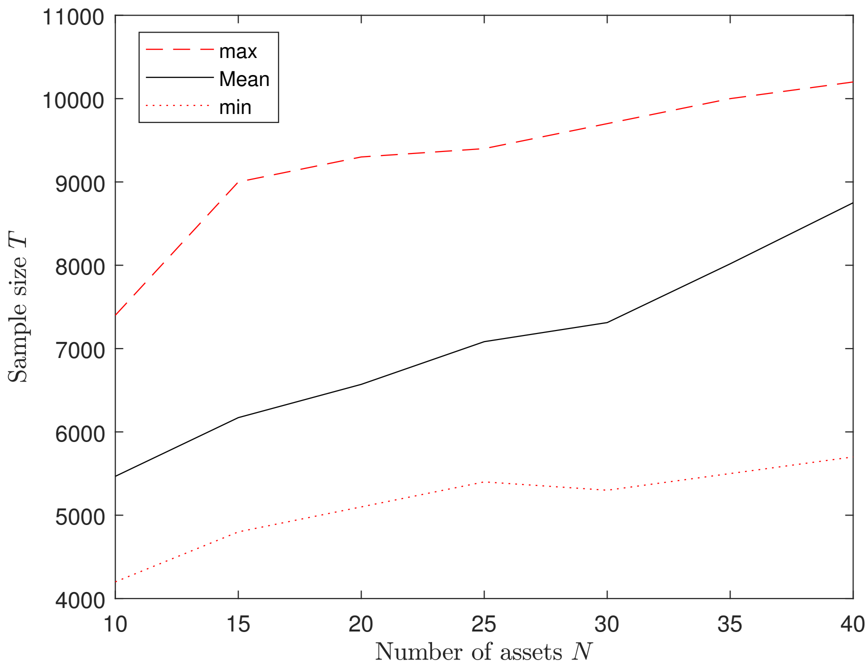


Table 2.18: $\frac{SR_r}{SR_{1/N}} = 1.2$ and $\frac{\mu_r}{\mu_{GMV}} = 1.25$

	Sample size			
	min	max	**mean**	std
$N = 10$	600	1000	**790**	114
$N = 15$	700	900	**760**	66
$N = 20$	600	1000	**803**	98
$N = 25$	600	1000	**820**	91
$N = 30$	700	1000	**837**	75
$N = 35$	750	1060	**868**	72
$N = 40$	770	1120	**883**	81

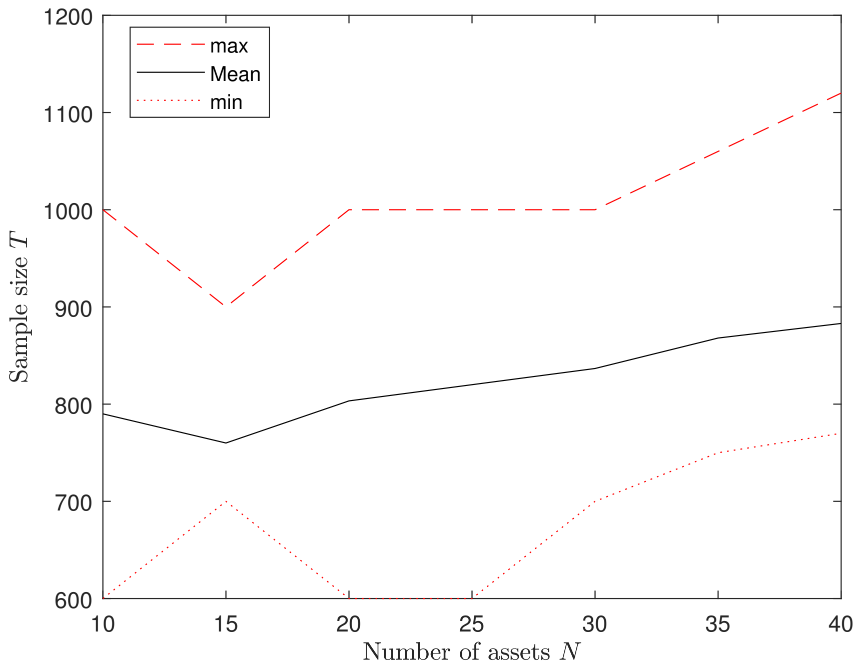


Table 2.19: $\frac{SR_r}{SR_{1/N}} = 1.2$ and $\frac{\mu_r}{\mu_{GMV}} = 1.5$

	Sample size			
	min	max	**mean**	std
$N = 10$	1000	3200	**2012**	408
$N = 15$	1400	3100	**2013**	364
$N = 20$	1300	3400	**2296**	439
$N = 25$	1500	4400	**2405**	437
$N = 30$	2000	4500	**2438**	420
$N = 35$	2250	4300	**2583**	391
$N = 40$	2750	4700	**3013**	367

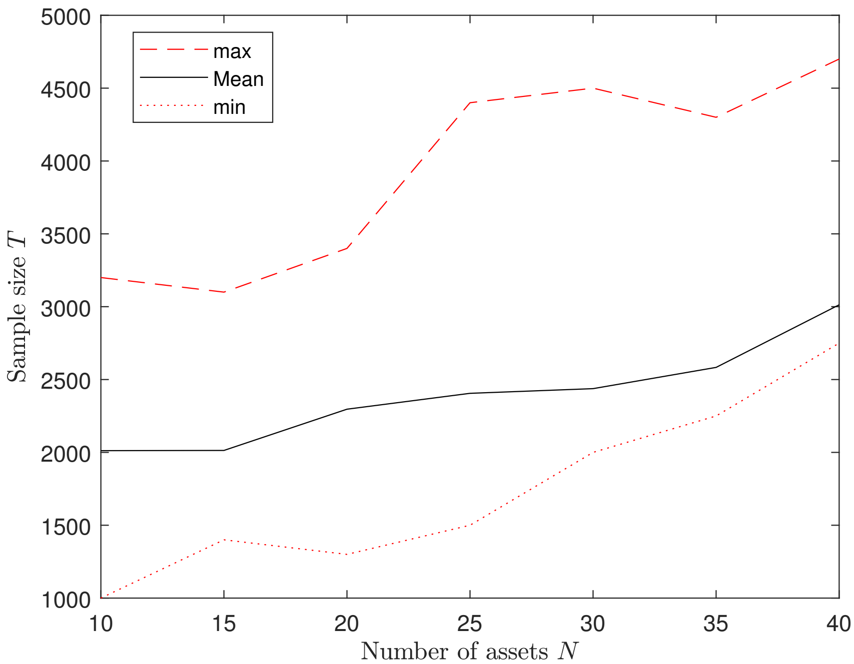


Table 2.20: $\frac{SR_r}{SR_{1/N}} = 1.2$ and $\frac{\mu_r}{\mu_{GMV}} = 1.75$

	Sample size			
	min	max	**mean**	std
$N = 10$	1500	3700	**2478**	427
$N = 15$	1800	4000	**2642**	571
$N = 20$	2300	5200	**3030**	574
$N = 25$	2100	6600	**3435**	617
$N = 30$	2400	6400	**3893**	601
$N = 35$	2600	6900	**3970**	611
$N = 40$	3100	7600	**4685**	621

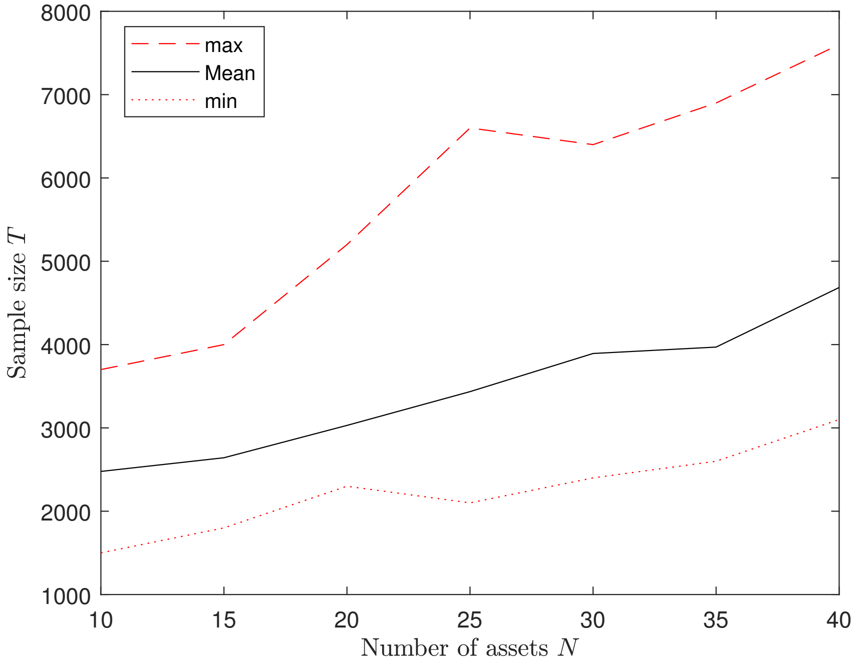


Table 2.21: $\frac{SR_r}{SR_{1/N}} = 1.2$ and $\frac{\mu_r}{\mu_{GMV}} = 2$

	Sample size			
	min	max	**mean**	std
$N = 10$	2000	4500	**2878**	402
$N = 15$	2200	6100	**3223**	601
$N = 20$	2300	6000	**3449**	571
$N = 25$	2500	6400	**3773**	592
$N = 30$	3000	7300	**4268**	607
$N = 35$	3200	8000	**4666**	625
$N = 40$	3600	8500	**5173**	612

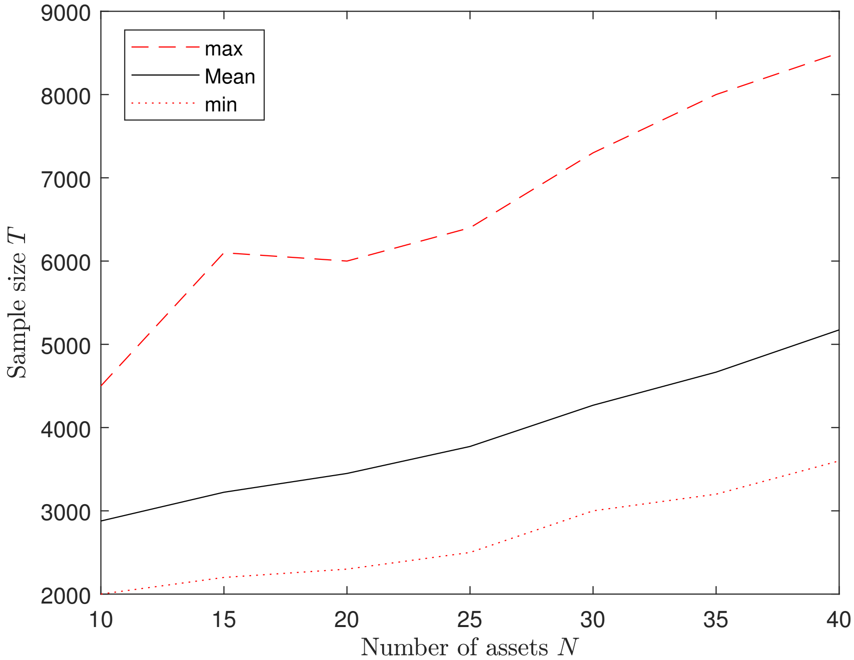


Table 2.22: $\frac{SR_r}{SR_{1/N}} = 1.3$ and $\frac{\mu_r}{\mu_{GMV}} = 1.25$

	Sample size			
	min	max	**mean**	std
$N = 10$	100	150	**140**	20
$N = 15$	120	170	**149**	11
$N = 20$	160	190	**178**	7
$N = 25$	190	230	**206**	16
$N = 30$	250	300	**255**	15
$N = 35$	270	290	**283**	9
$N = 40$	300	340	**316**	13

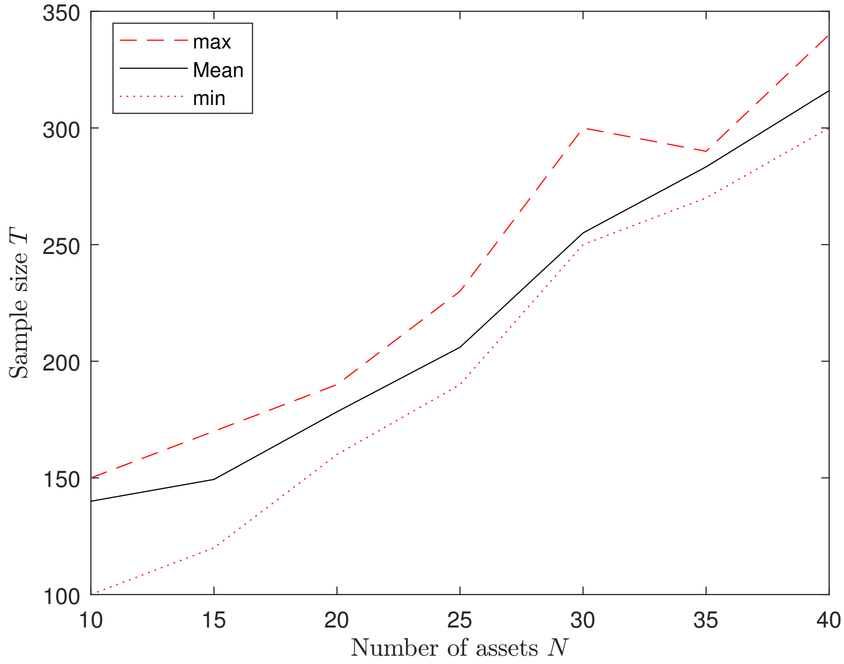


Table 2.23: $\frac{SR_r}{SR_{1/N}} = 1.3$ and $\frac{\mu_r}{\mu_{GMV}} = 1.5$

	Sample size			
	min	max	**mean**	std
$N = 10$	500	800	**699**	61
$N = 15$	650	950	**817**	73
$N = 20$	700	1100	**886**	84
$N = 25$	800	1200	**1020**	118
$N = 30$	900	1400	**1160**	143
$N = 35$	900	1600	**1197**	147
$N = 40$	1000	1500	**1277**	150

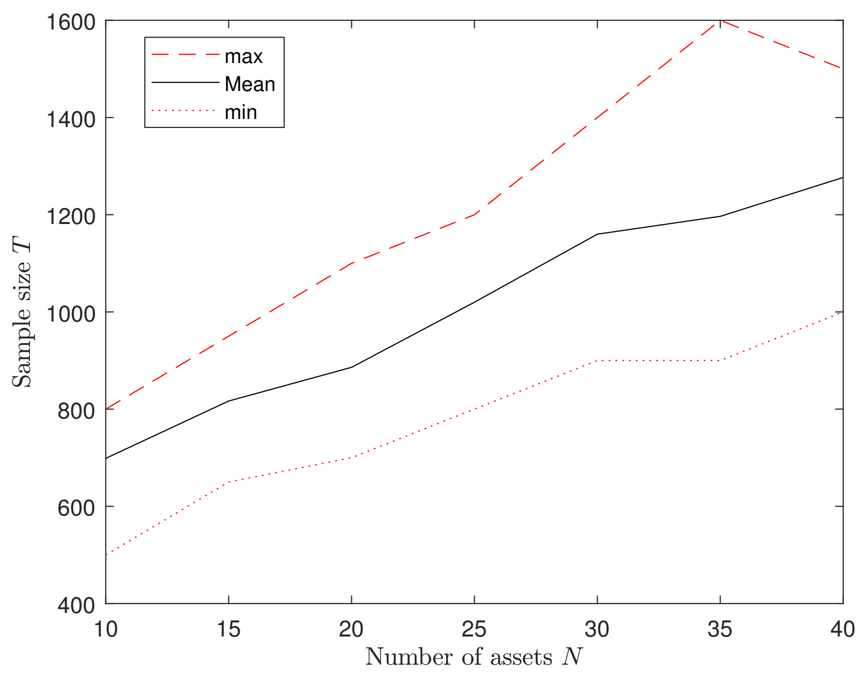


Table 2.24: $\frac{SR_r}{SR_{1/N}} = 1.3$ and $\frac{\mu_r}{\mu_{GMV}} = 1.75$

	Sample size			
	min	max	**mean**	std
$N = 10$	1000	1900	**1338**	208
$N = 15$	1000	2200	**1440**	213
$N = 20$	1100	2700	**1654**	254
$N = 25$	1200	3200	**1873**	306
$N = 30$	1400	3500	**2046**	311
$N = 35$	1600	3700	**2233**	336
$N = 40$	1800	4000	**2403**	327

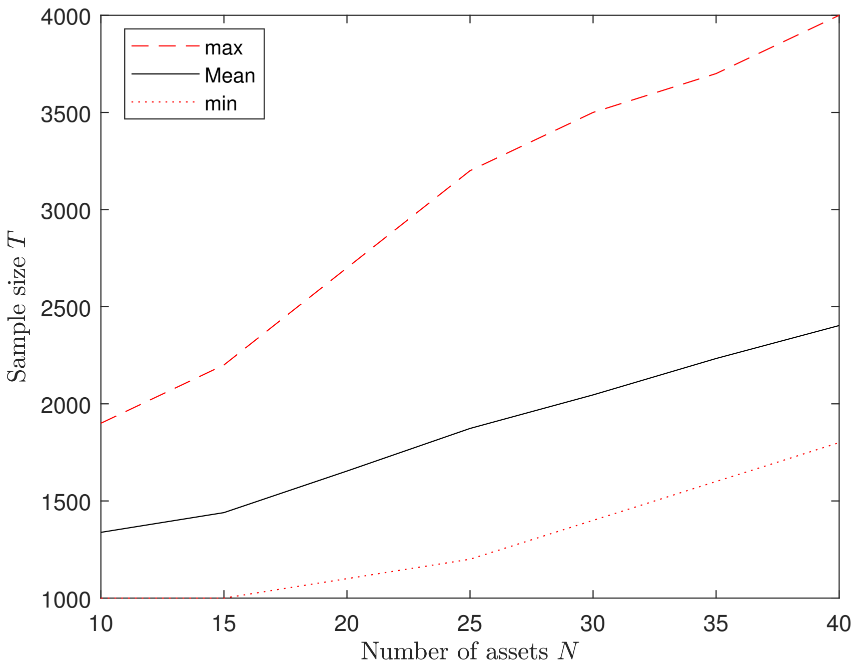


Table 2.25: $\frac{SR_r}{SR_{1/N}} = 1.3$ and $\frac{\mu_r}{\mu_{GMV}} = 2$

	Sample size			
	min	max	**mean**	std
$N = 10$	1200	2300	**1689**	214
$N = 15$	1400	2600	**1874**	236
$N = 20$	1400	2800	**1982**	255
$N = 25$	1750	3500	**2453**	283
$N = 30$	1750	3800	**2564**	314
$N = 35$	1900	4200	**2674**	325
$N = 40$	2000	4800	**2849**	359

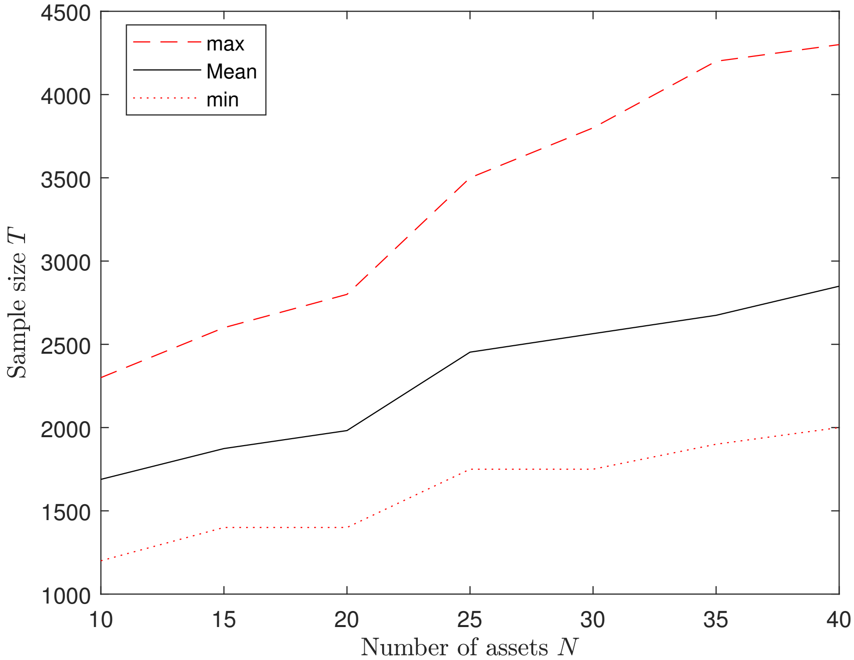


Table 2.26: $\frac{SR_r}{SR_{1/N}} = 1.4$ and $\frac{\mu_r}{\mu_{GMV}} = 1.25$

	Sample size			
	min	max	**mean**	std
$N = 10$	50	60	**53**	2
$N = 15$	60	70	**69**	3
$N = 20$	80	90	**89**	2
$N = 25$	100	110	**109**	3
$N = 30$	140	160	**145**	5
$N = 35$	150	190	**168**	11
$N = 40$	180	220	**180**	14

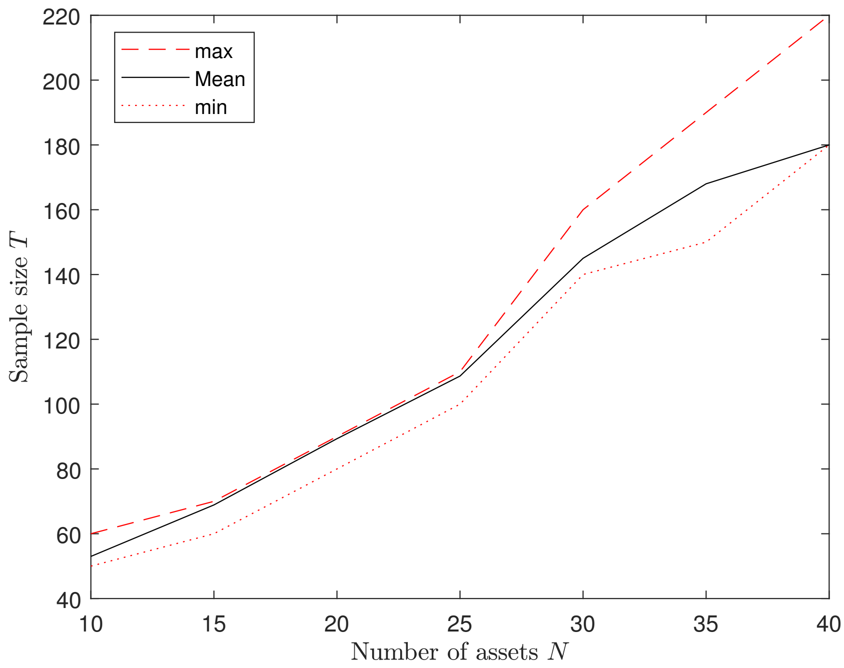


Table 2.27: $\frac{SR_r}{SR_{1/N}} = 1.4$ and $\frac{\mu_r}{\mu_{GMV}} = 1.5$

	Sample size			
	min	max	**mean**	std
$N = 10$	240	310	**274**	18
$N = 15$	290	360	**335**	25
$N = 20$	320	400	**364**	22
$N = 25$	400	450	**437**	22
$N = 30$	450	550	**505**	30
$N = 35$	460	580	**538**	34
$N = 40$	500	640	**579**	36

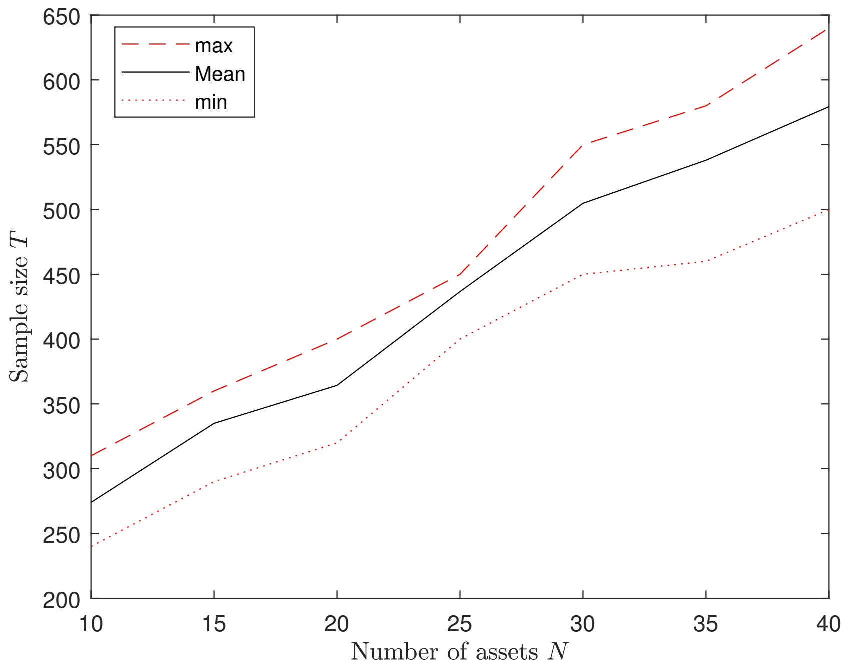


Table 2.28: $\frac{SR_r}{SR_{1/N}} = 1.4$ and $\frac{\mu_r}{\mu_{GMV}} = 1.75$

	Sample size			
	min	max	**mean**	std
$N = 10$	580	710	**657**	54
$N = 15$	600	750	**678**	67
$N = 20$	610	790	**694**	71
$N = 25$	660	880	**739**	73
$N = 30$	660	1100	**830**	132
$N = 35$	690	1000	**845**	126
$N = 40$	720	1400	**909**	159

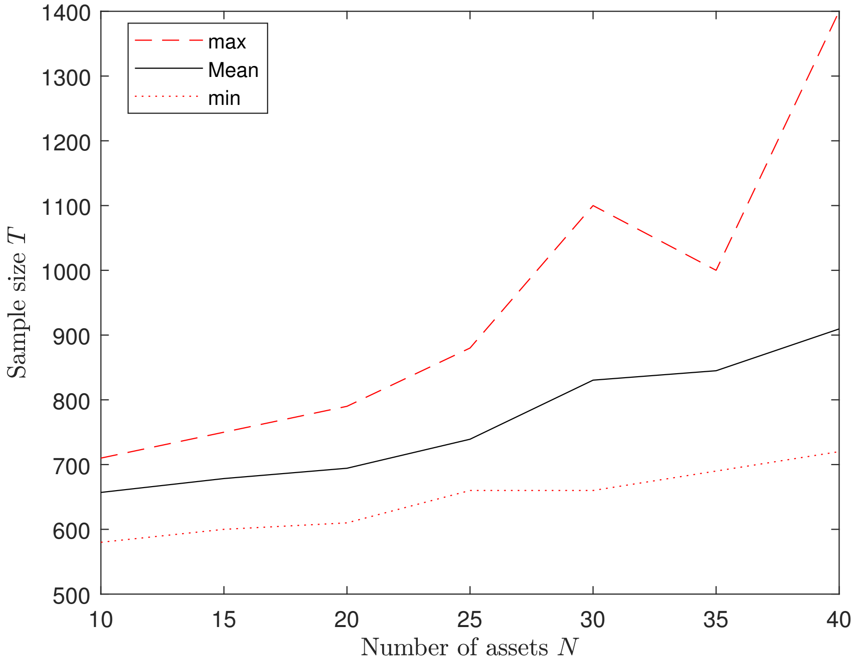
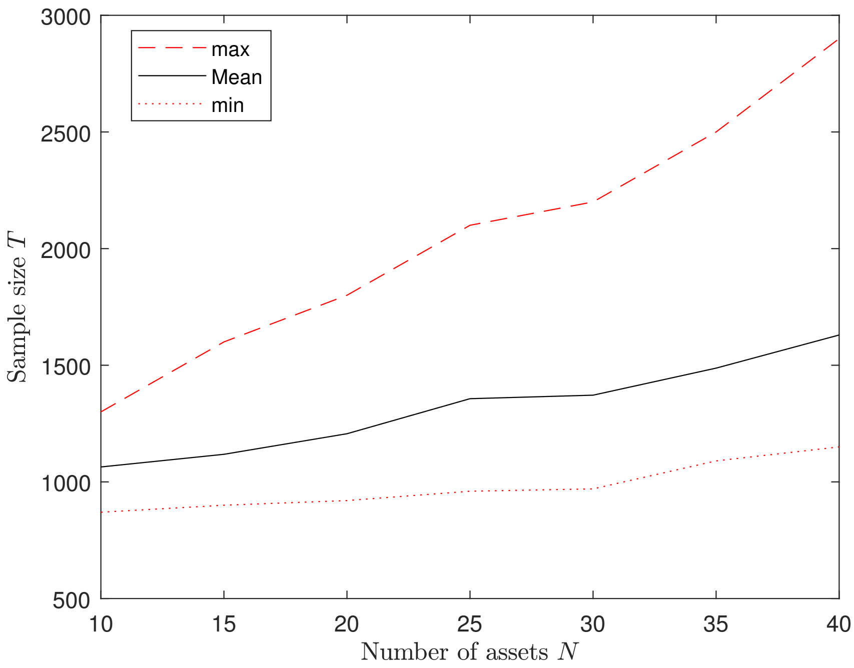


Table 2.29: $\frac{SR_r}{SR_{1/N}} = 1.4$ and $\frac{\mu_r}{\mu_{GMV}} = 2$

	Sample size			
	min	max	**mean**	std
$N = 10$	870	1300	**1064**	103
$N = 15$	900	1600	**1118**	121
$N = 20$	920	1800	**1206**	138
$N = 25$	960	2100	**1357**	133
$N = 30$	970	2200	**1372**	128
$N = 35$	1090	2500	**1488**	123
$N = 40$	1150	2900	**1630**	137



Appendix B: MVO vs $1/N$ under DGPs with structural breaks in mean, with means around $[-1\%, 1\%]$

Table 2.30: $\frac{\overline{\widehat{SR}_a}^B}{SR^B_{1/N}}$, $\frac{\overline{\hat{\sigma}}^B_a}{\sigma^B_{1/N}}$ and $\frac{\overline{\hat{\mu}}^B_a}{\mu^B_{1/N}}$ for portfolios of $N = 10$ assets and $T = 1,000$

	$\frac{\overline{\widehat{SR}_a}^B}{SR^B_{1/N}}$				$\frac{\overline{\hat{\sigma}}^B_a}{\sigma^B_{1/N}}$				$\frac{\overline{\hat{\mu}}^B_a}{\mu^B_{1/N}}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	0.97	1.20	**1.05**	0.07	0.82	0.97	**0.93**	0.05	0.93	1.03	**0.97**	0.03
$B = 50$	0.90	1.10	**0.98**	0.06	0.83	1.01	**0.95**	0.05	0.90	0.99	**0.93**	0.02
$B = 100$	0.81	1.02	**0.91**	0.06	0.86	1.03	**0.97**	0.05	0.83	0.98	**0.88**	0.04
$B = 150$	0.79	1.02	**0.90**	0.06	0.87	1.03	**0.98**	0.05	0.82	0.99	**0.87**	0.05
$B = 200$	0.78	1.02	**0.88**	0.07	0.88	1.04	**0.98**	0.05	0.81	0.99	**0.86**	0.05

Table 2.31: $\frac{\overline{\widehat{SR}_a}^B}{SR^B_{1/N}}$, $\frac{\overline{\hat{\sigma}}^B_a}{\sigma^B_{1/N}}$ and $\frac{\overline{\hat{\mu}}^B_a}{\mu^B_{1/N}}$ for portfolios of $N = 10$ assets and $T = 5,000$

	$\frac{\overline{\widehat{SR}_a}^B}{SR^B_{1/N}}$				$\frac{\overline{\hat{\sigma}}^B_a}{\sigma^B_{1/N}}$				$\frac{\overline{\hat{\mu}}^B_a}{\mu^B_{1/N}}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.02	1.24	**1.09**	0.06	0.80	0.96	**0.91**	0.05	0.96	1.01	**0.99**	0.01
$B = 50$	1.00	1.22	**1.07**	0.06	0.80	0.97	**0.92**	0.05	0.96	0.99	**0.98**	0.01
$B = 100$	0.92	1.12	**0.99**	0.06	0.83	1.00	**0.94**	0.05	0.91	0.99	**0.93**	0.02
$B = 150$	0.87	1.06	**0.95**	0.06	0.84	1.01	**0.95**	0.05	0.87	0.99	**0.91**	0.03
$B = 200$	0.83	1.03	**0.92**	0.06	0.85	1.02	**0.96**	0.05	0.85	0.99	**0.89**	0.04

Table 2.32: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 10$ assets and $T = 10,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.03	1.25	**1.09**	0.06	0.80	0.96	**0.91**	0.05	0.97	1.01	**0.99**	0.01
$B = 50$	1.02	1.24	**1.09**	0.06	0.80	0.96	**0.91**	0.05	0.98	1.00	**0.99**	0.01
$B = 100$	0.97	1.18	**1.04**	0.06	0.81	0.97	**0.92**	0.05	0.94	0.99	**0.96**	0.02
$B = 150$	0.92	1.13	**1.00**	0.06	0.82	0.99	**0.94**	0.05	0.91	0.99	**0.94**	0.03
$B = 200$	0.89	1.09	**0.97**	0.06	0.83	1.00	**0.95**	0.05	0.88	0.99	**0.92**	0.03

Table 2.33: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 1,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.02	1.14	**1.07**	0.03	0.91	0.95	**0.93**	0.01	0.93	1.04	**0.99**	0.03
$B = 50$	0.61	0.70	**0.64**	0.03	1.09	1.23	**1.19**	0.04	0.73	0.78	**0.76**	0.01
$B = 100$	0.48	0.56	**0.52**	0.03	1.13	1.29	**1.23**	0.04	0.61	0.68	**0.65**	0.02
$B = 150$	0.47	0.55	**0.52**	0.03	1.13	1.27	**1.22**	0.04	0.59	0.67	**0.63**	0.03
$B = 200$	0.44	0.52	**0.50**	0.03	1.14	1.30	**1.23**	0.04	0.56	0.65	**0.61**	0.03

Table 2.34: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 5,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.07	1.14	**1.09**	0.02	0.89	0.93	**0.91**	0.01	0.97	1.01	**1.00**	0.01
$B = 50$	0.86	0.95	**0.88**	0.03	0.97	1.05	**1.03**	0.03	0.90	0.92	**0.91**	0.01
$B = 100$	0.65	0.75	**0.69**	0.03	1.05	1.18	**1.14**	0.04	0.76	0.80	**0.78**	0.01
$B = 150$	0.58	0.66	**0.62**	0.03	1.07	1.21	**1.16**	0.04	0.68	0.74	**0.71**	0.02
$B = 200$	0.51	0.60	**0.56**	0.03	1.10	1.24	**1.19**	0.04	0.63	0.70	**0.66**	0.02

Table 2.35: $\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$, $\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$ and $\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$ for portfolios of $N = 40$ assets and $T = 10,000$

	$\frac{\overline{\widehat{SR}}_a^B}{SR_{1/N}^B}$				$\frac{\bar{\hat{\sigma}}_a^B}{\sigma_{1/N}^B}$				$\frac{\bar{\hat{\mu}}_a^B}{\mu_{1/N}^B}$			
	min	max	**mean**	std	min	max	**mean**	std	min	max	**mean**	std
$B = 0$	1.07	1.13	**1.10**	0.02	0.89	0.93	**0.91**	0.01	0.98	1.01	**1.00**	0.01
$B = 50$	0.96	1.02	**0.97**	0.02	0.93	0.99	**0.98**	0.02	0.94	0.95	**0.95**	0.01
$B = 100$	0.76	0.85	**0.79**	0.03	1.00	1.11	**1.08**	0.03	0.83	0.86	**0.85**	0.01
$B = 150$	0.66	0.75	**0.75**	0.03	1.04	1.16	**1.12**	0.04	0.75	0.80	**0.78**	0.01
$B = 200$	0.58	0.67	**0.62**	0.03	1.07	1.21	**1.16**	0.04	0.69	0.74	**0.72**	0.02

Chapter 3

On The Equivalence Of The Mean-Variance Criterion And Stochastic Dominance Criteria

3.1 Introduction

Back in the 18^{th} century Daniel Bernoulli proposed a solution for the St. Petersburg paradox. The solution was simply based on the assumption that an investor aims at maximizing his expected utility rather than his wealth. The utility function had to be logarithmic, meaning that it only covered the case where investors are risk averse. His most important assumption was that utility is both normative and descriptive. This simply means that "*An investor not only is obliged to choose between different goods to maximize his expected utility but also does so in reality*". This premise did not however explain why people choose to gamble. So, the main issue with Bernoulli's approach was it lacked generality.

Despite the fact that this finding was so important from a theoretical point of view it was not until 1944 that VN-M reinstated the theory around expected utility [87]. In fact, they laid down a set of sufficient and necessary axioms that the preferences of a decision maker need to satisfy in order to conclude that he makes his decisions based on the maximization of his expected utility. The utility theory is now well-established and covers all different types of risk attitude. Until this day, the MEUC is considered the new canon of economic theory.

However, in reality, we cannot apply direct utility maximization of an investor. The reason is that it is almost impossible to be aware of his exact utility function. Thus, all that we could hope for would be to know only some of the characteristics of the utility function of the investor, such that the utility function is increasing and concave (risk-averse investor). That is why highly prestigious works like that of Markowitz on Modern Portfolio Theory (1952) [64] as well as that of Arrow (1965) [4] and Pratt (1964) [76] focus on the first and second derivatives of the utility function of an investor. This is where the SDRs come to place. These rules constitute well-defined theorems determining the necessary and sufficient conditions under which an investor with specific risk preferences maximizes his expected utility. These conditions have to do with specific properties of the distribution of lotteries.

What we are going to focus more on is a different rule called the MVC developed by Markowitz [64], [65] (1952, 1959). This rule is a decision making criterion based only on the first two moments of the distributions of two lotteries. We shall discuss its value as well as the root causes for its criticism, like in Gandhi (1981) [30]. The main source of interest is: *"How this criterion differs from SDRs and when should an investor use it"*. We will see that there are some widespread misunderstandings concerning the necessary and sufficient conditions under which MVC coincides with MEUC. For example, there is a general agreement in the literature that the MVC is meaningful under two scenarios: (i) either lotteries are normally distributed, or (i) the investor has quadratic utility. As Baron (1977) [9] states, these assumptions are sufficient to justify the use of mean variance analysis in a manner consistent with the Von Neumann-Morgenstern axioms. Delving into the theory and the vast existing literature, we will see that there are misconceptions regarding both of these assumptions which need to be discussed. In particular, as we shall see there are also more interesting assumptions that make the MVC coincide with the MEUC that we will thoroughly analyze. Then, we are going to focus on the main subject of this work which has to do with the following idea from Levy and Markowitz (1979) [57] "*Assuming we have an approximately quadratic utility function, MV choices will almost* *maximize our expected utility function*". We will revisit this premise and will try to give our view based on MC Simulations and a thorough analysis of the idea. Although this subject is relatively old, the academic interest is still vivid. To name but a few of the most recent works, Markowitz (2010,2014) [63], [66], Malavasi et al. (2020) [60][1] and Schuhmacher et al. (2021) [80].

In the following sections, we are going to approach this subject carefully so as to not leave any unanswered questions. We start by setting the theoretical framework as constructed by VN-M. Based on it, we will define the MVC as well as the SD rules. From there on, we will delve into the necessary and sufficient conditions that connect the MVC with the SD rules and hence with the MEUC. We will study the assumptions of either normality or quadratic utility independently and discuss whether or not they are very restrictive. Next, we will discuss the more interesting cases of Elliptical and Skew-Elliptical families of distribution that seem to elevate the value of the MVC. Finally, we will examine the premise of Levy and Markowitz using MC Simulations together with a careful connection to the findings of the existing literature.

Our main findings split into two parts. (i) We show that, contrary to Elliptical distributions, under Skew-Elliptical distributions the MVC is not optimal for any risk-averse investor. In fact, our MC Simulations derive cases under which the MVC fails to deduce the right decision. (ii) The premise of Markowitz with respect to the approximately quadratic utility functions seems to be valid for Elliptical and Skew-Elliptical distributions. As we deviate more from normality (e.g. Extreme Value and Stable Pareto distributions) the set of approximately quadratic utility functions shrinks.

3.2 Theoretical Framework

An investor is given a number of different lotteries to choose from, each one with its own distribution. In order for the investor to make a rational choice his preferences should satisfy some specific axioms. Let $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$, where Z_i's are continuous random variables (lotteries). Let also, $\mathcal{P}$ be the set of probability distributions $F, G, Q : \mathcal{Z} \to [0, 1]$

[1]Malavasi et al. (2020) focus on the comparison between the SSD efficient set and the MV efficient set.

on $\mathcal{Z}$ (where F, G and Q represent the lotteries). We define a binary relation $\succeq$ on $\mathcal{P}$ representing the "preference" of the investor which should satisfy the following axioms

1. **Completeness**

$$F \succeq G \text{ or } G \succeq F$$

Investor is **obliged** to choose among F and G.

2. **Transitivity**

$$\text{If } F \succeq G \text{ and } G \succeq Q \Rightarrow \; F \succeq Q$$

3. **Continuity**
If $F \succ G \succ Q$, there exist $a, b \in (0, 1)$ such that

$$aF + (1 - a)Q \succ G \succ bF + (1 - b)Q$$

There are no "infinitely good" or "infinitely bad" prizes.

4. **Independence**
Let $a \in (0, 1)$. Then,

$$aF+(1-a)Q \succeq aG+(1-a)Q \Leftrightarrow F \succeq G$$

If we toss a coin between a fixed lottery Q and lotteries F and G our preference ($F \succeq G$) **should not** change. (Counterexample: **Allais paradox**)

According to VN-M, an investor's preferences will satisfy the above axioms if and only if his overall scope is to maximize his expected utility. In short, the investor will choose among different lotteries using the MEUC or equivalently the VN-M Representation Theorem.

Theorem 3.1 (MEUC)**.** *A relation $\succeq$ satisfies axioms 1-4 if and only if there exists a utility function $U : Z \to \mathbb{R}$, such that for every $F, G \in \mathcal{P}$*

$$F \succeq G \Leftrightarrow E[U(Z)] \geq E[U(Z)].$$

Moreover, U is unique up to a positive linear transformation, i.e. for some $a > 0$ and

$b \in \mathbb{R}$

$$\widetilde{U} = aU + b.$$

So, the MEUC implies that "*an investor with utility function U will prefer lottery F than lottery G if and only if his expected utility for F is larger than that of G*". The idea that the MEUC is the optimal investment criterion constitutes the cornerstone of financial economic theory. Based on it, we would like to go a step further and examine the preferences of a class of investors with respect to two lotteries.

For this, we need the notion of SD. More specifically, the SD definition exploits the MEUC in order to derive a conclusion for a specific set of investors U^*. Following Von Neumann-Morgenstern (1944) [87], we can define the notion of SD, as shown below.

Definition 3.2 (SD)**.** *Let two lotteries Z_1 and Z_2 with cumulative distribution functions F and G, respectively. We will say that Z_1 dominates Z_2 ($Z_1 D Z_2$) in U^*, or equivalently, F dominates G, FDG, if and only if*

$$E[U(Z_1)] \geq E[U(Z_2)],\ \forall U \in U^*$$

with a strong inequality for at least one $U_0 \in U^*$

The next step is to specify what types of investors could U^* include. In general, we can make the widely accepted and non-restrictive assumption that all investors are wealth maximizers ($U' \geq 0$), i.e. they belong to $U^* = \mathbf{U_1} = \{U : U' \geq 0\}$ and $U \in \boldsymbol{C^1}$[2]. For this set of investors, based on Definition 3.2, we can define the First-order of SD (hereafter; FSD). FSD is denoted by FD_1G.

The FSD involves the majority of investors, meaning that it applies SD on the largest class of investors possible. From that, we can proceed with narrowing the set of investors. The literature has strong evidence that the majority of investors are also risk-averse ($U'' \leq 0$), i.e. they belong to $U^* = \mathbf{U}_2 = \{U : U' \geq 0, U'' \leq 0\}$ and $U \in \boldsymbol{C^2}$[3]. So, based on Definition 3.2, it would be natural to define the Second-order Stochastic Dominance (hereafter; SSD) on $\mathbf{U}_2$. SSD is denoted by FD_2G.

Evidently, $\mathbf{U_2} \subset \mathbf{U_1}$ meaning that FSD implies SSD (i.e. $FD_1G \Rightarrow FD_2G$). Now, we can go even further to derive a Third-order of SD. But first, we need to give an intuitive

[2] $\boldsymbol{C^1}$: any function f for which the first derivative exists and is continuous.

[3] $\boldsymbol{C^2}$: any function f for which the first and second derivative both exist and are continuous.

description of the new set of investors. We can now include another empirically relevant case like those of wealth-maximizing and risk-averse investors. More specifically, as we shall see in Chapter 4, the empirical evidence has shown that investors exhibit DARA. Arrow (1965) [4] and Pratt (1964) [76] define ARA below,

$$r(x) = -\frac{U''(x)}{U'(x)}.$$

Since we need $r(x)$ to be decreasing we need

$$r'(x) = \frac{-U'''(x)U'(x) + (U''(x))^2}{(U'(x))^2} < 0,$$

provided that $U \in \boldsymbol{C^3}$. Based on that $U' \geq 0$ and $U'' \leq 0$, this is only possible if $U''' \leq 0$. Consequently, basd on Definitiion 3.2, the Third-order of Stochastic Dominance (hereafter; TSD) can be defined on $U^* = \mathbf{U_3} = \{U : U' \geq 0, U'' \leq 0, U''' \geq 0\}$ and $U \in \boldsymbol{C^3}$[4]. TSD is denoted by FD_3G.

Obviously, $\mathbf{U_3} \subset \mathbf{U_2} \subset \mathbf{U_1}$ meaning that FSD implies SSD and both imply TSD (i.e. $FD_1G \Rightarrow FD_2G \Rightarrow FD_3G$). Following the same rationale, we can narrow U^* even further but the aforementioned orders of SD are sufficient to discuss how the Markowitz's theory is connected to them.

Based on the definition of the SD, we could argue that we should stop right here, meaning that every time we need to examine whether or not one lottery stochastically dominates another one we should simply utilize the expected utility of each investor in U^* and see if the same inequality holds. However, if U^* is extremely large (as are the cases of $\mathbf{U_1}, \mathbf{U_2}$) it is impossible to check for each and every $U \in U^*$. How do we address this problem? By using the various SDRs as defined by Levy (1998) [54]. The idea is that any investor inside U^* can deduce SD between two lotteries, Z_1 and Z_2, if and only if he follows a specific decision rule that is defined based on the cumulative distributions of lotteries Z_1 and Z_2. This way we overcome the necessity of testing for each investor in U^* and focus only on the objective characteristics of the two lotteries under consideration.

Now, we are going to define the different types of SDRs. From Quirk and Saposnik (1962) [77], when $U^* = \mathbf{U_1}$ we can define the First-order SDR (hereafter; FSDR) as shown below.

[4] $\boldsymbol{C^3}$: any function f for which the first, second and third derivative exist and are continuous.

Theorem 3.3 (FSDR)**.** *For any two lotteries Z_1 and Z_2 with cumulative distributions F and G, FD_1G for all wealth maximizers $U \in \mathbf{U}_1$ ($U' \geq 0$) if and only if $F(x) \leq G(x)$ for all values x, and there is at least some x_0 for which a strong inequality holds. Namely,*

$$\underset{\forall x,\ \textit{with a strong inequality for at least one } x_0}{F(x) \leq G(x)} \quad \Leftrightarrow \quad \underset{\forall U \in \mathbf{U}_1,\ \textit{with a strong inequality for at least one } U_0 \in \mathbf{U}_1}{E[U(Z_1)] \geq E[U(Z_2)]} \tag{3.1}$$

FSDR says that as soon as the decision maker has an increasing utility function and needs to choose between two lotteries with cumulative distributions F and G, he will maximize his expected utility function by choosing the lottery with the smaller cumulative distribution function, i.e. F. This order of SD encompasses almost every rational investor, irrespective of his preferences. Two important necessary conditions for FSDR found in Levy (1998) [54] are shown below.

Condition 3.4. *(Necessary) If FD_1G, then the expected value of F must be greater than the expected value of G. $\mu_1 > \mu_2$ is a necessary condition for FSD. Equivalently,*

$$FD_1G \Rightarrow \mu_1 > \mu_2.$$

Condition 3.5. *(Necessary) If FD_1G, then the left tail of G must be "thicker". Equivalently,*

$$FD_1G \Rightarrow \min_F(x) \geq \min_G(x).$$

Let us turn now to the SSDR. So, under the additional assumption that investors are risk-averse Hadar and Russell [33] (1969) derive the SSDR, as shown below.

Theorem 3.6 (SSDR)**.** *For any two lotteries Z_1 and Z_2 with cumulative distributions F and G, FD_2G for all risk-averters $U \in \mathbf{U}_2$ ($U' \geq 0, U'' \leq 0$) if and only if*

$$\int_a^x [G(t) - F(t)]dt \geq 0, \quad \forall x \in [a, b],$$

with a strict inequality for at least one x_0. Equivalently,

$$\underset{\forall x,\ \textit{with a strong inequality for at least one } x_0}{\int_a^x [G(t) - F(t)]dt \geq 0} \quad \Leftrightarrow \quad \underset{\forall U \in \mathbf{U}_2,\ \textit{with a strong inequality for at least one } U_0 \in \mathbf{U}_2}{E[U(Z_1)] \geq E[U(Z_2)]} \tag{3.2}$$

SSDR says that as soon as the decision maker has an increasing and concave utility function and needs to choose between two lotteries with cumulative distributions F and

G, he will maximize his expected utility function by choosing the cumulative distribution that has a larger area for at least one point $x_0 \in [a, b]$, i.e. F.

As in the case of FSDR, Levy (1998) [54] derives the the necessary and sufficient conditions for SSDR.

Condition 3.7. *(Necessary) If FD_2G, then the expected value of F must be greater than or equal to the expected value of G. $\mu_1 \geq \mu_2$ is a necessary condition for FSD. Equivalently,*

$$FD_2G \Rightarrow \mu_1 \geq \mu_2.$$

Condition 3.8. *(Necessary) If FD_2G and $\mu_1 = \mu_2$ then the variance of F must be less than or equal to the variance of G. Equivalently,*

$$FD_2G \text{ and } \mu_1 = \mu_2 \Rightarrow \sigma_1 \leq \sigma_2.$$

Condition 3.9. *(Necessary) If FD_2G, then the left tail of G must be "thicker". Equivalently,*

$$FD_2G \Rightarrow \min_F(x) \geq \min_G(x).$$

So, first of all, FSD implies SSD which is logical as the set of investors that satisfy $U' \geq 0$ is a superset of those investors with the additional constraint of $U'' \leq 0$. Secondly, compared to Condition 3.4, Condition 3.7 does not require a strict inequality and we see from Condition 3.8 that under concave utility functions we derive an extra necessary condition concerning the variances of lotteries.

Last but not least, we can derive the Third-order SDR (hereafter; TSDR), which was introduced by Whitmore (1970) [88].

Theorem 3.10 (TSDR)**.** *For any two lotteries Z_1 and Z_2 with cumulative distributions F and G, FD_3G for all risk-averters with DARA $U \in \mathbf{U}_3$ $(U' \geq 0, U'' \leq 0, U'' \geq 0)$ if and only if*

$$(i) \quad \int_a^x \int_a^z [G(t) - F(t)]dtdz \geq 0, \ \forall x \in [a, b]$$

and

$$(ii) \quad \mu_1 \geq \mu_2,$$

with at least one strong inequality.

Below we get the necessary and sufficient conditions for TSDR as found in Levy (1998) [54].

Condition 3.11. *(Necessary) If FD_3G, then the expected value of F must be greater than or equal to the expected value of G. $\mu_1 \geq \mu_2$ is a necessary condition for FSD. Equivalently,*

$$FD_3G \Rightarrow \mu_1 \geq \mu_2.$$

Condition 3.12. *(Necessary) If FD_3G and $\mu_1 = \mu_2$ then the variance of F must be less than or equal to the variance of G. $\sigma_1 < \sigma_2$ is a necessary condition for FSD. Equivalently,*

$$FD_3G \text{ and } \mu_1 = \mu_2 \Rightarrow \sigma_1 < \sigma_2.$$

Condition 3.13. *(Necessary) If FD_3G and $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$ then the skewness of F, namely s_1, must be greater than the skewness of G, namely s_2. $s_1 > s_2$ is a necessary condition for FSD. Equivalently,*

$$FD_3G \text{ and } \mu_1 = mu_2 \text{ and } \sigma_1 = \sigma_2 \Rightarrow s_1 > s_2.$$

Condition 3.14. *(Necessary) If FD_3G, then the left tail of G must be "thicker". Equivalently,*

$$FD_3G \Rightarrow \min_F(x) \geq \min_G(x).$$

Condition 3.12 does not require equality between variances and Condition 3.13 says that if means and variances remain unchanged then FD_3G necessitates that lottery Z_1 is more positively skewed than lottery Z_2.

Each of these rules pertains to a specific class of investors. The higher the order of the SDR the narrower the set of investors. We saw that each of these rules deduces SD only by focusing on the objective characteristics of the two lotteries and not on the subjective characteristics of each investor. Moreover, these rules do not specify the types of distributions the two lotteries follow.

The theory presented in this section can be better interpreted through a concise example. In the following example we assume that Z_1 and Z_2 are discrete random variables. Let the distribution of lotteries Z_1, Z_2 be defined as in the following table.

Table 3.1: Example

Z_1	$P(Z_1 = z_1)$	Z_2	$P(Z_2 = z_2)$
5	0.4	10	0.4
10	0.6	20	0.6

where z_1 and z_2 are the outcomes of Z_1, Z_2, respectively.

So, the means and variances of the two lotteries are, respectively,

$$\mu_1 = 8, \quad \sigma_1^2 = 6$$
$$\mu_2 = 16 \quad \sigma_2^2 = 24.$$

From the above calculations, we observe that μ_2 is twice as large as μ_1 while at the same time σ_2^2 is quadruple of σ_1^2. Can lottery Z_1 dominate by any order of SD lottery Z_2? The answer is no. Because, we saw that under any order of SD a necessary condition is $\mu_1 > \mu_2$ (for FSD) and $\mu_1 \geq \mu_2$ (for SSD and TSD). So, it is Z_2 that might stochastically dominate Z_1. To determine the order of SD we need to derive the probability distributions of Z_1 and Z_2. Namely,

Table 3.2: Example

Z_1	$F(z)$	Z_2	$G(z)$
5	0.4	5	0
10	1	10	0.4
20	1	20	1

So, we deduce that $G(z) \leq F(z)$, $\forall z$, which means that $Z_2 D_1 Z_1$. As a result, despite the fact that the variance of lottery Z_2 is four times as large as that of lottery Z_1, any investor inside $\mathbf{U_3} \subset \mathbf{U_2} \subset \mathbf{U_1}$, regardless of his level of risk-aversion, will prefer Z_2.

This last example not only helps in better grasping the concept of SD rules but also indicates that we could probably derive other decision rules that are based on the moments of distributions. A moment-based criterion could simplify even more the decision making of an investor. However, Liu (2004) proved the following theorem.

Theorem 3.15. *There is no specific set of moment relationships between the first n moments of lotteries Z_1, Z_2 with cumulative distribution functions F and G that determines, FD_1G, or FD_2G, or FD_3G.*

In other words, according to Theorem 3.15, we should not expect finding any connection between a moment-based criterion and one of the SDRs, without making any further assumption with respect to the risk preferences of the investor or the specific type of the distribution of lotteries Z_1, Z_2. An interesting case that we are going to analyze, extensively, is that of the MVC, introduced by Markowitz (1952, 1959) [64] and [65].

3.2.1 Mean-Variance Criterion

The notions we discussed in the previous section are so fundamental that any further theoretical finding should be consistent with the MEUC to be meaningful. In his doctorate thesis in 1952 [64], Markowitz proposed a new criterion (rule) with respect to an investor's decision making. This criterion, known as the MVC, suggested that all that the investor needs to know in order to decide between two different lotteries is their first two moments. Particularly, the investor needs to either maximize the expected value of his chosen portfolio (lottery) for a specific level of risk, or equivalently, minimize the risk of the portfolio for a specific level of return. This is formulated as shown below.

Definition 3.16 (MVC)**.** *Let Z_1 and Z_2 be two lotteries with means μ_1, μ_2 and standard deviations σ_1, σ_2, respectively. Then, Z_1 will satisfy the MVC with respect to Z_2, $Z_1 MV Z_2$, if and only if*

1. $\mu_1 \geq \mu_2$
2. $\sigma_1 \leq \sigma_2$

It is important to detect that through the above definition we do not have any information about the kind of investors that would prefer lottery Z_1 instead of lottery Z_2. Before discussing the details around this rule we should first highlight why it is so important. The main reason is that if we could specify the class of investors for which this rule is optimal we would only need the first two moments of the two lotteries. So, any further information with respect to the distributions of the two lotteries would simply be irrelevant.

According to Markowitz, during his doctoral defence he received a rather interesting comment from Milton Friedman, that "*his contribution was not economics*". Potentially, what drove Friedman to make that statement was that Markowitz, at that point, had only made a suggestion that this criterion is meaningful for investors with an expected utility depending only on μ and σ, with an increasing and a decreasing relation respectively. A theoretical result was still needed to be found to justify the connection between the MVC and the MEUC. In other words, although his idea was intuitive, in the sense that we would expect that (risk-averse) investors desire higher means and detest higher variances, the set of investors for which the MVC coincides with the MEUC was still unknown.

Consider the following example. Assume that we have two risk-averse investors one with $U_1(x) = \ln x$ and the other one with $U_2(x) = \sqrt{x}$. Let also, two probability distributions F and G of two discrete random variables Z_1, Z_2 defined as shown in the following table.

Table 3.3: Example

Z_1	$F(z_1)$	Z_2	$G(z_2)$
5	0.80	7	0.99
30	0.20	150	0.01

From Table 3.3, we get $\mu_1 = 10 > \mu_2 = 8.4$ and $\sigma_1^2 = 100 < \sigma_2^2 = 203$, and thus MVC is satisfied. Now, the expected utility functions derive

$$E[U_1(Z_1)] = 1.9678 < E[U_1(Z_2)] = 1.9766$$

$$E[U_2(Z_1)] = 3.0731 > E[U_2(Z_2)] = 2.9230.$$

So, we see that the investor who has a logarithmic utility function will not decide based on the MVC, since if he does so he will select the wrong lottery. On the other hand, the investor with the square root utility function should decide based on the MVC for that specific MV-pair. This example, suffices to conclude that the MVC is not optimal for all risk-averse investors. In the following paragraphs we will discuss which types of investors did Markowitz have in his mind.

The above concerns, drove Markowitz to formulate the MVC under the VN-M theoretical framework in 1959 [65]. He developed his idea by assuming three rational investors each with his own utility function. The first one having a logarithmic utility $\ln(1+R_p)$, the second one the square root utility $\sqrt{1+R_p}$ and the third one the cubic root utility $\sqrt[3]{1+R_p}$, where R_p represents the portfolio return. A common characteristic of these three utility functions is that they are all increasing and concave, meaning that Markowitz focused on risk-averse investors. He then applied a Taylor expansion of 2nd-order around 0, which is defined as

$$U(R_p) \simeq U(0) + U'(0)R_p + \frac{1}{2}U''(0)R_p^2.$$

By applying the above Taylor expansion to each one of the three utility functions we obtain

$$\begin{aligned}
U_1(R_p) &= \ln(1+R_p) \\
&\simeq R_p - \frac{1}{2}R_p^2 \\
U_2(R_p) &= \sqrt{1+R_p} \\
&\simeq 1 + \frac{1}{2}R_p - \frac{1}{8}R_p^2 \\
U_3(R_p) &= \sqrt[3]{1+R_p} \\
&\simeq 1 + \frac{1}{3}R_p - \frac{1}{9}R_p^2
\end{aligned}$$

So, each utility function is now expressed as a quadratic approximation. Also, if we apply the expected values on each utility we get

$$\begin{aligned}
E[U_1(R_p)] &\simeq \mu_p - \frac{1}{2}(\mu_p^2 + \sigma_p^2) \\
E[U_2(R_p)] &\simeq 1 + \frac{1}{2}\mu_p - \frac{1}{8}(\mu_p^2 + \sigma_p^2) \\
E[U_3(R_p)] &\simeq 1 + \frac{1}{3}\mu_p - \frac{1}{9}(\mu_p^2 + \sigma_p^2)
\end{aligned}$$

So, all expected utility functions become a function of only the mean and variance of portfolio returns. Markowitz (2010) [66] reasoned that under a no short-selling assumption (i.e. restricting portfolio returns from getting below -100%), "*for a relatively large range of returns the quadratic approximations are very similar to the respective utility functions*". In fact, the following table suggests that.

Table 3.4: Quadratic Approximations of $\ln(1+R_p)$, $\sqrt{1+R_p}$, $\sqrt[3]{1+R_p}$

R_p	$\ln(1+R_p)$	Approx.	$\sqrt{1+R_p}$	Approx.	$\sqrt[3]{1+R_p}$	Approx.
−60%	−.92	−.78	.63	.66	.74	.76
−50%	−.69	−.63	.71	.72	.79	.81
−40%	−.51	−.48	.77	.78	.84	.85
−30%	−.36	−.35	.84	.84	.89	.89
−20%	−.22	−.22	.89	.90	.93	.93
−10%	−.11	−.11	.95	.95	.97	.97
0%	.00	.00	1.00	1.00	1.00	1.00
10%	.10	.10	1.05	1.05	1.03	1.03
20%	.18	.18	1.10	1.10	1.06	1.06
30%	.26	.26	1.14	1.14	1.09	1.09
40%	.34	.32	1.18	1.18	1.12	1.12
50%	.41	.38	1.22	1.22	1.14	1.14
60%	.47	.42	1.26	1.26	1.17	1.16
70%	.53	.46	1.30	1.29	1.19	1.18
80%	.59	.48	1.34	1.32	1.22	1.20
90%	.64	.50	1.38	1.35	1.24	1.21
100%	.69	.50	1.41	1.38	1.26	1.22

The blue coloured numbers represent "good" approximations, in that they differ from the real value of the utility about −.03 to .03

Indeed, the above results indicate the point made by Markowitz. Especially, square and third root utility functions are very well approximated by a quadratic. The interested reader can refer to the Appendix 3.5, at the end of this chapter, where we plot multiple different utility functions together with their 2nd-order Taylor approximation. But why did Markowitz develop his idea this way? The answer to this question comes in the form of the next theorem proven by Markowitz.

Theorem 3.17 (Markowitz 1959)**.** *Let $E[R_p] = \mu_p$ and $E[f(R_p)]$, where R_p represents the portfolio returns and $f(R_p)$ a rule which associates a number f to each value of R_p. An individual maximizes the expected value of a utility function $U(R_p) = aR_p + bf(R_p)$, if and only if*

(i) he maximizes the expected value of some utility function, and

(ii) his preferences are based solely on μ_p and $E[f(R_p)]$.

The above theorem basically states that an investor with a quadratic utility function that maximizes his utility function will act on the basis of $E[R_p]$ and $E[R_p^2]$. The converse is also true. An investor that bases his preferences solely on $E[R_p]$ and $E[R_p^2]$ and maximizes his expected utility implies that he has a quadratic utility function. So, the appropriate set of investors to which Markowitz was referring to, was the set of quadratic utilities, or as we saw in the previous analysis the set of approximately quadratic utility functions. In the following sections, we will further analyze this subject.

From this point on, the literature has done an extensive amount of research on Markowitz's framework. It is generally argued that the MVC is meaningful under two alternative hypotheses; (i) the investor's preferences are characterized by a quadratic utility function or, (ii) the distribution of returns is normal. Both of these cases have been shown to be unrealistic and so in general problematic (see Arrow (1965) [4] and Pratt (1964) [76] for further details). Starting with the assumption of quadratic preferences we actually defer to increasing ARA, which is contrary to the empirical evidence of decreasing ARA (see Chapter 4). With respect to the Gaussian nature of returns, many empirical findings like those in [58], [61] and [62] have shown that returns are far from normal, displaying fat-tails, meaning that a Stable-Pareto family of distributions would be more appropriate to assume.

Now that we have properly set the building blocks that lead to Markowitz's MVC, we can now delve into the fine points of this decision rule. In the following sections, we are going to review and comment on the literature which examines the necessary and sufficient conditions under which the MVC becomes the optimal decision rule. Next, we will carefully test the idea of Markowitz that the MVC under any quadratic approximation of a utility function is almost equivalent to the MEUC.

3.2.2 MVC Relation to MEUC: With Known Distribution

Having defined the MVC as well as the different orders of SDRs, the question that arises naturally is whether or not there is some kind of connection between these SDRs and the MVC. The importance of this connection stems from the strong theoretical foundation of the SD rules, because if there is not some kind of connection between them and MVC, there is no base in using the MVC.

In 1969, Levy and Hanoch [35] focused on determining when the MVC becomes a necessary and sufficient condition for MEUC. Before them, there were the works of Tobin (1958) [86] and later Feldstein (1969) [27], who concentrated on the type of distribution that makes MVC coincide with MEUC. Tobin, suggested that for any two-parameter distribution MVC coincides with MEUC. The problem with his assertion was an assumption he made, first noticed in [27], in trying to prove it. That, for any two-parameter distribution function with μ and σ we can standardize, i.e.

$$Z = \frac{X - \mu}{\sigma}.$$

But this is not generally true, as for this to hold we need a distribution function for which the location and scale parameters are μ and σ, respectively. Some counterexamples of two-parameter distributions are the log-normal, Beta and Gamma. Feldstein noticed that and argued that Tobin's analysis works just for normal distributions. Levy and Hanoch, impacted by Tobin, revolved around the case where the MVC under any two-parameter distribution becomes necessary and sufficient. In fact, they highlighted that the MVC under any two parameter distribution is only a sufficient condition for MEUC. To see that, the following example from the paper of Hanoch and Levy will help.

Remark 3.18. *Let two random variables X and Y with continuous uniform distributions F and G (i.e. two-parameter distributions). Let also X have a constant density function in $x_1 \leq x \leq x_2$ and Y have a constant density function in $y_1 \leq y \leq y_2$, with $x_1 > y_1$ and $x_2 > y_2$. From this, $F(x) \leq G(x)$ and so F dominates G by FSD. Moreover,*

$$\mu_1 = \frac{1}{2}(x_1 + x_2) > \frac{1}{2}(y_1 + y_2) = \mu_2.$$

Thus, $\mu_1 > \mu_2$ is necessary for dominance. However, the relation between the variances of the two distributions plays no role, meaning that any wealth-maximizer will choose F even if he is highly risk-averse.

In order for MVC to become necessary and sufficient for MEUC, Hanoch and Levy showed that we need to have two two-parameter distributions with an intersection point. This is formalized in the following theorem.

Theorem 3.19. *Let F and G be two distinct distributions with means μ_1, and μ_2, and standard deviations σ_1, σ_2, respectively, such that $F(x) = G(y)$, for all x and y which satisfy $\frac{x-\mu_1}{\sigma_1} = \frac{y-\mu_2}{\sigma_2}$. Let $\mu_1 \geq \mu_2$, and $F(x_1) > G(x_1)$ for some x_1 (i.e., $F(x)$ and $G(x)$ intersect). Then, F dominates G for all concave $U(x)$, if and only if $\sigma_1 \leq \sigma_2$.*

The assumption that "$F(x) = G(y)$, for all x and y which satisfy $\frac{x-\mu_1}{\sigma_1} = \frac{y-\mu_2}{\sigma_2}$" is very restrictive, and at the same time it is hard to interpret. However, without it, it would be impossible to generalize for all two-parameter distributions.

Levy, knowing that the above theorem is rather complicated decided that he should restate the theorem under only normal distributions. In fact, in his book [53] (1998) the new theorem is structured as shown below.

Theorem 3.20. *Let two lotteries Z_1 and Z_2 whose cumulative distributions are F and G, respectively, with $Z_1 \sim \mathbb{N}(\mu_1, \sigma_1)$ and $Z_2 \sim \mathbb{N}(\mu_2, \sigma_2)$. Then, F dominates G by SSD if and only if F dominates G by the MVC with at least one strong inequality.*

The nice properties of normal distributions as well as the assumption of "at least one strong inequality" capture the intersection between the two distributions. So, under normality the MVC coincides with the SSDR. Practically this means that any risk averse investor ($U' > 0,\ U'' < 0$), assuming that lotteries are normally distributed should make a decision based either on MVC or SSDR. This explains why the literature insists on the assumption of normality when referring to the MV framework.

However, in the end of Chapter 6 Levy notes: "*actually, the MV coincides with the elliptic family of distributions where the normal distribution belongs to this family*". This statement needs to be investigated thoroughly when it comes to what are the necessary and sufficient conditions for MVC to coincide with MEUC, under any elliptical distribution.

Feldstein argued that one important mistake that Tobin made was that he assumed that any linear combination of random variables following a two-parameter distribution follows the same two-parameter distribution. However, Feldstein pointed that a linear combination of normally distributed random variables remains normal but if we take for example a Gamma distribution any linear combination will have a one-parameter distribution with equal mean and variance. So, Feldstein concluded that the only admissible candidate is a normal distribution. However, Agnew (1971) [1] wrote a comment on Feldstein's assertion, claiming that Tobin's Separation Theorem is valid also for non-normal distributions. In particular, he asks the question "*If* $X_1, \ldots, X_n$*, are random variables with finite second moments and if all non-trivial linear combinations* $a_1X_1 + \ldots + a_nX_n$*, have the same distribution except for location and scale, then that distribution must be normal. True or false?*". Agnew basically argues that for normality to be the only candidate, the random variables should be stochastically independent. Otherwise, even for uncorrelated random variables the above assertion is false, i.e. there are non-normally distributed random variables that their linear combinations follow the same distribution. A specific example is the standardized bilateral exponential distribution (or else Laplace distribution with $\mu = 1$ and $\beta = 1$), which belongs to the elliptic family of distributions.

Later, Chamberlain in 1983 [17] (also Owen-Rabinovicth [70]) introduced two theorems, regarding the relation between elliptical distributions and the MVC, giving substance to Agnew's assertion. But first we need to define the spherical and elliptical distributions.

Definition 3.21 (Spherical distributions)**.** *A random vector* X *of dimension* n *is* ***spherically distributed*** *about the origin if its probability density function* f *satisfies the following*

$$f(X) = f(MX),$$

where $M^\top M = MM^\top = I_n$*.*

Equivalently, a spherical distribution is invariant under orthogonal linear transformations that leave the origin fixed. Likewise, an elliptical distribution is defined as shown below.

Definition 3.22 (Elliptical distributions)**.** *A random vector* $X(n \times 1)$ *is* ***elliptically distributed*** *if*

$$X = \mu + AY,$$

where $Y(k \times 1)$ *is a spherically distributed random vector,* A *is a* $(n \times k)$ *matrix such that* $AA^\top = \Sigma$ *(with* Σ *representing the scale matrix), and* $\mu(n \times 1)$ *is the location vector.*

Remark 3.23.

- *All symmetric elliptical distributions are symmetric around μ. So,*

$$E[(X-\mu)^i] = 0, \text{ for } i = 3, 5, 7\dots$$

- *All symmetric elliptical distributions are determined exactly by their mean and variance*

- *Any linear combination of elliptically distributed variables is still elliptical*

- *Under elliptical distributions, variance measures risk.*

- *Some elliptical distributions are: Normal, Student's t, Laplace, Logistic, Exponential, etc.*

Now, the first theorem of Chamberlain, considers the case where portfolio returns are made up of risky assets and a risk-free asset. Namely,

Theorem 3.24 (MV-utilities under elliptical distributions)**.** *The distribution of portfolio returns $R_p = w'R + (1-w)R_f$ is determined by its mean μ_p and variance σ_p^2 for every w if and only if there is a non-singular matrix T such that*

$$z = T(R-\mu),$$

is spherically distributed about the origin.

We saw that a linear transformation of a spherical random vector is elliptically distributed, which means that the asset returns R are elliptically distributed and since any linear combination of elliptical distribution is also elliptical that also makes R_p being elliptically distributed. Thus, the above theorem states that if there is a riskless asset in the investor's portfolio and the distribution of the risky assets is elliptical, the distribution of the portfolio's returns will be determined only by μ_p and σ_p. Accordingly, that derives the following result,

$$E[U(R_p)] = f(\mu_p, \sigma_p),$$

for some concave f.

Does that implicate that the MVC coincides with the MEUC? The answer is no. In order for this to hold, we should prove that under elliptical distributions the following equivalence holds

$$Z_1 MV Z_2 \Leftrightarrow F D_2 G.$$

Fortunately, the "necessity" side has been proven by Chamberlain, namely $Z_1 MV Z_2 \Rightarrow F D_2 G$. In particular, Chamberlain showed that for any concave utility function, i.e. $U \in \mathbf{U_2}$, the expected utility is increasing in mean and decreasing in variance. In other words, the MVC implicates the MEUC. But what about the "sufficiency" side, namely $F D_2 G \Rightarrow Z_1 MV Z_2$? For this, we will need to make use of Conditions 3.7 and 3.9. Condition 3.9, known as the left-tail necessary condition for the SSD, entails that $\sigma_1 \leq \sigma_2$, since we are talking about elliptical distribution which are known to be determined by their mean and variance. Accordingly, Condition 3.7 entails that $\mu_1 \geq \mu_2$. Thus, we get also the "sufficiency" side. And so now, we can claim that under elliptical distributions

$$Z_1 MV Z_2 \Leftrightarrow F D_2 G.$$

So, for any two lotteries which are elliptically distributed, the optimal rule for a risk-averse investor is the MVC. In other words, the Theorem 3.20 can be restated as shown below.

Theorem 3.25 (MVC-SSD under elliptical distributions)**.** *Let two lotteries Z_1 and Z_2 with F and G denoting their cumulative distributions, respectively. Let also Z_1, Z_2 be elliptically distributed with means μ_1, μ_2, respectively, and, standard deviations σ_1, σ_2, respectively. Then, F dominates G by SSD if and only if F dominates G by the MVC with at least one strong inequality.*

The above theorem states that, regardless of the distribution being normal, or logistic, or Laplace, or any other type of elliptical distribution, the investor should use the MVC to make his decisions. This elevates the value of the MVO method developed by Markowitz. However, in practice, we can see in the following remark that this family of distributions is quite limited since in order for the MVC to be meaningful we need skewness to be equal to zero.

Remark 3.26. *Some valid cases of symmetric elliptical distributions are:*

- *Student's-t: If* $df > 3$, *then* $\mu = 0$ $\sigma^2 = \frac{df}{df-2}$, $s = 0$, $\kappa = 3 + \frac{6}{df-4}$, *if* $df > 4$, *otherwise undefined*

- *Laplace:* $\mu = \mu$, $\sigma^2 = 2b^2$, $s = 0$, $\kappa = 6$, $b > 0$

- *Logistic:* $\mu = \mu$, $\sigma^2 = \frac{a^2\pi^2}{3}$, $s = 0$, $\kappa = \frac{21}{5}$, $a > 0$

- α*-stable:* $\mu = \mu$, $\sigma^2 = 2c^2$, $s = 0$, $\kappa = 3$, *if* $\alpha = 2$ *(Gaussian case)*

Duchin and Levy (2004) [55] conducted an empirical study to determine how this new finding from Chamberlain correlates with real data. They used monthly returns for 5 portfolios spanning from 1926 to 2001. Namely, common stocks, small stocks, long-term corporate bonds, long-term government bonds and Treasury bills. Then, they tested which of the following candidate distributions: Normal, Beta, Exponential, Extreme value, Gamma, Logistic, Lognormal, Student-t, Skew-Normal, Stable Paretian and Weibull, best fits the data. They found strong evidence pointing to the logistic distribution, which belongs to the symmetric elliptical family of distributions. Based on Theorem 3.25, they argued that this indicates that the MVC is the optimal decision rule for these portfolios.

Although it is clear that the MVC is optimal under elliptical distributions, many research papers and academic books still consider the MV-framework only under either quadratic preferences or normality. Markowitz (2010) [66] has observed that and emphatically states: "*I never-at any time!-assumed that return distributions are Gaussian*". True, the literature has often misinterpreted under what conditions the MVC is valid, but even under elliptical distributions the MVC is still far from being truly useful when dealing with real stock or portfolio returns. More specifically, families of distributions which contain more non-normal cases are more interesting, since they are known to describe better empirical data.

One very recent work from Schuhmacher et al (2021) [80] tries to broaden the family of distributions for which the MVC is relevant. The authors show that, in the presence of a risk-free asset, the return distribution of every portfolio is determined by its mean and variance if and only if asset returns follow a specific Skew-Elliptical distribution. A Skew-Elliptical distribution is defined as shown below.

Definition 3.27 (Skew-Elliptical GLS distributions). *A random vector X of dimension n is said to have a Skew-Elliptical generalized location-scale (hereafter; GLS) distribution with constant $r \in \mathbb{R}$, if its components X_i $(i = 1, \ldots, n)$, can be written as*

$$X_i = r + \beta_i Y + \gamma_i Z_i,$$

where, conditional on Y, the vector $Z = (Z_1, \ldots, Z_n)'$ is spherically distributed and Y is a real-valued random variable with $E[Y] \neq 0$ and $Var[Y] = 1$. The coefficients β_i, γ_i are real numbers with $\beta_i \neq 0$ for at least one $i = 1, \ldots, n$. Also, Z and Y are linearly independent.

Remark 3.28.

- *All Skew-Elliptical distributions are determined exactly by their mean and variance. More specifically, based on Definition 3.27 we derive*

$$\begin{aligned} E[X_i] &= r + \beta_i E[Y] + \gamma_i E[Z_i] = r + \beta_i E[Y] \\ Var(X_i) &= \beta_i^2 Var(Y) + \gamma_i^2 Var(Z_i) = \beta_i^2 + \gamma_i^2. \end{aligned}$$

Solving for β_i and γ_i we get

$$\begin{aligned} \beta_i &= \frac{E[X_i] - r}{E[Y]} \\ |\gamma_i| &= \sqrt{Var(X_i) - \left(\frac{E[X_i] - r}{E[Y]}\right)^2}. \end{aligned}$$

- *Any linear combination of Skew-elliptically GLS distributed variables is still Skew-Elliptical*

- *Some Skew-Elliptical distributions are: Skew-Normal, Skew-t, Skew-Cauchy, Skew-logistic, etc.*

Similar to Chamberlain (1983) [17], Schuhmacher et al. (2021) [80] proved the following theorem.

Theorem 3.29 (MV-utilities under Skew-Elliptical distributions). *Assume there exists at least one $i = 1, \ldots, n$ such that $E[R_i] \neq R_f$, where R_i is the ith element of the risky* *asset vector R. In the presence of a risk-free asset, R_f, the distribution of portfolio returns $R_p = w'R + (1 - w)R_f$ is determined by its mean and variance for every $w \in \mathbb{R}^n$ with $w'\mathbb{1} = 1$ if and only if the asset returns R have a Skew-Elliptical GLS distribution.*

Theorem 3.29 states that lotteries which follow a Skew-Elliptical distribution have a MV-utility. So, we derive the following result

$$E[U(R_p)] = f(\mu_p, \sigma_p).$$

Following the same rationale as in Chamberlain's work, the fact that the expected utility is only a function of the mean and the variance of the portfolio returns does not implicate that the MVC is necessary and sufficient for the MEUC. In other words, one needs to prove that under skew-elliptical distributions the following holds

$$Z_1 MV Z_2 \Leftrightarrow F D_2 G.$$

Contrary to Chamberlain (1983) [17], Schuhmacher et al. (2021) [80] do not show that f is increasing in mean and decreasing in variance, for any $U \in \mathbf{U_2}$. So, with regards to the "sufficiency" side, this might mean that FD_2G does not necessarily implicate $Z_1 MV Z_2$. Moreover, when it comes to proving the "necessity" side we cannot make use of Condition 3.9, as we did earlier for the elliptical family of distributions, since this condition only applies to distributions that are not skewed. In fact, later on we will see through MC Simulations that under Skew-Elliptical distributions there are some cases that violate the "necessity" side. So, contrary to elliptical distributions, there is no theoretical proof that under Skew-Elliptical distributions the MVC is the optimal decision rule for any risk-averse investor. We only know that under Skew-Elliptical distributions the expected utility of the investor is a function of mean and variance. As a result, there might be cases in which even though the MVC is satisfied between two lotteries, namely $Z_1 MV Z_2$, some type of investor inside $\mathbf{U_2}$ might prefer lottery Z_2. This will become evident in our MC Simulations in the Quadratic approximations subsection.

3.2.3 MVC Relation to MEUC: With Known Preferences

An alternative to searching for a good candidate distribution is to make an assumption on the utility function of the investor. A widely used premise is that of quadratic utility.

In that case, the expected utility becomes a function of only μ and σ. Moreover, it is increasing in μ and decreasing in σ. The price we pay for this kind of assumption is that, (i) quadratic preferences constitute a very restrictive class and, (ii) by assuming quadratic utility we are led to increasing absolute risk aversion (ARA), which is counter-intuitive. So, one should be careful when trying to avoid an assumption with respect to the distribution of returns, as he will be left with a class of utility functions that is questionable for its realism as well as for its usefulness. However, this is the specific class of investors that Markowitz pointed to.

Hanoch and Levy (1969) [35], presented an example through which they argued that the MVC under quadratic preferences is only sufficient for MEUC. This, can be formally shown through the following proposition from Hanoch and Levy (1970) [34].

Proposition 3.30. *Assuming quadratic preferences, the MVC is only a sufficient condition for the MEUC*

Proof. (Sufficiency) Following [34], let the following quadratic utility function

$$U(x) = 2Kx - x^2, \text{ with } x < K$$

where $K > 0$, $U'(x) = 2(K - x) > 0$ and $U''(x) = -2 < 0$. Let two lotteries x_1 and x_2 for which we derive

$$\begin{aligned}
\Delta E[U] &= E[U(x_1)] - E[U(x_2)] \\
&= 2K\mu_1 - E[x_1^2] - 2K\mu_2 - E[x_2^2] \\
&= 2K\mu_1 - (\mu_1^2 + \sigma_1^2) - 2K\mu_2 - (\mu_2^2 + \sigma_2^2) \\
&= 2K\Delta\mu - (\Delta\mu^2 + \Delta\sigma^2) \\
&= 2\Delta\mu(K - \bar{\mu}) - \Delta\sigma^2
\end{aligned}$$

where $\Delta\sigma^2 = \sigma_1^2 - \sigma_2^2$, $\bar{\mu} = \frac{\mu_1+\mu_2}{2}$. Assuming that $\mu_1, \mu_2 < K$, we have that $\bar{\mu} < K$. Then, $\Delta E[U] > 0$ if we assume that $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$, which is exactly the MVC.

(Necessity) Let $\Delta E[U] > 0$. Does that imply $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$? The answer is no. In fact, from $\Delta E[U] > 0$ we have that

$$\begin{aligned}
&2\Delta\mu(K - \bar{\mu}) > \Delta\sigma^2, \\
&\Delta\mu > 0
\end{aligned}$$

Thus, even if $\Delta\sigma^2 < 0$ (i.e. $\sigma_1 > \sigma_2$), since $2\Delta\mu(K - \bar{\mu}) > 0$, the above inequality holds. $\square$

The above proof led Hanoch and Levy to identify the right rule which is both necessary and sufficient for MEUC. The rule is called quadratic dominance rule and we can see below how it coincides with MEUC.

Theorem 3.31. *Assuming quadratic preferences, the quadratic dominance rule as defined below*

1. $\mu_1 \geq \mu_2$
2. $2\Delta\mu\Big(\max(x_1, x_2) - \bar{\mu}\Big) - \Delta\sigma^2 \geq 0,$

is both necessary and sufficient for MEUC.

The proof of this theorem is evident from the previous proof of the proposition. The reason the authors chose to replace K with $\max(x_1, x_2)$ is that in this way the rule constitutes a smaller set than if we had K. This set happens to be the smallest and thus the optimal set. The new rule under quadratic preferences is both necessary and sufficient for MEUC, mainly because it also includes cases where $\sigma_1 > \sigma_2$.

Johnstone et al. (2011) [41] proved that if one wants to avoid constraining the distribution of portfolio returns it is necessary to assume quadratic preferences to apply the MVC. This is formulated as shown below.

Theorem 3.32. *The use of MVC, on the class of all distributions, implies that the decision maker's utility function must be quadratic.*

This theorem basically says that under the MV-framework we should not look for any other set of utilities other than the quadratic, assuming that we do not constrain the family of distributions, which justifies the use of quadratic utilities. Still, the quadratic family of utilities is very restrictive. But in Markowitz's words (2010) [66]: "*Nor did I ever assume that the investor's utility function is quadratic*". So, although the literature has adopted the quadratic utility as the only appropriate class of investors for which the MV-framework is relevant, Markowitz claims that we should not be fixated just on the quadratic utility function. In particular, as we have already discussed, Markowitz attempted to upgrade the MVC by discussing its validity even under a wider class of utility functions that happen to be approximately quadratic. We will discuss this premise in the following section.

3.3 Methodology and MC Simulations

3.3.1 Approximately Quadratic Utility Functions

As we analyzed earlier, Markowitz (1959) chose three specific utilities in his work in order to discuss his idea about the quadratic approximations, namely $\log(1+Z)$, $\sqrt{1+Z}$ and $\sqrt[3]{1+Z}$. Those three utilities are not only concave but they also satisfy one additional property, namely $U''' \geq 0$. To see why this extra property is crucial, we have to take a 2nd-order Taylor series on the utility function, as shown below

$$Q_Z = U(\mu) + U'(\mu)(Z-\mu) + \frac{U''(\mu)}{2}(Z-\mu)^2$$
$$E[Q_Z] = U(\mu) + \frac{U''(\mu)}{2}\sigma^2$$

So, $E[Q_Z]$ will increase with respect to μ and decrease with respect to σ if the following holds

$$\frac{\partial E[Q_Z]}{\partial \sigma} = U''(\mu)\sigma < 0, \text{ if } U'' < 0$$
$$\frac{\partial E[Q_Z]}{\partial \mu} = U'(\mu) + \frac{U'''(\mu)}{2}\sigma^2 > 0, \text{ if } U' > 0 \text{ and } U''' \geq 0$$

Therefore, any utility function that is a part of $\mathbf{U_3} = \{U : U' > 0, U'' < 0, U''' \geq 0\}$ and at the same time is almost quadratic, will be increasing in mean and decreasing in variance. Are $\log(1+Z)$, $\sqrt{1+Z}$ and $\sqrt[3]{1+Z}$ almost quadratic? According to Markowitz's Table 3.4, we see that $\sqrt{1+Z}$ and $\sqrt[3]{1+Z}$ are approximately quadratic for any value around -60% and 100%, but with regards to $\log(1+Z)$ the quadratic approximation is good only for values around -40% and 50%. As a result, assuming that two lotteries Z_1, Z_2 take values in $[-60\%, 100\%]$ (or in $[-40\%, 50\%]$) and that $Z_1 MV Z_2$, we could say that the investors with utility functions $\sqrt{1+Z}$ and $\sqrt[3]{1+Z}$ (or $\log(1+Z)$) should prefer Z_1. Does the inverse also hold? Based on Proposition 3.30, we deduce that an approximately quadratic utility function can qualify the MVC to be only a sufficient condition for the

MEUC. So, according to Markowitz, without any further assumption on the distributions followed by Z_1 and Z_2, the MVC will be sufficient for the MEUC if and only if the utility function of the investor is approximately quadratic for a sufficiently wide range of values. In the end of this chapter there is an Appendix which includes several plots of different utility functions together with their quadratic approximation. Markowitz realized that this reasoning does not suffice to support his assertion. Thus, he resorted to a quite different approach.

In 1979, Levy and Markowitz [57] revisited this subject by doing an empirical analysis. Firstly, they restated Markowitz's premise as follows: "*an investor that chooses carefully from among the mean-variance efficient set, will almost maximize his expected utility, if and only if his utility function is approximately quadratic*", i.e. it is almost perfectly approximated by a 2nd-order Taylor expansion. Consequently, they introduced a way to identify the size of the set containing the approximately quadratic utility functions. Firstly, the utility functions have to be a part of $\mathbf{U_3}$. Such utility functions are the following, $\log{(1+Z)}$, $(1+Z)^a$ with $a = 0.1, 0.3, 0.5, 0.7, 0.9$ and $-e^{-a(1+Z)}$ with $a = 0.1, 0.5, 1, 3, 5, 10$. Secondly, they collected the annual returns of 149 mutual funds during the period 1958 through 1967. Following that, they calculated $Corr(E[U(Z)], E[Q_Z])$, for each of the above utility functions. The idea was simple. If the correlation of the expected utility and the expected value of the quadratic approximation is close to 1, that would indicate that $E[U(Z)]$ and $E[Q_Z]$ move in the same direction, which is the actual point of interest. This approach, overcomes the limitations in Markowitz's initial attempt to promote his idea of approximately quadratic utilities. Going back to their results, the authors found evidence of

$$Corr(E[U(Z)], E[Q_Z]) \simeq 1,$$

for all the parametrizations of the utility functions, except for $b = 5, 10$ which represent the extremely risk-averse investors. Additional empirical evidence came from their joint work with Kroll in 1984 [47]. Based on these findings, they argued that the above utility functions are almost quadratic for almost all of their parametrizations. Thus, if the MVC holds, the above investors should decide based on it.

In the following section we are going to thoroughly analyze our approach on this subject. Up to this point, the context of our discussion around the connection between the MVC and MEUC includes either an assumption with respect to the set of utility functions or the type of distribution of the lotteries. With that being said, although Levy and Markowitz did provide some supportive evidence of their premise, we believe that in order for it to be confirmed we need to clarify whether we need an extra assumption with respect to the type of distribution. Namely, since we are assuming approximately quadratic utility functions we are obliged to research on whether or not we need an extra assumption on the kind of distribution under which the MVC is sufficient for the MEUC. Otherwise, the premise about quadratic approximations cannot be strongly supported. For our analysis, we need to define $\mathbf{U_3^*} = \{U : U' > 0, U'' < 0, U''' \geq 0 \text{ and } U(Z) \simeq Q_Z\}$ to be the set that contains all those utility function that are part of $\mathbf{U_3}$ and at the same time are almost quadratic.

3.3.2 Methodology

An important question that needs to be addressed is under what conditions does the assumption of approximately quadratic utility functions hold. Or, in more general, what is the range of the MVC? Whether or not a utility function is well-approximated by a 2nd-order Taylor series should depend on the utility function we use (the subjective characteristics of the investors) as well as the type of the assumed distribution (the objective characteristics of the lotteries). A simple approach like that on Table 3.4 is inadequate. The reason is that non-normal or skewed distributions might conflict with quadratic approximations, in terms of the validity of the MVC. In other words, we should specify for which distributions we have $U(Z) \simeq Q_Z$. In a recent review of his work, Markowitz (2010) [66], claimed that the idea presented in Levy and Markowitz (1979) [57] was targeting any type of distribution. This last information helps us formulate a mathematical proposition connecting the MVC to approximately quadratic utility functions. Before doing that we need to highlight the following. First, we should take into consideration the fact that the MVC is a decision criterion between two lotteries. So, the premise of Markowitz should be restated accordingly. Second, based on Proposition 3.30, the MVC is only a sufficient condition for MEUC, for any quadratic utility function. Third, the set of investors we refer to is $\mathbf{U_3^*} = \{U : U \in \mathbf{U_3} \text{ and } U(Z) \simeq Q_Z\}$. Altogether, we get the following corollary.

Corollary 3.33 (MVC under Quadratic Approximation)**.** *For any two lotteries Z_1 and Z_2, with any cumulative distributions F and G, the following holds*

$$Z_1 MV Z_2 \Rightarrow E[U(Z_1)] \geq E[U(Z_2)],\ \forall U \in \mathbf{U_3^*} = \{U : U \in \mathbf{U_3} \text{ and } U(Z) \simeq Q_Z\}.$$

Proof. If $Z_1 MV Z_2$ then for any $U \in \mathbf{U_3^*}$ we have

$$\begin{aligned} E[U(Z_1)] - E[U(Z_2)] &\simeq E[Q_{Z_1}] - E[Q_{Z_2}] \\ &= U(\mu_1) + \frac{U''(\mu_1)}{2}\sigma_1^2 - U(\mu_2) - \frac{U''(\mu_2)}{2}\sigma_2^2 \geq 0 \end{aligned}$$

□

The above corollary states that for any two lotteries that satisfy the MVC, i.e. $Z_1 MV Z_2$, all investors inside $\mathbf{U_3^*}$ will maximize their expected utility functions by choosing lottery Z_1. But the success of Corollary 3.33 relies on $\mathbf{U_3^*}$ being sufficiently large. This is what we will try to determine. We already know from Theorem 3.25 that for any elliptical symmetric distribution the MVC becomes equivalent to the MEUC. So, for these types of distributions the additional limiting assumption of approximately quadratic utility functions is unnecessary. In other words, we would like to examine the validity of the above corollary for asymmetric distributions and even for very non-normal cases which are considered to characterize daily or even monthly stock returns. Jondeau and Rockinger (2006) [43], using empirical data, supported that cubic or even quartic approximations are better approximations of expected utility, under large departure from normality. But their work does not approach the work of Levy and Markowitz the way we do. As long as the quadratic approximation consistently results in the same decision making between two lotteries as the direct MEUC, there is no reason in searching for more precise approximations of the utility function. One way to research that is by applying MC Simulations. The simulations enable us to apply different types of distribution with specific characteristics. This way we can identify more clearly under what conditions the premise of Markowitz is valid.

3.3.3 MC Simulations

As we have already analyzed we only consider utility functions that belong to $\mathbf{U_3}$. The main reason being that a 2nd-order Taylor series, of any utility function inside $\mathbf{U_3}$ is only a function of μ and σ, while being increasing in μ and decreasing in σ. Moreover, any investor inside $\mathbf{U_3}$ prefers a utility function with higher positive skewness. Our main scope, is to calculate the percentage that MVC implies the MEUC, for utility functions belonging to $\mathbf{U_3}$. By doing that, we expect that skewness plays an important role in the decision making of the investors inside $\mathbf{U_3}$. This way, we can detect whether the MVC is sufficient for the MEUC or if the higher skewness is more desirable by the investors inside $\mathbf{U_3}$, leading to the failure of Corollary 3.33.

The utility functions we are going to use are a combination of the ones used by Levy and Markowitz (1979) and Ederington (1995). The utilities are presented in the following table.

Table 3.5: Utility functions inside $\mathbf{U_3}$

$(1+Z)^a$ with $a = \{0.01, 0.1, 0.5, 0.9\}$
$\log(a+Z)$ with $a = \{0.9, 1\}$
$-e^{-a(1+Z)}$ with $a = \{0.7, 1, 3, 5, 8, 10, 15, 20\}$
$-(1+Z)^{-a}$ with $a = \{0.01, 0.3, 0.5, 1, 3, 5, 8, 10, 15, 20\}$

Before going into the simulations we have to highlight a few things about the level of risk-aversion of each utility function. We can measure the level of risk-aversion of each utility function by the absolute risk aversion. Namely,

$$ARA_{(1+Z)} = \frac{1-a}{1+Z}$$
$$ARA_{log} = \frac{1}{a+Z}$$
$$ARA_{exp} = a$$
$$ARA_{-(1+Z)} = \frac{1+a}{1+Z}$$

In our case, lotteries Z represent stock returns which means that the range of values is $[-1, 1]$. So, in general we can sort the utility functions in terms of their level of risk-aversion as follows $(1+Z)^a$, $\log(a+Z)$, $-e^{-a(1+Z)}$ and $-(1+Z)^{-a}$, with the last one describing the more risk risk-averse investor. Parameter-wise the *log*-utility function characterizes the more risk-averse investors when a gets closer to 0.9. Accordingly, $(1+Z)^a$ is more risk-averse for a's closer to 0.01. For $-e^{-a(1+Z)}$ and $-(1+Z)^{-a}$ the higher a gets, the more risk-averse the investors are. One would expect that for more risk-averse investors the skewness of an asymmetric distribution together with the existence of more extreme jumps will impact their decision making.

Now, as we said, we are going to apply MC Simulations in order to have the absolute control in terms of the DGP that generates our data. The simulations enable us not only to choose the type of distribution that generates our data but also enable us to control the levels of differences between the parameters μ_1, μ_2 and σ_1, σ_2. Evidently, the larger the differences, μ_1/μ_2 and σ_1/σ_2 are, the less the effect of a more skewed or even a more non-normal distribution will be on the MVC's efficiency. Thus, the MC Simulations will act as a stress test on the premise of Levy and Markowitz, as we will consider specific cases under which the premise might fail even for less risk-averse investors. In the following section, we aim to measure exactly the efficiency rate of the MVC under some specific cases that we consider.

The methodology we will follow for the MC Simulations is analyzed in the following steps. First, we choose the distribution from which we will generate data for two lotteries, Z_1 and Z_2.

$$Z_1 \sim D(p_1) \quad Z_2 \sim D(p_2),$$

where p_1 and p_2 represent the parameters of each distribution.

In our analysis, we use five types of distribution, the Gaussian, the Laplace, the Skew-Normal, the Extreme Value and the Stable Pareto. These are considered good candidates as they are regularly used to fit multiple frequencies of stock returns. More specifically, Linden (2001) [59] finds evidence that the Laplace distribution fits well into a sample of daily and weekly observations of individual stocks. The Skew-Normal distribution is employed in Harvey et al. (2004) [36] to model multivariate returns. Levy and Duchin (2004) [56] find evidence that Extreme Value distributions fit best stock returns with longer horizons. Finally, there is empirical evidence found in Mandelbrot (1963) [61], Manegna and Stanley (1995) [62] and in Levy (2005) [58], that short-term stock returns are better described by a Stable Pareto distribution. The Gaussian as well as Laplace distributions are expected to derive a 100% success of the MVC inferring the MEUC, based on Theorem 3.20. The Skew-Normal represents an interesting case as it belongs to the skew-elliptical family that Schuhmacher et al. [80] were referring to. The other two distributions are gradually more skewed and in general more non-normal. So, the last three distributions are considered more interesting. The next step, is to control the differences between the means, the variances and the skewnesses of the two lotteries Z_1, Z_2. We generate data in such a way that we have absolute control on these differences. This is important as it makes it easier to see the effect of skewness on the investors' decision making. Moreover, the data we generate will always make sure that $Z_1 MV Z_2$ and not the other way. We replicate this step multiple times. The data we generate each time are approximately $100,000$ observations. For each distribution and each case of differences in the parameters we generate approximately 100 MV-pairs. So, we can be certain that the findings are robust.

We set the parameters for each type of distribution based on real data. Namely, we first fit the type of distribution we want on the daily stock returns of 850 different NYSE stocks. And then, we use these fitted parameters as a basis to generate our own data. This way we are more confident that the different cases we consider are realistic.

The results produced by the MC Simulations can be found in the following tables.

Table 3.6: Percentage of MVC⇒MEUC for $(1+Z)^a$

Distribution	Parameters Differences	$a=0.01$	$a=0.1$	$a=0.5$	$a=0.9$
Normal	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01$	100%	100%	100%	100%
Laplace	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01$	100%	100%	100%	100%
SkewN	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=3$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=3$	100%	100%	100%	100%
Extreme	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=3$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=3$	81%	86%	91%	94%
Stable	$1.1<\frac{\mu_1}{\mu_2}\leq 1.15,\ 1.1<\frac{\sigma_2}{\sigma_1}\leq 1.15,\ 1.5\leq\frac{s_2}{s_1}\leq 3$	96%	98%	100%	100%
	$1.05<\frac{\mu_1}{\mu_2}\leq 1.1,\ 1.05<\frac{\sigma_2}{\sigma_1}\leq 1.1,\ 1.5\leq\frac{s_2}{s_1}\leq 3$	94%	98%	100%	100%
	$1.01\leq\frac{\mu_1}{\mu_2}\leq 1.05,\ 1.01\leq\frac{\sigma_2}{\sigma_1}\leq 1.05,\ 1.5\leq\frac{s_2}{s_1}\leq 3$	76%	81%	95%	100%

Table 3.7: Percentage of MVC⇒MEUC for $log(a + Z)$

Distribution	Parameters Differences	$a = 0.9$	$a = 1$
Normal	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01$	100%	100%
Laplace	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01$	100%	100%
SkewN	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 1.5$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 3$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 1.5$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 3$	100%	100%
Extreme	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 1.5$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 3$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 1.5$	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 3$	81%	85%
Stable	$1.1 < \frac{\mu_1}{\mu_2} \leq 1.15,\ 1.1 < \frac{\sigma_2}{\sigma_1} \leq 1.15,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	93%	96%
	$1.05 < \frac{\mu_1}{\mu_2} \leq 1.1,\ 1.05 < \frac{\sigma_2}{\sigma_1} \leq 1.1,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	90%	94%
	$1.01 \leq \frac{\mu_1}{\mu_2} \leq 1.05,\ 1.01 \leq \frac{\sigma_2}{\sigma_1} \leq 1.05,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	62%	76%

Table 3.8: Percentage of MVC⇒MEUC for $-e^{-a(1+Z)}$

Distribution	Parameters Differences	$a = 0.7$	$a = 1$	$a = 3$	$a = 5$	$a = 8$	$a = 10$	$a = 15$	$a = 20$
Normal	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05$	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01$	100%	100%	100%	100%	100%	100%	100%	100%
Laplace	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05$	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01$	100%	100%	100%	100%	100%	100%	100%	100%
SkewN	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 1.5$	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 3$	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 1.5$	100%	100%	100%	100%	100%	100%	91%	18%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 3$	100%	100%	100%	100%	100%	100%	4%	0%
Extreme	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 1.5$	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2} = 1.05,\ \frac{\sigma_2}{\sigma_1} = 1.05,\ \frac{s_2}{s_1} = 3$	100%	100%	100%	100%	6%	0%	0%	0%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 1.5$	100%	100%	99%	80%	7%	0%	0%	0%
	$\frac{\mu_1}{\mu_2} = 1.01,\ \frac{\sigma_2}{\sigma_1} = 1.01,\ \frac{s_2}{s_1} = 3$	94%	49%	1%	0%	0%	0%	0%	0%
Stable	$1.1 < \frac{\mu_1}{\mu_2} \leq 1.15,\ 1.1 < \frac{\sigma_2}{\sigma_1} \leq 1.15,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	100%	98%	93%	88%	81%	74%	56%	52%
	$1.05 < \frac{\mu_1}{\mu_2} \leq 1.1,\ 1.05 < \frac{\sigma_2}{\sigma_1} \leq 1.1,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	98%	92%	82%	77%	59%	55%	41%	36%
	$1.01 \leq \frac{\mu_1}{\mu_2} \leq 1.05,\ 1.01 \leq \frac{\sigma_2}{\sigma_1} \leq 1.05,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	92%	73%	43%	38%	19%	14%	10%	3%

Table 3.9: Percentage of MVC⇒MEUC for $-(1+Z)^{-a}$

Distribution	Parameters Differences	$a=0.01$	$a=0.3$	$a=0.5$	$a=1$	$a=3$	$a=5$	$a=8$	$a=10$	$a=15$	$a=20$
Normal	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Laplace	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
SkewN	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=3$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%	100%	100%	100%	100%	82%	15%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=3$	100%	100%	100%	100%	100%	100%	100%	76%	0%	0%
Extreme	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	$\frac{\mu_1}{\mu_2}=1.05,\ \frac{\sigma_2}{\sigma_1}=1.05,\ \frac{s_2}{s_1}=3$	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=1.5$	100%	100%	100%	100%	92%	49%	0%	0%	0%	0%
	$\frac{\mu_1}{\mu_2}=1.01,\ \frac{\sigma_2}{\sigma_1}=1.01,\ \frac{s_2}{s_1}=3$	72%	68%	55%	37%	0%	0%	0%	0%	0%	0%
Stable	$1.1<\frac{\mu_1}{\mu_2}\leq 1.15,\ 1.1<\frac{\sigma_2}{\sigma_1}\leq 1.15,\ 1.5\leq\frac{s_2}{s_1}\leq 3$	96%	94%	92%	88%	86%	82%	70%	63%	52%	47%
	$1.05<\frac{\mu_1}{\mu_2}\leq 1.1,\ 1.05<\frac{\sigma_2}{\sigma_1}\leq 1.1,\ 1.5\leq\frac{s_2}{s_1}\leq 3$	94%	91%	88%	80%	78%	64%	54%	50%	38%	33%
	$1.01\leq\frac{\mu_1}{\mu_2}\leq 1.05,\ 1.01\leq\frac{\sigma_2}{\sigma_1}\leq 1.05,\ 1.5\leq\frac{s_2}{s_1}\leq 3$	76%	57%	48%	33%	28%	24%	17%	12%	8%	2%

Before getting into the results we need to highlight some important features of our generated data. Notice that for the Skew-Normal as well as the Extreme Value distribution the differences between the means and standard deviations we consider are small as we found out that for larger differences the MVC works fine. On the contrary, under a Stable Pareto distribution the MVC fails to entail the MEUC even for distances between means and standard deviations that are as high as 15%. This is due to the extreme characteristics of the Stable Pareto distribution, which is known to exhibit sudden large jumps and thus creates higher challenges for the MVC. In terms of the Stable Pareto distribution, the reason we allow the differences between the parameters to move inside a specific range is because it is harder to control the data produced, since the distribution has undefined moments.

Now, the MC Simulations indicate that $(1+Z)^a$ and $\log(a+Z)$ utility functions are only mildly affected by the Skew-Normal and Extreme Value distributions when the differences between the parameters are very close in value. This shows, that the less risk averse investors can generally trust the MVC for their decision making, even for mildly non-normal skewed distributions. But, in the Stable Pareto case when the differences between the means and the standard deviations hover around 1% and 10% the investors should take into consideration the skewnesses of the two generated processes in order to make better decisions. With regards to $-e^{-a(1+Z)}$ and $-(1+Z)^{-a}$ the issues with the sufficiency of the MVC for the MEUC are evident even in the case of the Skew-Normal distribution. More specifically, the extremely risk-averse investors with $a = 15, 20$ will make very wrong decisions when the means and standard deviations differences are very close in value. The results are even worse for the Extreme Value distribution case. In particular, we see that the impact on the MVC is evident even for less risk-averse investors and is far worse as the differences in the parameters get closer. But the more interesting results come from the Stable Pareto case, which signifies that these types of investors will need more information on the lotteries' characteristics, besides the means and variances, in order to make their decision. This is evident even in cases where the differences between the parameters get as high as 15%.

As our results indicate, the premise of Levy and Markowitz fails under the Stable Pareto family of distributions. More specifically, we deduce that even though $Z_1 MV Z_2$, the investors under consideration prefer Z_2 due to it having a higher positive skewness than that of Z_1. We should also highlight here that we set the parameter α to lie around 1.4 and 1.6. This is due to the fact that approximately 70% of the 850 NYSE stocks have a fitted parameter α inside this range. The further away we get from $\alpha = 2$, the more non-normal the distribution gets. But why do we focus mainly on the findings derived from the Stable Pareto family of distributions? The answer here relies on the Generalized Central Limit Theorem (G-CLT). Based on it, a random variable $R = \sum_{i=0}^{n} \xi_i$ will be α-stable with $\alpha < 0 \leq 2$ if and only if ξ_i's are iid and $a_n(\xi_1 + \ldots + \xi_n) - b_n \to R$, with $a_n > 0$ and $b_n \in \mathbb{R}$. According to Mandelbrot (1963) the random variables ξ_i's have infinite variance. So, it is common to use the Stable Pareto family of distributions to describe daily and weekly stock returns. Thus, the fact that under Stable Pareto distributions the premise of Levy and Markowitz fails is highly relevant.

There are also more extreme non-normal cases for which we derive even worse results. Namely, if we let the parameter α of the Stable Pareto distribution to be between 1.2 and 1.4 we end up with the following Tables.

Table 3.10: Percentage of MVC⇒MEUC for $(1+Z)^a$

Distribution	Parameters Differences	$a=0.01$	$a=0.1$	$a=0.5$	$a=0.9$
	$1.1 < \frac{\mu_1}{\mu_2} \leq 1.15,\ 1.1 < \frac{\sigma_2}{\sigma_1} \leq 1.15,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	93%	95%	100%	100%
Stable	$1.05 < \frac{\mu_1}{\mu_2} \leq 1.1,\ 1.05 < \frac{\sigma_2}{\sigma_1} \leq 1.1,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	71%	74%	93%	100%
	$1.01 \leq \frac{\mu_1}{\mu_2} \leq 1.05,\ 1.01 \leq \frac{\sigma_2}{\sigma_1} \leq 1.05,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	33%	37%	67%	93%

Table 3.11: Percentage of MVC⇒MEUC for $log(a+Z)$

Distribution	Parameters Differences	$a=0.9$	$a=1$
	$1.1 < \frac{\mu_1}{\mu_2} \leq 1.15,\ 1.1 < \frac{\sigma_2}{\sigma_1} \leq 1.15,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	90%	93%
Stable	$1.05 < \frac{\mu_1}{\mu_2} \leq 1.1,\ 1.05 < \frac{\sigma_2}{\sigma_1} \leq 1.1,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	67%	71%
	$1.01 \leq \frac{\mu_1}{\mu_2} \leq 1.05,\ 1.01 \leq \frac{\sigma_2}{\sigma_1} \leq 1.05,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	30%	33%

Table 3.12: Percentage of MVC⇒MEUC for $-e^{-a(1+Z)}$

Distribution	Parameters Differences	$a=0.7$	$a=1$	$a=3$	$a=5$	$a=8$	$a=10$	$a=15$	$a=20$
	$1.1 < \frac{\mu_1}{\mu_2} \leq 1.15,\ 1.1 < \frac{\sigma_2}{\sigma_1} \leq 1.15,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	99%	96%	77%	45%	41%	35%	30%	26%
Stable	$1.05 < \frac{\mu_1}{\mu_2} \leq 1.1,\ 1.05 < \frac{\sigma_2}{\sigma_1} \leq 1.1,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	96%	90%	43%	30%	26%	21%	17%	14%
	$1.01 \leq \frac{\mu_1}{\mu_2} \leq 1.05,\ 1.01 \leq \frac{\sigma_2}{\sigma_1} \leq 1.05,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	78%	56%	0%	0%	0%	0%	0%	0%

Table 3.13: Percentage of MVC⇒MEUC for $-(1+Z)^{-a}$

Distribution	Parameters Differences	$a=0.01$	$a=0.3$	$a=0.5$	$a=1$	$a=3$	$a=5$	$a=8$	$a=10$	$a=15$	$a=20$
	$1.1 < \frac{\mu_1}{\mu_2} \leq 1.15,\ 1.1 < \frac{\sigma_2}{\sigma_1} \leq 1.15,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	95%	91%	87%	82%	65%	41%	37%	31%	27%	22%
Stable	$1.05 < \frac{\mu_1}{\mu_2} \leq 1.1,\ 1.05 < \frac{\sigma_2}{\sigma_1} \leq 1.1,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	64%	60%	54%	50%	35%	24%	21%	18%	14%	11%
	$1.01 \leq \frac{\mu_1}{\mu_2} \leq 1.05,\ 1.01 \leq \frac{\sigma_2}{\sigma_1} \leq 1.05,\ 1.5 \leq \frac{s_2}{s_1} \leq 3$	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Evidently, for those more extreme non-normal cases the investors inside $\mathbf{U_3}$ will be led to an even higher degree of wrong decisions, assuming they make their decisions based solely on the MVC.

To sum up, we conclude that the main issue with the premise of Levy and Markowitz lies with the more non-normal cases. These cases can better be described by the Extreme Value and Stable Pareto distributions. However, under the assumption of Gaussian, Laplace or Skew-Normal distributions the premise of Levy and Markowitz seems valid for all the utility functions that we put to test. In other words, under those distributions the above utility functions are almost quadratic. The Extreme Value distribution represents the first step in testing the premise under more non-normal cases. The MVC seems to resist the pressure for less risk-averse investors. The more noteworthy findings are derived by the Stable Pareto distribution, which, according to Mandelbrot, represents an appropriate description of the movement of daily stock returns. In this case, letting the means' and standard deviations' differences to hover around 1% and 15% results in erroneous decisions made by almost all parametrizations of the four investors. So, as expected, the greater the non-normality of the distribution of lotteries, the more information the decision makers will need, with respect to the characteristics of the distribution of each lottery, in order to make the right decision. As a result, we conclude that the premise of Levy and Markowitz works appropriately for Elliptical or Skew-Elliptical distributions. However, as we have thoroughly discussed, under the elliptical family of distributions the MVC is equivalent to the MEUC, for any concave utility function. So, the additional assumption of approximately quadratic utility functions is unnecessary. With regards to the Skew-Elliptical family of distributions we find some extreme cases for which only the very risk-averse investors might need to know the level of skewness of each lottery. Lastly, when departing from normality the premise of Levy and Markowitz is problematic for all the utility functions that we put to test. At this point we may give an answer to our initial question concerning the size of $\mathbf{U_3^*}$. In particular, as we have shown, the Elliptical family of distributions is expected to deliver MVC⇒MEUC for the four utility functions we put to test, without the need of the utility functions being approximately quadratic. So, such cases should not be considered to comprise $\mathbf{U_3^*}$. On the other hand, under the more non-normal distributions the four utility functions cannot be included in $\mathbf{U_3^*}$.

3.4 Conclusions

Since its conception in 1952, the MVC has gone through an extensive amount of criticism when it comes to its realism and usefulness. The main argument has always been that the underlying assumptions of either (i) quadratic preferences, or (ii) Gaussian distributions, are unrealistic. Markowitz (2010, 2014) [63] and [66] insists that the literature has misinterpreted his model. This stimulated us to revisit the MVC, so as to examine

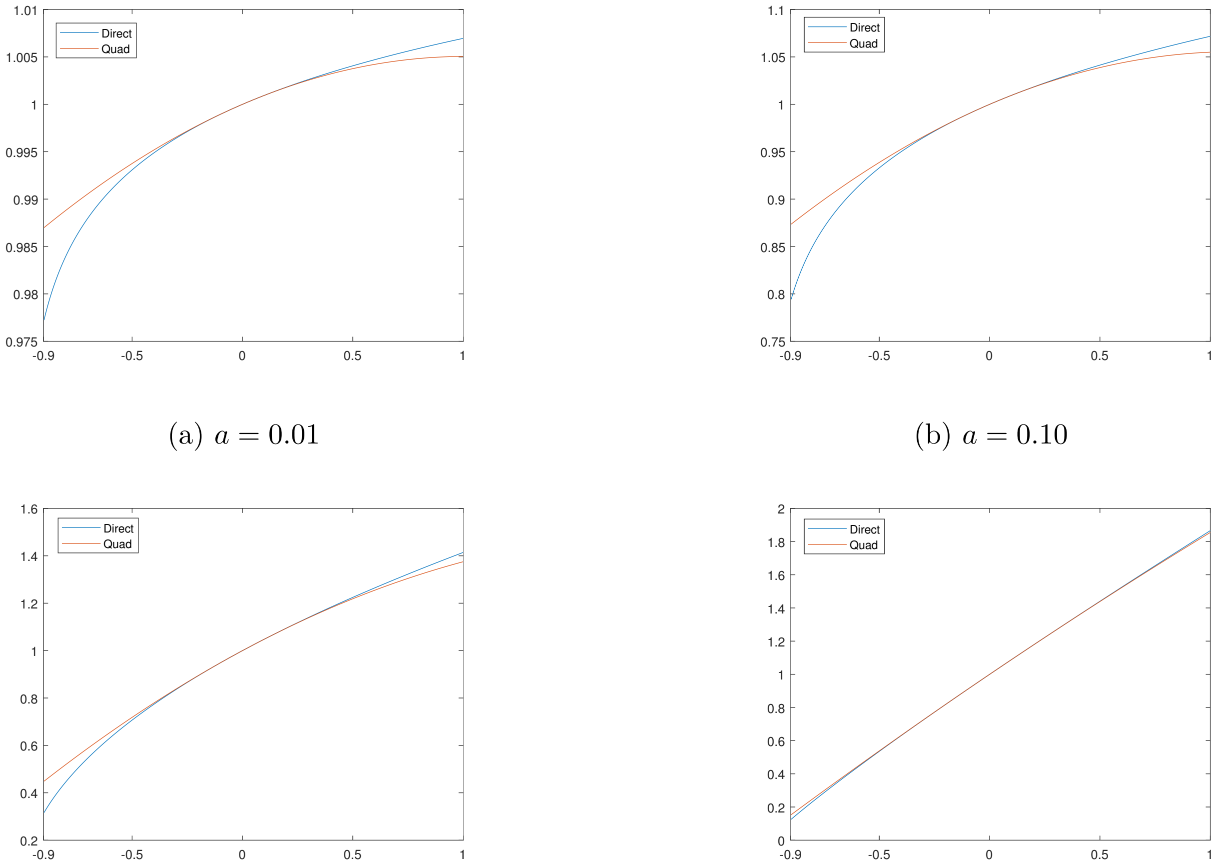
Markowitz's remark as well as to clarify how it is associated with the SD rules.

We analyzed thoroughly the literature to clarify which are the conditions that make the MVC coincide with the MEUC. We found that the elliptical family of distributions can replace the assumption of normality, based on the findings of Chamberlain (1983). However, the more recent findings of Schuhmacher et al. (2021) with respect to the Skew-Elliptical family are not consistent with our numerical results. In particular, we show that under Skew-Normal distributions there are very-risk averse investors that are part of $\mathbf{U_3}$, that may require to know the level of skewness of each lottery. Thus, we concluded that under Skew-Elliptical distributions we cannot claim that the MVC coincides with the MEUC.

From there, we went on to identify the class of investors for which the MVC coincides with the MEUC. We investigated Markowitz's premise in 1959, arguing that we only need approximately quadratic utility functions to make the MVC equivalent to the MEUC. We argued that the evidence from [57] and [65] does not suffice to support that premise. In fact, we proposed the use of MC Simulations in order to test Markowitz's premise under multiple types of distributions, for a specific choice of investors. We found out that under a Skew-Normal distribution, the MVC is equivalent to the MEUC except for the very risk-averse investors. But for more non-normal distributions, like the Extreme Value and Stable Pareto, even less risk-averse investors will make a high percentage of wrong decisions if they use only the information coming from the MVC. Thus, based on our findings, we deduced that Markowitz's premise seems to work for Skew-Elliptical distributions, but this is not the case for more non-normal distributions.

3.5 Appendix

Figure 3.1: $(1+Z)^a$ vs its Quadratic approximation around 0, for $Z \in [-0.9, 1]$



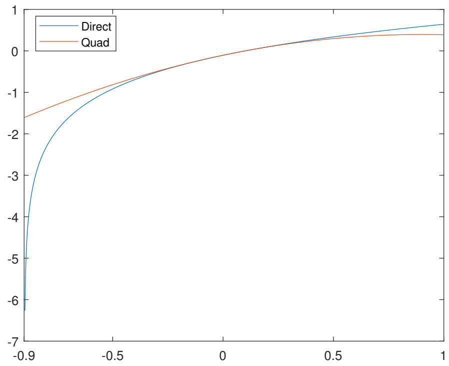
(a) $a = 0.01$

(b) $a = 0.10$

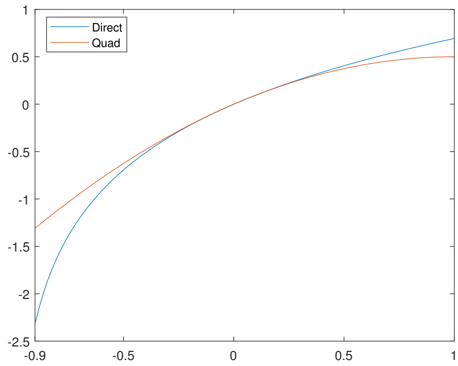
(c) $a = 0.50$

(d) $a = 0.90$

Figure 3.2: $log(a + Z)$ vs its Quadratic approximation around 0, for $Z \in [-0.9, 1]$

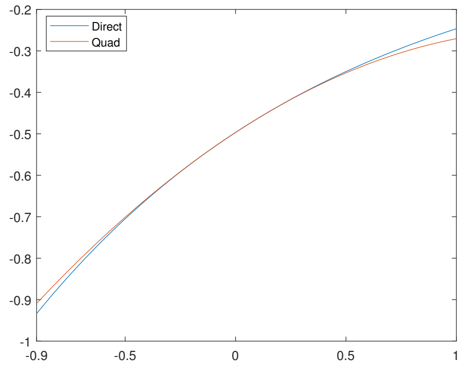


(a) $a = 0.90$

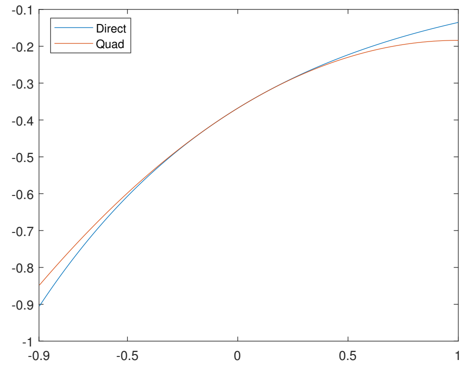


(b) $a = 1$

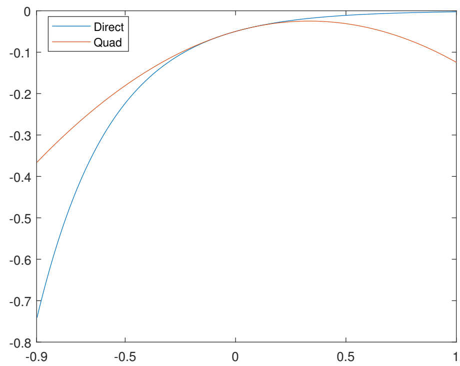
Figure 3.3: $-e^{-a(1+Z)}$ vs its Quadratic approximation around 0, for $Z \in [-0.9, 1]$



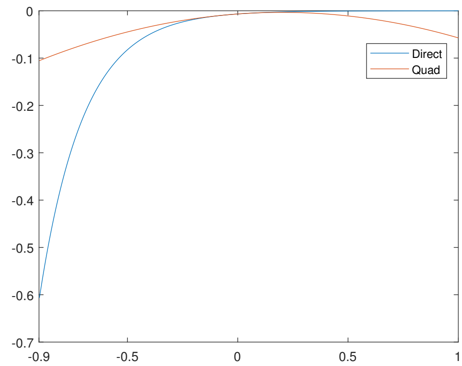
(a) $a = 0.70$



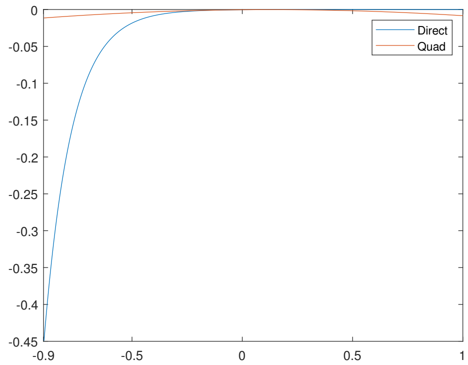
(b) $a = 1$



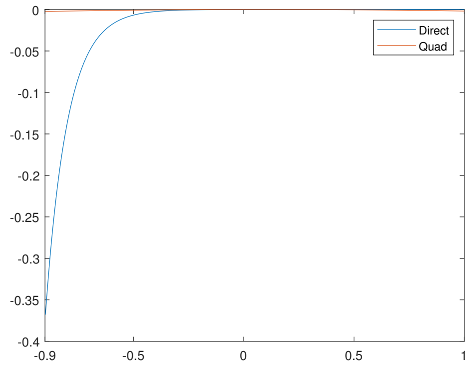
(c) $a = 3$



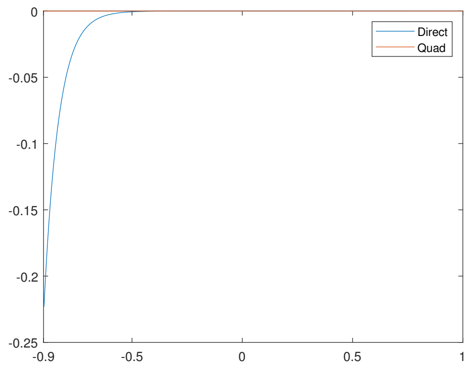
(d) $a = 5$



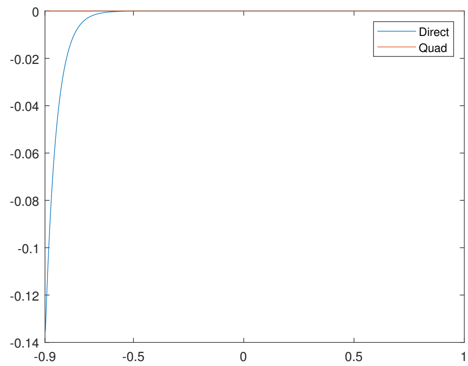
(e) $a = 8$



(f) $a = 10$

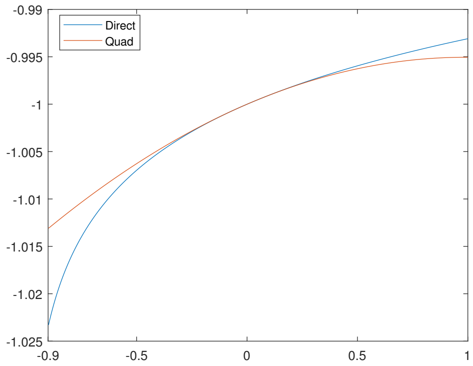


(g) $a = 15$

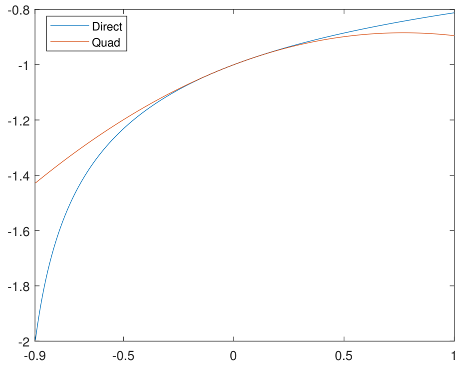


(h) $a = 20$

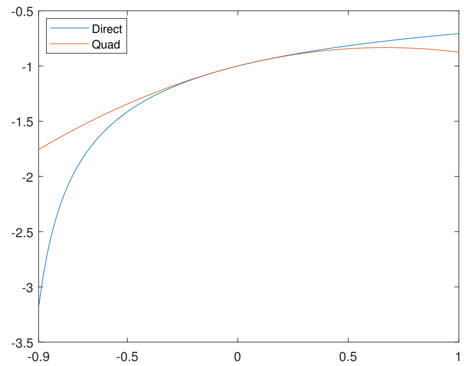
Figure 3.4: $-(1+Z)^{-a}$ vs its Quadratic approximation around 0, for $Z \in [-0.9, 1]$



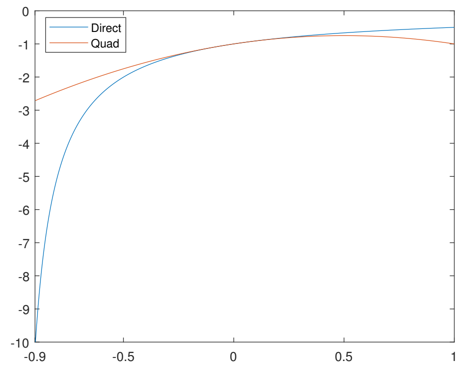
(a) $a = 0.01$



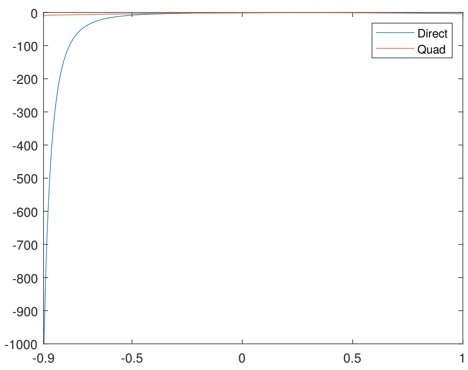
(b) $a = 0.30$



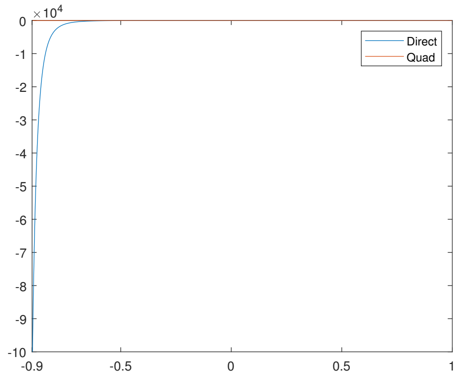
(c) $a = 0.50$



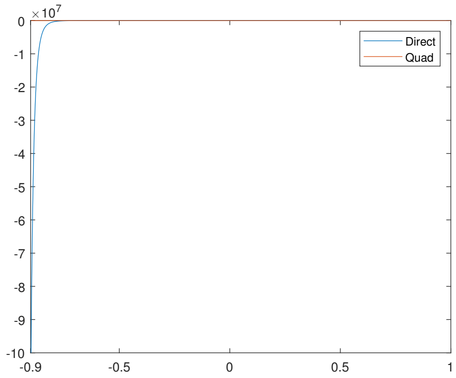
(d) $a = 1$



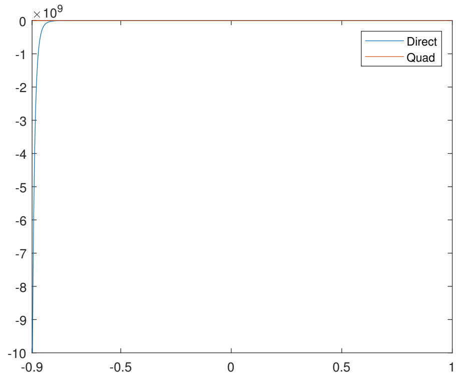
(e) $a = 3$



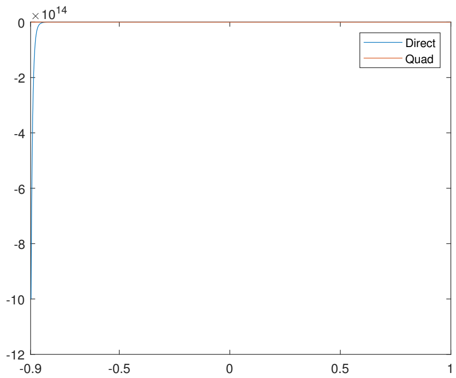
(f) $a = 5$



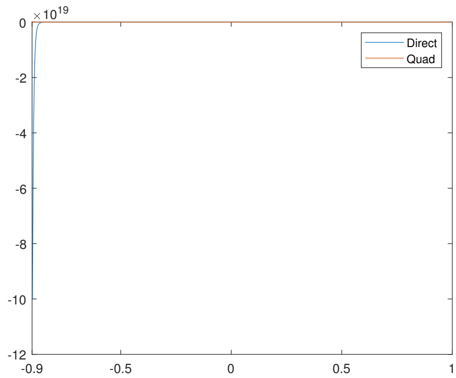
(g) $a = 8$



(h) $a = 10$



(i) $a = 15$



(j) $a = 20$

Chapter 4

Dynamic Estimates Of The Arrow-Pratt Absolute And Relative Risk Aversion Coefficients

4.1 Introduction

The risk attitude of investors is considered to be a crucial factor in Portfolio and Decision Theory. A rational investor in the Von Neumann-Morgenstern sense [87], will always choose the lottery that maximizes his/her expected utility function. Based on that decision criterion, Arrow and Pratt (1964, 1965) [4], [76] developed the concepts of ARA and RRA as ways to indicate the level of risk-aversion of the investor. Determining the trends of ARA and RRA is crucial for pointing at a specific class of utility functions.

The existing literature has done a pile of work on the recovery of degree of ARA and RRA of investors. In terms of ARA, the literature universally finds evidence of decreasing ARA (DARA) as wealth increases. This is in agreement with Arrow (1965) [4] who claimed that the DARA "*seems supported by everyday observation*". On the other hand, the empirical works on RRA derive mixed conclusions. Some of the earliest studies that found evidence of DRRA with respect to wealth are [23], [32], [52], [68], [69] and [71]. While evidence of increasing RRA (IRRA) was found in [8], [11], [26] and [83]. Most of these works, based their findings on cross-section data, which implicitly assumes that all investors have the same utility function, a rather unrealistic assumption. Works like that of [8], [11], [26] and [52] used both experimental and econometric methods to extract RRA. A totally different approach was used by Jackwerth and Carsten (2000) [37] as well as Perignon and Villa (2002) [71] who utilized information from Options markets. The most recent part of the literature, argues that investors reveal constant RRA (CRRA). Some of the works supporting this evidence are [16], [18] and [79]. The advantage of these studies is that they all use panel data. So, the implicit assumption of a common utility function among investors is no longer made.

Our work revolves around the definitions of ARA and RRA and the concept of a lottery that is not "fair". Considering the fact that markets are not "fair" lotteries the value of such an approach becomes evident. We proceed with deriving the formulae of the absolute and relative risk premium which subsequently lead to ARA and RRA when the underlying lottery is not "fair". These formulae can be used to extract the risk attitude of investors across different markets, specifically, in terms of ARA and RRA.

Using monthly returns and market cap from CAC 40, EURO, S&P 500 and STOXX 600 as well as 10-year Treasury yields, we find evidence of DARA and IRRA for 2012 to 2022, for all markets. Moreover, our results capture the impact of the QE program announced by Fed in March 2020. We also test our findings using a rolling-window approach with 5 and 15-year window sizes to assess their robustness. The new results do not deviate from our prior conclusions. Then, we use data for different time periods, namely, 1993-1998 and 1999-2005 to determine whether or not the investors' risk attitude is consistent through years. Our findings, point at DARA and DRRA for 1993-1998 and CRRA/IRRA and DARA for 1999-2005. As a result, we conclude that through the years the markets became more risk-averse with respect to their level of wealth.

The evidence of DARA and IRRA for 2012-2022 is satisfied by a specific class of utility functions. This leads us to our next question which has to do with measuring the impact of a wrong assumption with respect to the utility function of an investor. For that, we consider an asset manager who mistakenly assumes that his client is characterized by a quadratic utility function while in reality the client's true utility function is logarithmic. Assuming that the asset manager diversifies the investor's wealth between a risky and a risk-free asset we propose a way to measure the differences in the asset manager's decisions on behalf of the investor for different utility functions. The results predicate that the asset manager will mislead his client considerably, in case he assumes quadratic instead of logarithmic utility functions.

Our main findings split into four parts. (i) We derive a closed-form expression for the degree of RRA for a lottery with nonzero mean. (ii) We apply our formula on four different markets, namely CAC 40, EURO, S&P 500 and STOXX 600, to deduce that all of them exhibit DARA and IRRA. (iii) We also find evidence that the slope of the RRA has changed through different time periods, while the slope of ARA has always been decreasing. (iv) Finally, we introduce a way to measure the differences in portfolio diversification among different utility functions.

4.2 Theoretical Framework

4.2.1 Degree of Risk Aversion for a "Fair" Lottery

In this section, we delve into the different notions underlying Decision Theory. Consider the following case. A decision maker (or investor) with level of wealth w_0 is asked whether or not he wants to participate in a game (or lottery) represented by Z. Assuming that his binary relation ($\succeq$) satisfies the axioms posed by VN-M for any lottery, we can safely say that his decision criterion arises from the Representation Theorem of VN-M. Namely, he should always choose the lottery that maximizes his expected utility function. Now, in order to participate in the game the investor demands a risk premium represented by $\rho(w_0, Z)$. The risk premium is based on the notion of Certainty Equivalent (hereafter; CE). The CE is defined below.

Definition 4.1. *The* ***CE*** *of lottery Z is the amount z_0 which makes the decision maker indifferent between lottery Z and this certain amount z_0. Alternatively, if the decision maker owns lottery Z, then z_0 represents the minimum amount he is willing to sell the lottery. Mathematically,*

$$\zeta_{z_0} \sim Z \overset{VN-M}{\Leftrightarrow} U(z_0) = E[U(Z)] \Leftrightarrow z_0 = U^{-1}(E[U(Z)]), \quad \forall Z \in \mathcal{P}$$

where $\mathcal{P} = \{P : P$ is a probability function on $\mathcal{Z}\}$ and $\mathcal{Z} = \{Z : Z$ is a random variable$\}$.

From the definition, we see that the CE results from the decision maker's subjective characteristics, but is also dependent on the objective (probabilistic) characteristics of lottery Z. Thus, the level of "fear" the decision maker has with respect to risk will determine the magnitude of CE for bearing the risk of Z. This leads to the definition of risk premium.

Definition 4.2. *The* ***risk premium*** *is the difference between the expected return of lottery Z and the CE the decision maker asks for it.*

$$\rho(Z) = E[Z] - z_0 = \mu_Z - z_0.$$

The notion of risk premium translates in two ways. (i) If the decision maker owns lottery Z, $\rho(Z)$ represents the expected units of return he is willing to "sacrifice" in order to avoid its risk. (ii) Alternatively, if the decision maker considers owning lottery Z, $\rho(Z)$ represents the expected units of return he "demands" in order to bear the risk of lottery Z. In our frame, we will focus exclusively in the case of risk-averse investors with an increasing utility function $U' > 0$. That implies that the following should hold:

- The utility function U is concave, i.e. $U'' < 0$
- The risk premium is positive $\rho(Z) > 0, \ \forall Z$

Now, assume that on a later stage of his life the investor has a new higher level of wealth w_0^+ ($w_0^+ > w_0$). Assume also that the investor's utility remains unchanged. What is the new risk premium $\rho(w_0^+, Z)$ the investor demands assuming that the objective characteristics of lottery Z remain unchanged? Although, it can be the same we shall consider the case where

$$\rho(w_0^+, Z) \neq \rho(w_0, Z).$$

Since the mean and variance of lottery Z remain the same what is the causative factor for the change in ρ? There should be a function of w_0 which we will express by $r(w_0)$ that affects the risk premium. Based on Arrow and Pratt this function is called local degree of ARA of the investor with respect to his level of wealth. This measure is defined below.

Definition 4.3. *The local degree of* ***Absolute Risk Aversion****, $r(w)$, at a level of wealth w is defined as*

$$r(w) = -\frac{U''(w)}{U'(w)}$$

Interpretation of ARA:

- For risk-averse investors we have $U'' < 0$. The larger $|U''|$, the more concave the utility function of the investor will be. In other words, the investor will be more risk-averse.

- According to VN-M, any positive linear transformation of U should still satisfy the Representation Theorem and by extension lead to the same $r(w)$. That justifies the division by U'.

- ARA overcomes an important limitation of the risk premium. Namely, in order to compare the degree of risk-aversion of two investors, we would need to compare their risk premiums for any lottery Z and any level of wealth w. On the contrary, $r(w)$ needs to be tested just for any level of wealth w.

- This degree of risk aversion is measured in absolute terms. Thus, lottery Z represents an absolute amount (expressed in dollars).

Now that we specified the factor that generates the differences in the aforementioned example, we will derive $\rho(w_0, Z)$ in terms of $r(w_0)$. In fact, we assume that Z is a "fair" lottery in the sense that $\mu_Z = 0$ and σ_Z is relatively small. Let also, the final level of wealth of the decision maker expressed by,

$$W_1 = w_0 + Z.$$

Using the definition of CE together with 2nd degree Taylor series we derive,

$$\rho(W_1) = \rho(w_0, Z) = \rho(w_0 + Z) = \frac{1}{2} r(w_0) \sigma_Z^2. \tag{4.1}$$

Interpretation of $\rho(w_0, Z)$:

- The subjective factor, $r(w_0)$, remains unchanged regardless of any changes in the objective characteristics of lottery Z. However, the risk premium varies for different lotteries. This is due to the objective factor of the lottery, σ_Z^2.

- If $r(w_0)$ is decreasing (DARA), increasing (IARA) or constant (CARA) for any level of wealth w_0, then $\rho(w_0, Z)$ will also be decreasing, increasing, or constant, respectively.

Within the current context, lottery Z is independent of any change in w_0. Consider the following case. A decision maker with $w_0 = 1000\$$ is asked to participate in a coin toss game with the following set of outcomes. If the coin lands "Heads" the decision maker will earn 100\$ while if it lands "Tails" he will have to pay 100\$. The decision maker decides that this lottery is too risky for his risk appetite. Assume that the decision maker's degree of risk aversion is DARA. So, at a later stage of his life his wealth is $100,000\$$ and thus he feels comfortable to participate in exactly the same game. Now, assume that the same decision maker is asked to participate in a different coin toss game Z' in which if the coin lands "Heads" he will earn 10% of his wealth w_0 while if it lands "Tails" he will have to pay 10% of w_0. That means that for $w_0 = 1000\$$ he will earn/lose 100\$ while for $w_0 = 100,000\$$ earn/lose $10,000\$$. Obviously, this concept may alter the decision made by the decision maker. In particular, the fact that his wealth increased does not necessarily mean that he will participate in Z' since this new game changes with his level of wealth, i.e. $Z' = w_0 R$, where R is a lottery expressed in percentages.

To deal with such cases, Arrow and Pratt defined a new measure called the local degree of RRA.

Definition 4.4. *The local degree of* ***Relative Risk Aversion****,* $\lambda(w)$*, at a level of wealth* w *is defined as*

$$\lambda(w) = -w\frac{U''(w)}{U'(w)} = wr(w).$$

Interpretation of RRA:

- RRA depends on ARA, since $\lambda(w) = wr(w)$. If for example $r(w)$ is decreasing at a faster rate than the rate at which w is increasing, $r(w)$ will also be a decreasing function of w (DRRA). On the other side, if $r(w)$ is decreasing at a slower rate than the rate at which w is increasing, $r(w)$ will be an increasing function of w (IRRA).

- This degree of risk aversion is measured in relative terms. Thus, we define a new lottery $R = Z/w_0$ which represents a relative amount (expressed as a fraction of 100 (percentage)).

As shown above, under this new measure, we define a new "fair" lottery R, such that $Z = Rw_0$, with $\mu_Z = \mu_R = 0$ and σ_Z, σ_R be relatively small. Then, the "relative" risk premium for W_1 will be,

$$\begin{aligned}
&\rho(W_1) = \rho(w_0, Rw_0) = \frac{1}{2} r(w_0)\sigma^2_{Rw_0} \\
&\frac{\rho(w_0, Rw_0)}{w_0} = \frac{1}{2}\frac{w_0^2 r(w_0)\sigma^2_R}{w_0} \\
&\tilde{\rho}_{w_0}(R) = \frac{1}{2}\lambda(w_0)\sigma^2_R.
\end{aligned} \tag{4.2}$$

Based on the above analysis, the literature has done extensive research on the Decision Making of investors under risk. More specifically, as we will see in the following subsections the literature focused on identifying the slope of the ARA and the RRA of investors with respect to their level of wealth. We will see that previous research works have followed multiple approaches in trying to arrive at more reliable conclusions. Following our overview, we will introduce a totally unique approach on this subject.

4.2.2 Decreasing Absolute Risk Aversion Literature

In 1965, Arrow [4] posed the hypothesis that investors reveal DARA and IRRA. In fact, he claimed that the DARA "*seems supported by everyday observation*". However, in terms of IRRA Arrow said that: "*the hypothesis of increasing relative risk aversion is not easily confrontable with intuitive evidence*". From that point on, an abundance of research works have attempted to recover the relative risk preferences of the investors, while DARA is universally accepted. Extracting RRA, has been proven to be particularly challenging. We will review the research work on RRA in the subsequent sections.

Haim Levy (1994) [52], applying time-series analysis found that "*only six subjects out of 62 significantly contradict the DARA property*". As a matter of fact, Levy concluded that the evidence is significantly strong and supports Arrow's assertion. We will analyze Levy's approach in the next section where we review his findings on RRA.

Some other studies that find supportive evidence of DARA are [8], [26] and [37]. Going on, the literature considered DARA to be the norm and focused more on RRA. This universally accepted property of individual risk preferences plays a crucial role in many applications of the expected utility theory.

4.2.3 Decreasing Relative Risk Aversion Literature

Let us focus on the bibliography that shows evidence of DRRA. Cohn et al. (1975) [23] focused on the information from 588 different portfolio allocations provided by customers of a retail brokerage firm. In particular, each customer provided information with respect to the share he had in common stocks, corporate bonds, etc., in percentages. The authors established two alternative classifications for long-term fixed-income securities. In the first one, they treated Savings Account, Checking Account, Personal Residence, Personal Property and Other Assets as the "risk-free" assets. While, in the second Classification they also included Preferred Stock, Corporate Bonds and Government Bonds in this category. However, there is no compelling reason to treat Corporate Bonds as riskless assets. The level of wealth was also given two different definitions. With the first one being the Total Assets and the second one being Total Assets less Personal Residence. Although, they acknowledged the fact that a preferred proxy of wealth would be the Net Worth of each customer, which was not possible as the customers did not provide any data on their Liabilities. Such an approach could conceivably alter their conclusions. The authors found strong evidence of DRRA even when they controlled for demographic factors like the age of each customer as well as his/her marital status. In their study, they split the customers into different wealth groups, in order to determine whether or not their findings are consistent between different groups. The fact that their target group pertains to active investors only, means that they do not cover a complete cross-section of investors at each wealth level. Finally, the aforementioned results depend on cross-sectional data which means that the authors did not study the portfolio allocation of each individual in time. This implies that they draw conclusions based on information from individuals with different utility functions.

Morin and Suarez (1983) [68], used portfolio allocation data of 9,962 different private households taken from the Survey of Consumer Finances (SCF) in Canada. The authors argue that the SCF database used is more representative of investors and broad-based in terms of range of wealth covered. One major improvement in their methodology is that they use Net worth as a proxy of wealth. This is due to the private households providing information with respect to their debt obligations. In their framework, the "risk-free" assets include Cash, Deposit Account Balances, Canada Savings Bonds, and Personal Property. As in Cohn et al., the households were split into different wealth groups. What is interesting is that for poorer individuals they found evidence of IRRA while for the middle-wealth individuals DRRA was supported and lastly for the wealthiest group the findings pointed to CRRA. However, these findings are based on interpersonal comparisons of utility. A similar approach was done by Guiso et al. (1996) [32]. In fact, Guiso et al. revisited the RRA subject by doing a cross-sectional analysis on a random sample of 8,274 Italian households who provided information with respect to their portfolio allocation proportions. Their findings showcased that the households revealed DRRA.

In 1994, Haim Levy [52] introduced a different approach in determining the slope of RRA of individuals. Levy conducted an experiment in which 62 of his MBA students participated. Each student was given an initial investment allotment of $30,000$$ "paper" money and was offered stocks of 20 pure equity firms as well as a "risk-free" rate of 2%. The major difference between Levy's approach and the two aforementioned works is that he ran time-series regressions for each individual and did not test cross-sectionally. What this means, is that through his approach the results are not based on interpersonal utility functions. Now, in terms of RRA, Levy argued that the results indicate a decreasing trend with respect to wealth. On top of that, Levy also did robustness checks studying the subjects' portfolio allocation when the only available assets are one risky and one risk-free asset. In this frame, the conclusions did not deviate from the previous findings. Even though this new approach overcomes some previous limitations an important caveat of this work is the small sample size.

A whole different approach was proposed from Perignon and Villa (2002) [71] in 2002. More specifically, they attempted to extract information from Put and Call Options on CAC 40 with regards to the slope of RRA. The two authors, based their methodology on Ait-Sahalia and Lo (2000) [2] who defined a pure-exchange economy. Namely, in this economy and in equilibrium, the investor optimally invests all his wealth in the single risky stock at every instant prior to the terminal date and then consumes the terminal value of the stock at time T. This implicitly says that the level of wealth of the individual is exactly equal to the stock price at each time. In other words, the representative investor consumes only at the final date and maximizes the expected utility of the terminal wealth by choosing the amount invested in the stock at each intermediary date. In this framework, Ait-Sahalia and Lo derive the "implied" RRA formula. Empirically, Perignon and Villa, obtain an "implied" RRA for the CAC 40 stock index and they conclude that as the index price (i.e. the wealth) increases the "implied" RRA decreases.

4.2.4 Increasing Relative Risk Aversion Literature

Contrary to the previous works, Siegel and Hoban (1982) [83] find that by restricting the sample to higher wealth target groups will lead to DRRA or CRRA. But they argue that "*the use of a broader based sample and a more comprehensive measurement of wealth alters the conclusions and a pattern indicative of increasing relative risk aversion emerges*". Siegel and Hoban used data from the US National Longitudinal Surveys (NLS). More specifically, they acquired a sample of 2,881 different sets of asset holdings of individual households. The proxy for level of wealth was Net Worth. As for the "riskless" assets they used Cash, Deposit Account Balances, and U.S. Savings Bonds. Splitting into different groups of level of wealth they concluded that for any group of individuals there is evidence of IRRA rather than DRRA. We should keep in mind that this paper constitutes a cross-sectional analysis as in Cohn et al. (1975) [23] and Morin and Suarez (1983) [68].

In 2003, Eisenhauer and Ventura [26] found evidence of both DARA and IRRA. The authors asked different households in Italy the following question: "*You are offered the opportunity of acquiring a security permitting you, with the same probabilities, either to gain 10 million lire or to lose all the capital invested. What is the most you are prepared to pay for this security?*". The answer to this question is denoted by z and we can interpret it as the CE of the respective household. The lower z is the more risk-averse the household. The proxy of level wealth, denoted by w, was the average income of each household from 1993 to 1995. Now, the authors observed that 1,624 of the households interpreted the 10 million lire as a gross gain meaning that in the favourable case they get $w - z + 10$ while in the other case their wealth becomes $w - z$. A number of 1,730 households interpreted the 10 million lire as a net gain meaning that in the favourable case they get $w + 10$ while in the other case their wealth remains to be equal to w. In any case, they found that investors exhibit DARA and IRRA. Again, the most obvious limitations are that the results are based on a hypothetical survey question as well as on a cross-sectional analysis.

Two additional works finding evidence of IRRA are those of Bar and Shira (1997) [8] and Barsky et al. (1997) [11]. The first one was based on a set of gamble questions that had to do with the retirement decisions. The second one extracts the risk attitude of Israeli farmers.

4.2.5 Constant Relative Risk Aversion Literature

The research works finding evidence of CRRA are more recent and better substantiated. The first work finding evidence of CRRA was that of Friend and Blume (1975) [29]. Using cross-section regressions based on household-level data on asset holdings the authors conclude that CRRA characterizes household behaviour. The main drawbacks of this study are that it is based on cross-sectional data and the data is only focused on high-wealth households.

We will now discuss about the work of Sahm (2012) [79] which is the latest work of our review but it had started since 2006. Sahm initialized the use of panel data instead of cross-section data with regards to determining the slope of RRA in terms of wealth. In particular, he used gamble responses across the 1992 to 2002 waves of the Health and Retirement Study (HRS). The gamble under study is the following: "*Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50-50 chance the second job would double your total lifetime income and a 50-50 chance that it would cut it by a third. Which job would you take - the first job or the second job? Individuals who accept the first risky job then consider a job with a larger downside risk of one-half, while those who reject the first risky job are asked about a job with a smaller downside risk of one-fifth. If they reject the first two risky jobs, individuals consider a third risky job that could cut their lifetime income by one-tenth. Likewise, if they accept both risky jobs, individuals consider a third* *risky job that could cut their lifetime income by three-quarters.*" Sahm finds no effect of wealth changes on relative risk aversion and thus concludes CRRA.

The next work we are going to discuss is that of Brunnermeier and Nagel (2008) [16]. The authors use household-level panel data from the Panel Study of Income Dynamics (PSID), covering a period of 20 years between 1984 and 2003. The data includes Asset holdings, Income and households characteristics. To identify how wealth changes are related to market participation they use probit regressions. In particular, the authors first split the data in two subperiods, 1984 to 1999 and 1999 to 2003. The reason is that for 1984–1999 the time-span between successive waves of the PSID with wealth information is $k = 5$ years, and for the 1999–2003 sample the time-span between successive waves of the PSID with wealth information is $k = 2$ years. In practice, the probit regressions derive the probability of the households who did not participate in the stock market at time $t - k$ to enter until time t. They also estimate the probability that a household that is participating at $t - k$ to exit the stock market until t. The authors find that for both subperiods, namely 1984-1999 and 1999-2003, there is a 1% probability to enter the market when there is a 10% increase in the wealth. Accordingly, the probability of exiting the stock market when there is an increase in the wealth is extremely low. Thus, they conclude that the households' relative risk aversion remains constant with respect to a change in wealth (CRRA). A major improvement of this work is that it uses panel data. This way you avoid the implicit assumptions made in the cross-sectional analysis. Namely, assuming that the distributions of wealth and preferences are independent.

Another important work is that of Chiappori and Paiella (2011) [18]. Similar to [16] and [79], they use panel data to showcase evidence of CRRA. In fact, they prove that previous studies that supported their findings on cross-sectional analysis are led to erroneous conclusions. Citing the authors: "*without a priori restrictions on the joint distribution of wealth and preferences, the form of individual preferences simply cannot be recovered from cross-sectional data. In fact, any form of individual preferences is compatible with any observed, joint distribution of wealth and risky asset shares provided that one can freely choose the joint distribution of wealth and preferences.*". In other words, estimating the joint distribution of wealth and risky asset shares using cross-section data derives the joint distribution of wealth and preferences only and only if you pre-assume that preferences are the same for each investor. This however is an over-simplification of reality, which substantiates the use of panel data instead of cross-section data. The data used is from the Survey of Household Income and Wealth (SHIW), which is a large-scale household survey run every two years by the Bank of Italy from which they get asset allocations for 1,332 households. Also, they exclude all the households with a change in wealth less than 25%. As in [16] and [79], they find evidence of CRRA.

4.3 Degree of Risk Aversion for a Lottery with Non-Zero Mean

In the previous section, we derived the "absolute" and "relative" risk premium for a "fair" lottery. In reality however, stock prices are not a fair lottery meaning that $E[Z] = \mu_Z \neq 0$. Equivalently, stock returns are not a "fair" lottery and so $E[R] = \mu_R \neq 0$. Deriving ARA under $\mu_Z \neq 0$ leads to RRA through $\lambda(w_0) = w_0 r(w_0)$. So, for a non-fair lottery with $\mu_Z \neq 0$ we have

$$\rho(W_1) = \rho(w_0 + Z) = E[w_0 + Z] - z_0 = w_0 + \mu_Z - z_0,$$

where z_0 represents the CE of lottery Z with w_0 level of wealth (or equivalently the CE of lottery $w_0 + Z$). Thus, the CE will be

$$z_0 = w_0 + \mu_Z - \rho(w_0 + Z).$$

According to Definition 3.1, the following equation must be satisfied

$$E[U(Z)] = U(z_0).$$

In our case,

$$E[U(w_0 + Z)] = U(w_0 + \mu_Z - \rho(w_0 + Z)).$$

At this point, we use Taylor expansions of 2nd-order around w_0 for both sides of the equation. For the left side of the equation we derive

$$\begin{aligned} E[U(w_0 + Z)] &\simeq U(w_0) + U'(w_0)E[w_0 + Z - w_0] + \frac{U''(w_0)}{2}E[(w_0 + Z - w_0)^2] \\ &\simeq U(w_0) + U'(w_0)\mu_Z + \frac{U''(w_0)}{2}(\sigma_Z^2 + \mu_Z^2). \end{aligned} \tag{4.3}$$

Accordingly, for the right side of the equation we get

$$\begin{aligned} U(w_0+\mu_Z-\rho(w_0+Z)) &\simeq U(w_0)+U'(w_0)(w_0+\mu_Z-\rho(w_0+Z)-w_0)+\frac{U''(w_0)}{2}(w_0+\mu_Z-\rho(w_0+Z)-w_0)^2 \\ &\simeq U(w_0)+U'(w_0)(\mu_Z-\rho(w_0+Z))+\frac{U''(w_0)}{2}(\mu_Z-\rho(w_0+Z))^2 \\ &\overset{\text{drop 3rd term}}{\simeq} U(w_0)+U'(w_0)(\mu_Z-\rho(w_0+Z)). \end{aligned} \tag{4.4}$$

Now, combining (4.3) and (4.4) we get

$$\begin{aligned} &U(w_0)+U'(w_0)\mu_Z+\frac{U''(w_0)}{2}(\sigma_Z^2+\mu_Z^2) \simeq U(w_0)+U'(w_0)(\mu_Z-\rho(w_0+Z)) \\ &\frac{U''(w_0)}{2}(\sigma_Z^2+\mu_Z^2) \simeq -U'(w_0)\rho(w_0+Z) \\ &\rho(w_0+Z) \simeq -\frac{U''(w_0)}{2U'(w_0)}(\sigma_Z^2+\mu_Z^2) \\ &\rho(w_0+Z) \simeq \frac{1}{2}r(w_0)(\sigma_Z^2+\mu_Z^2). \end{aligned}$$

So, compared to (4.1) the risk premium will satisfy the following equation

$$\boxed{\rho(w_0+Z) \simeq \frac{1}{2}r(w_0)(\mu_Z^2+\sigma_Z^2)} \tag{4.5}$$

The main change is that the risk-premium is now positively related also to μ_Z^2. In other words, a higher mean in absolute terms, would increase the required ρ. Now solving for the ARA term $r(w_0)$ we get

$$w_0+\mu_Z-z_0 \simeq \frac{1}{2}r(w_0)\mu_Z^2+\frac{1}{2}r(w_0)\sigma_Z^2$$

$$\boxed{r(w_0) \simeq \frac{2(w_0+\mu_Z-z_0)}{\mu_Z^2+\sigma_Z^2}} \tag{4.6}$$

The above formula represents a closed-form solution for ARA when Z is not a "fair" lottery. In relative terms, Z can be expressed in the following way

$$R=\frac{Z}{w_0} \Leftrightarrow Z=Rw_0.$$

Therefore, we have

$$\begin{aligned} \rho(w_0+Z) &\simeq \frac{1}{2}r(w_0)w_0^2(\sigma_R^2+\mu_R^2) \\ \rho(w_0+Z) &\simeq \frac{1}{2}r(w_0)(\sigma_Z^2+\mu_Z^2). \end{aligned}$$

And so, the "relative" risk premium for $\mu_R \neq 0$ takes the following form

$$\begin{aligned}\frac{\rho(w_0, Rw_0)}{w_0} = \tilde{\rho}(w_0, R) &\simeq \frac{1}{2} r(w_0) w_0 (\sigma_R^2 + \mu_R^2) \\ &\simeq \frac{1}{2} \lambda(w_0)(\sigma_R^2 + \mu_R^2)\end{aligned}$$

$$\boxed{\tilde{\rho}(w_0, R) \simeq \frac{1}{2} r(w_0) w_0 (\mu_R^2 + \sigma_R^2)} \tag{4.7}$$

Now, we can derive the RRA term $\lambda(w_0)$.

$$\boxed{\boldsymbol{\lambda(w_0) \simeq \frac{2(1 + \mu_R - \tilde{z}_{Rw_0})}{\mu_R^2 + \sigma_R^2}}} \tag{4.8}$$

Here, $\tilde{z}_{Rw_0} = z_0/w_0$ defines the "relative" CE expressed as a fraction of 100 (percentage). The main question at this point is how we interpret the terms in formula 4.8. Our main purpose is to use 4.8 to extract the level of RRA of the market. Since the degree of RRA is measured in "relative" terms we know that lottery $R = Z/w_0$ is expressed as a fraction of 100 (percentage). Thus, we could assume that R represents the market returns. Accordingly, μ_R and σ_R represent the mean and standard deviation of market returns, respectively. Likewise, since $\tilde{z}_{Rw_0}$ is the "relative" CE which is required by the investor (market), we assume that it represents the risk-free rate. Finally, w_0, which is the level of wealth of the investor (market), is assumed to define the market capitalization of the market.

4.4 Empirical Results

The previous sections showcased that there is no definitive answer as to if and how much RRA is affected by a change in the level of wealth of investors. In our view, the largest part of the literature has focused on determining the slope of RRA using an approach with multiple perils. Namely, most studies have to deal with demographic and socio-economic characteristics or over-simplifications. Even controlling for those characteristics may lead to sub-optimal conclusions. The most recent part of the literature finds evidence of CRRA using panel data. We believe that these studies are more meaningful and better substantiated.

In our framework, we try to avoid the issues noted in cross-sectional data and at the same time simplify even further the extraction of RRA from real data. More specifically, we believe that one should focus on determining the market's ARA and RRA through the formulae we derived. These formulae are free of any need of including and subsequently controlling for any subjective characteristic of the investors such as educational status or age or even level of wealth. The only subjective characteristic lies in the CE (or equivalently, risk-free rate) which we will measure using 10-year Treasury yields, since this is considered a logical proxy of the risk-free rate in the literature. The results that we will analyze in the following paragraphs have also been verified using 3-month Treasury bill rates.

We use monthly returns and market cap from four different markets, namely CAC 40, EURO, S&P 500 and STOXX 600, spanning from 1991 to 2021 excluding 2008. The reason why we exclude 2008 is because this period is known for extreme levels of volatility which may affect our concluding remarks. We believe that the advantage of using monthly returns instead of daily returns is that they do not exhibit volatility clustering. This means that μ_R and σ_R can be estimated recursively. Estimations start from 2012 using all the past data and continue up until 2022. As a proxy of risk-free rate we use the respective 10-year Treasury yields of each market. The reason we want to extract RRA for all these indexes is to determine whether or not the results are consistent across different markets.

The first issue we want to address is how do ARA and RRA of the aforementioned market indices change with respect to an increase in their market cap. From Figure 4.5, we deduce that for all market indices ARA is decreasing (DARA). In fact, the linear correlation between the market capitalization and the degree of ARA is almost perfectly negative for all markets. What this says, is that in all these indices the investors become less risk averse in absolute terms. So, as the level of wealth of a market (i.e. the market cap) increases investors are willing to take on more risk. This evidence is strongly supported by the findings of the literature. Figure 4.6, presents the RRA of each market with respect to each level of wealth. We observe that for the European indices RRA displays an increasing trend with high positive correlation. When it comes to S&P 500, we see that the correlation is -36%, but we should not focus entirely on it. Taking a closer look, we can see that up until the level of a market cap of approximately 27 trillion dollars the RRA exhibits an increasing trend. More specifically, the correlation between the market cap and the degree of RRA up until this point is approximately 90%. In fact, we argue that the sudden drop at a market cap of approximately 27 trillion dollars is due to an important economic event that took place during the first quarter of 2020, which we will discuss, in detail, in the following paragraph. Thus, we maintain that we should not be distracted by a one-off event.

A closer look at Figure 4.1 justifies the sharp drop observed in the level of relative risk aversion of the US market. In this figure, we plot RRA and market cap through time. In February 2020, the Fed Chair Mr. Jerome H. Powell, in an attempt to soothe the market which was extremely nervous due to the COVID-19 pandemic, hinted a forthcoming rate cut. In March 2020, Fed announced the new QE program through which it would purchase $600 billion in bank debt, U.S. Treasury notes, and mortgage-backed securities (MBS) from member banks.[1] At the same time, it cut down the federal funds rate by a total of 1.5 percentage points. To comprehend the impact of such an action by the central bank of United States we should refer to equation (4.7). The level of RRA is negatively related to the CE (or equivalently risk-free rate) which is set by the investors. So, assuming that the objective (probabilistic) characteristics of the lottery remain unchanged, as CE gets smaller RRA should increase. Equivalently, the investors would become more risk-averse and so they would require a smaller CE in order to avoid the risky lottery. But this is not the case in the graph. In fact, RRA plummets. To determine the reason behind this drop we should break down the graph in two parts. Up until February 2020, it was the market that had been setting the CE. But in March 2020, Fed set a new lower level of risk-free rate. This new rate was not the result of a change in the investors' level of risk-aversion. It actually resulted from the artificial cut in interest rates by the Federal Reserve Bank. This led to a group of less risk-averse bondholders shifting to the stock market. So, the drop in RRA gives the impression that all bondholders became less risk-averse, which is not true. Assuming that Fed had not intervened, in which case would we have an analogous shift from bonds to stocks by the bondholders? In case the bondholders became more risk-averse. Thus, we conclude that the QE program forced

[1]For further details refer to https://www.federalreserve.gov/

investors to move into risky assets and acted "as-if" the bondholders became more risk-averse. Furthermore, it seems that this action affected the other markets as well, which proves that markets are interconnected when it comes to important economic events.

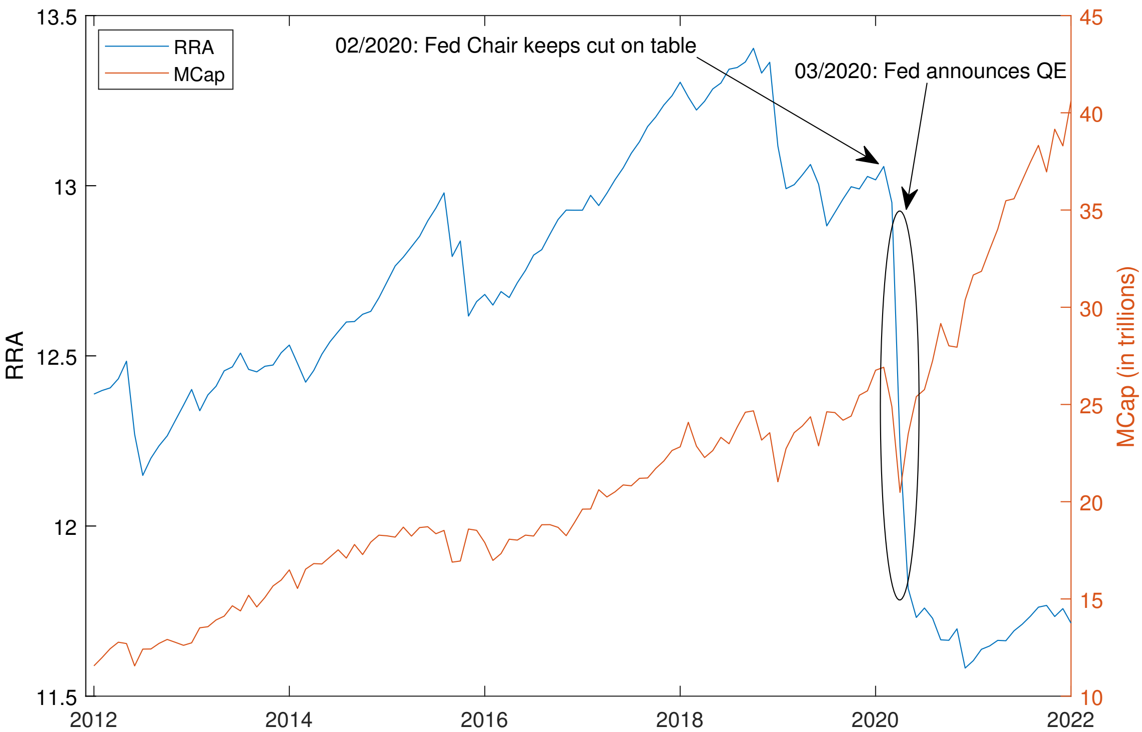


Figure 4.1: RRA: S&P 500

The discussion made in the previous paragraphs leads to the following conclusion. The investors reveal DARA and IRRA for 2012-2022. It needs to be highlighted that this conclusion applies for this time period and should not be treated as a generalization on the investors' risk preferences. In other words, we imply that market participants may have had different risk preferences during different time periods. In fact, we will see later on that for different time periods the slope of the RRA of the different market indices varies. These findings indicate that, currently, we could assume that the investors' class of utility functions should satisfy both DARA and IRRA. The following utility functions are a part of this class.

$$U(w) = (w + a)^c,\ a > 0,\ 0 < c < 1, \quad U(w) = -(w + a)^{-c},\ a > 0,\ c > 0,$$
$$U(w) = \log(w + a),\ a > 0$$

Any utility function from this group serves as an appropriate proxy for the true utility function of the markets. We proceed as follows. First, we will compare our evidence of DARA and IRRA to a rolling-window approach. This will serve as a robustness test of our evidence based on recursive estimations. Then, we will extract the investors' risk preferences through S&P 500 for different time periods, namely, 1993-1998 and 1999-2005.

4.4.1 Rolling-Window Approach

As previously stated, our estimations with regards to μ_R and σ_R were done recursively. One could argue that such an approach is problematic, reasoning that the estimations should better be updated using more recent data. For that reason, we also estimate μ_R and σ_R by employing a rolling-window approach. We test three different window sizes, $M = 60$ (5-years), $M = 120$ (10-years) and $M = 180$ months (15-years). Such window sizes are regularly applied by the literature for monthly data, as in DeMiguel et al. (2009) [24].

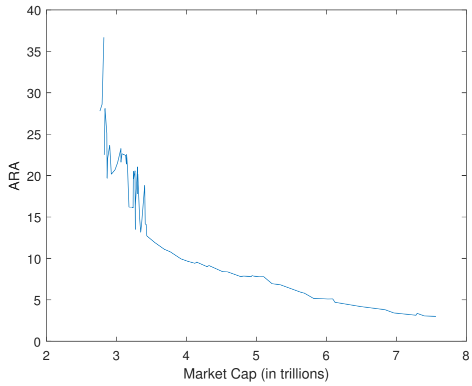
The graphs in Figures 4.8 and 4.9 showcase that for a window size of $M = 60$ months, or equivalently, a 5-year window, there is strong evidence of DARA and IRRA. In fact, the correlation between the ARA and the market cap is highly negative and smaller than -30% for all markets. Accordingly, the correlation between the RRA and the market cap is above 40% for all markets except for S&P 500, for which the magnitude of correlation is mainly affected by the QE program announced by Fed. When we increase the window size to $M = 120$ motnhs, or equivalently, a 10-year window, Figures 4.10 and 4.11, point to the same direction with the exception of S&P 500, in which case it seems that the RRA has a decreasing trend. Lastly, in Figures 4.12 and 4.13, it is evident that we have similar conclusions to those made in the 5-years case. More specifically, for a window size of $M = 180$ months, or equivalently, a 15-year window, the evidence of DARA and IRRA is even stronger, with correlations being higher in absolute terms. Overall, our results for 2012-2022 reveal strong evidence of DARA and IRRA.

4.4.2 RRA and ARA for S&P 500 in Different Time Periods

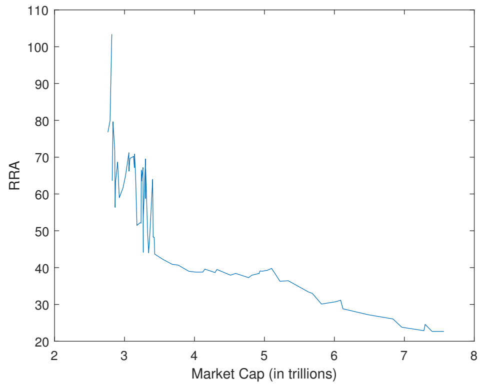
Now, we are going to study ARA and RRA of S&P 500 in different time periods. Doing that will help us determining the risk attitude of US investors across the years. On top of that, we would like to check whether or not the evidence of DRRA and CRRA in earlier studies is supported by our model. For this, we will recover ARA and RRA for 1993-1998 and 1999-2005.

Figure 4.2 displays the level of ARA and RRA for 1993-1998. What is worth noticing is that investors' risk attitude was rather different than currently is, revealing DRRA instead of IRRA. This result supports the evidence of DRRA found in [32], [52], [68] and [71] for earlier time periods. Although some of these works refer to different time periods or different markets they are indicative of the investors' risk attitude overall. Now, the RRA graph on Figure 4.3 is harder to interpret. The graph is indecisive between CRRA and IRRA. Studies like those of [16], [18] and [79] found evidence of CRRA for this time period. However, our graph does not offer a clear conclusion. In terms of the ARA, both Figures reveal DARA which has been validated by most empirical studies, for any time period.

Overall, our empirical evidence demonstrates the importance of studying the risk attitude of investors across different periods. It seems that the market's utility function varies through years. More specifically, our findings point to a utility function satisfying DRRA and DARA for years 1993 through 1998 meaning that investors were willing to take on more risk as they became wealthier. For 1999 to 2005, investors started being more cautious as they reached higher levels of wealth, revealing CRRA/IRRA and DARA. Currently, the different markets seem to become continuously more risk-averse as their market caps grow even further, revealing evidence of IRRA and DARA.

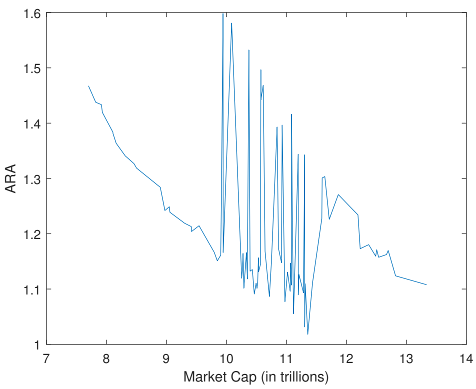


(a) ARA

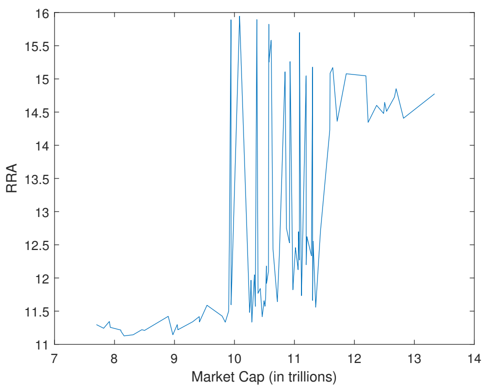


(b) RRA

Figure 4.2: ARA and RRA: 1993-1998



(a) ARA



(b) RRA

Figure 4.3: ARA and RRA: 1999-2005

Following our empirical evidence in the previous sections, a subsequent question would be: How much does a wrong assumption with respect to the utility function of an investor, made by an asset manager, will affect the structure of his portfolio? Could we find an easy and intuitive way to measure the difference between the level of risk that the investor should incur, based on his "true" utility function, and the level of risk that the asset manager recommends, based on his perception of the investor's utility function? We approach these questions in the following section.

4.4.3 Measuring the Differences in Portfolio Diversification among Different Utility Functions

Consider the following case. A financial advisor designs a survey to extract the risk preferences of an investor. The asset manager misinterprets the investor's answers and concludes that the investor reveals IARA and IRRA. In reality however, the investor reveals DARA and IRRA. How much would the decisions made by the asset manager on behalf of the investor be affected? Assuming IARA and IRRA leads to a very well-known utility function, the quadratic $U(w) = w - bw^2$ with $b > 0$. This utility function is considered by the literature to be a sufficient condition for Markowitz's Mean-Variance Optimization method. So, by assuming a quadratic utility the asset manager should use Markowitz's method to diversify the investor's portfolio. In the following paragraph we are going to illustrate an intuitive way to measure the effect of assuming a quadratic rather than a DARA and IRRA utility function like one of those shown below.

$$U(w) = (w+a)^c,\ a > 0,\ 0 < c < 1, \qquad U(w) = -(w+a)^{-c},\ a > 0,\ c > 0,$$

$$U(w) = \log(w+a),\ a > 0 \qquad U(w) = -e^{-c(w+a)},\ a > 0,\ c > 0$$

Assume that at time $t = 0$ the investor's wealth is w_0. The asset manager decides to invest w_s in a risky asset with return R and $w_0 - w_s$ in a riskless asset with rate of return R_f. Both w_0 and w_s are measured in percentages. So, $w_0 = 100\%$ while w_s can take any value. The idea is that if $w_s > w_0$, the asset manager can borrow money to over-invest in the risky asset, while if $w_s < 0$ the asset manager will short-sell the risky asset. So, at point $t = 1$ the investor's wealth will be

$$\begin{aligned} W_1 &= w_s(1+R) + (w_0 - w_s)(1+R_f) \\ &= w_0(1+R_f) + w_s(R - R_f). \end{aligned}$$

Based on VN-M Representation Theorem the asset manager will simply need to maximize the expected utility function of the investor for W_1.

$$\max_{W_1} E[U(W_1)] \Leftrightarrow E[U'(W_1)] = 0$$

Solving the above equation derives w_s. This percentage will differ among different utility functions which means that we can measure the differences between them. In our case, we will find w_s^{quad} from the quadratic utility function $U(w) = w - bw^2$ with $b = 0.2, 0.3, 0.4$ and compare it with w_s^{log} from $U(w) = \log w$. Below, we derive the two quantities. Namely, for $U(w) = w - bw^2$ we have

$$w_s^{quad} = \frac{\mu_R - R_f - 2b(1+R_f)(\mu_R - R_f)}{2b(\mu_R^2 + \sigma_R^2 - 2\mu_R R_f + R_f^2)}$$

In terms of $U(w) = \log(w)$, we are unable to apply the approach we propose directly on it, since we will end up with a fraction equal to 0 that cannot be solved. We can omit this hurdle by taking the Taylor expansion of 2nd-order around 0 as follows.

$$\begin{aligned} U(w_s) &\simeq U(0) + w_s U'(0) + \frac{w_s^2}{2} U''(0) \\ &\simeq \log(1+R_f) + w_s \frac{R - R_f}{1+R_f} - w_s^2 \frac{(R-R_f)^2}{2(1+R_f)^2} \end{aligned}$$

Thus, we obtain

$$w_s^{log} = \frac{(1+R_f)(\mu_R - R_f)}{\mu_R^2 + \sigma_R^2 - 2\mu_R R_f + R_f^2}$$

To measure the difference between the two utility functions we need to use real data. For that reason, we are going to use monthly returns from S&P 500 and 10-year Treasury yields spanning from 2012 to 2022. Then, we will plot only the weight invested in the risky asset, namely w_s, for both the quadratic and the logarithmic utility functions, with respect to time.

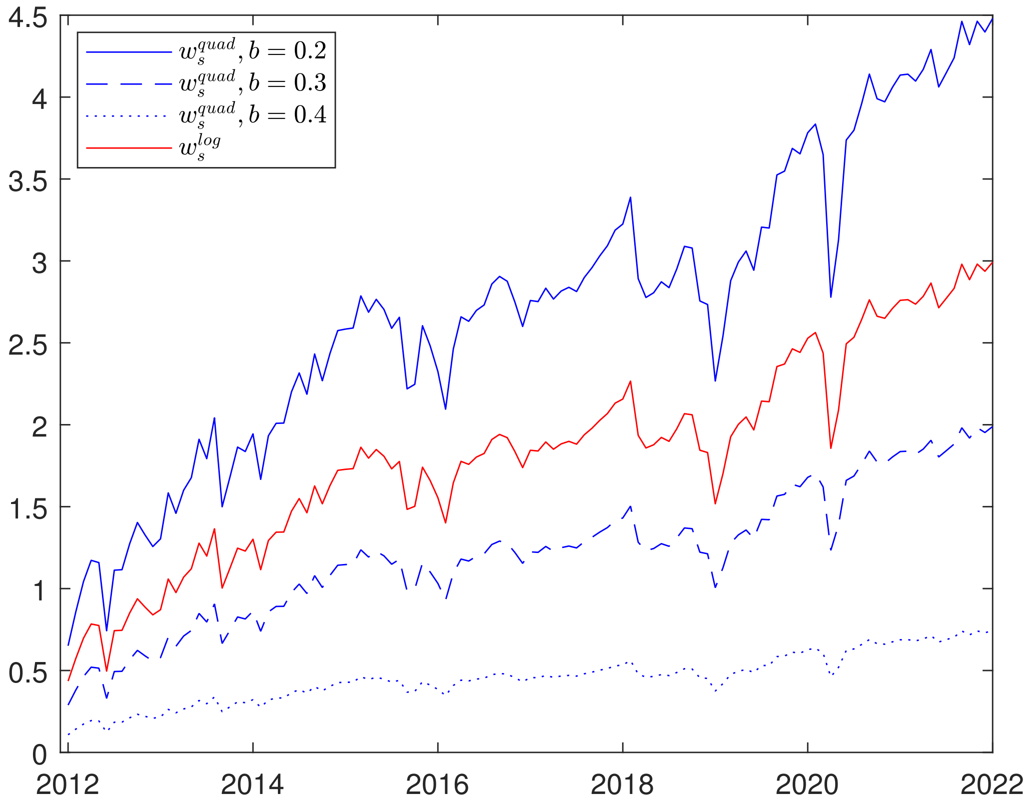


Figure 4.4: Quadratic vs Logarithmic utility's weight in the risky asset

The graph indicates that in case the asset manager had assumed one of the depicted parametrizations of the quadratic utility function for the investor, he would advise him to put significantly different portions of his wealth in S&P 500 each month. Specifically, if $b = 0.2$ the asset manager would underestimate the "true" level of risk-aversion of the investor, by advising him to put 50% more on S&P 500 than what his "true" level of risk tolerance suggests. While if $b = 0.3, 0.4$, the asset manager would overestimate the "true" level of risk-aversion of the investor, by almost 50% to 100%. So, the asset manager's misjudgment would lead the investor to incur substantially more or less risk than he could actually tolerate. We should highlight here that even if we consider the case in which the "true" utility function of the investor is quadratic with $b = 0.3$, the weight invested in the risky asset, namely w_s, is substantially higher through all years, compared to the $b = 0.4$ case. So, even if the asset manager correctly detects that the investor's utility is quadratic, there is still plenty of room to underestimate (or overestimate) his true level of risk-aversion.

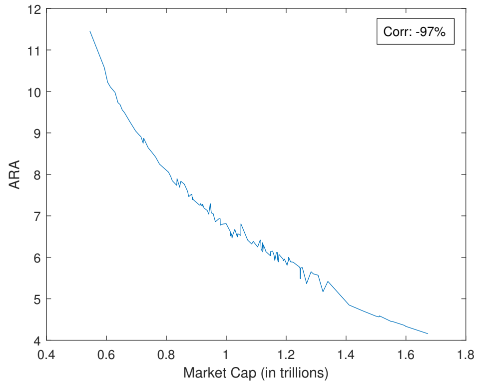
To conclude, we designed a simple and intuitive method to measure the differences between alternative utility functions, with relative accuracy.

4.5 Conclusions

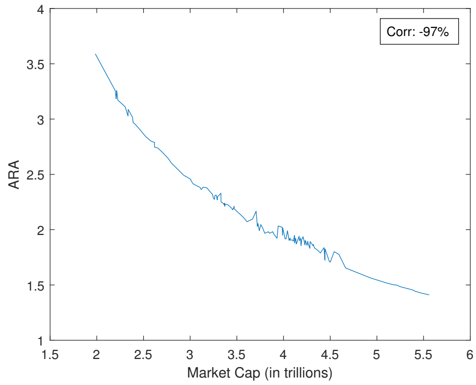
In this work, we overviewed the various ways the literature has extracted the ARA and the RRA of investors through the years. We saw that, the literature universally accepts DARA but is not conclusive with respect to the RRA of investors. In fact, one part of previous research works reveals evidence of DRRA, another part finds evidence of IRRA and the most recent works point to CRRA. We asserted that using cross-section data as was regularly done in earlier research works, conceals risks. More specifically, cross-sectional analysis implicitly says that all investors have the same utility function, which is a rather unrealistic assumption. Studies basing their evidence on either panel data or Options markets data are better supported.

We proposed a different approach in extracting both ARA and RRA. This approach, is based on carefully analyzing the Theory of Arrow-Pratt together with the different notions that underlie the Decision Theory under risk. In particular, we derived closed-form expressions for the ARA and the RRA of investors, for non-“fair” lotteries. We proceeded with collecting data from different markets which we then applied on our formulae. Our findings, pointed out that through 2012-2022, European as well as US markets revealed evidence of DARA and IRRA. We further showcased that our formula captures important economic events as in the case of the QE program announced by Fed in March 2020, which caused a sharp drop in the investors’ level of relative risk-aversion. Our results were further tested using a rolling-window approach for different window sizes $M = 60$, $M = 120$ and $M = 180$ months. The rolling-window results supported our evidence of DARA and IRRA for 2012-2022. Then, we found out that for different time periods, namely, 1993-1998 and 1999-2005 RRA may vary. More specifically, for 1993-2005 we found strong evidence of DARA and DRRA while for 1999-2005 we found strong evidence of DARA but in terms of RRA our results were indecisive between CRRA and IRRA. In the last part of our work, we proposed a simple way to measure the effect of a wrong assumption with respect to the utility function of an investor. As it became clear, an investor with a logarithmic utility function is led to a very different portfolio structure compared to an investor with a quadratic utility function.

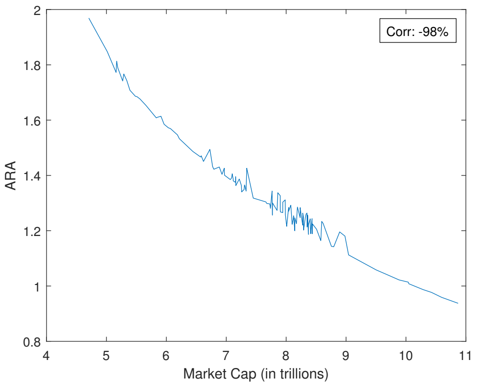
Appendix



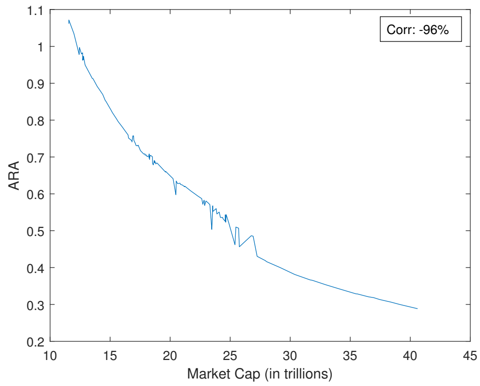
(a) CAC 40



(b) EURO

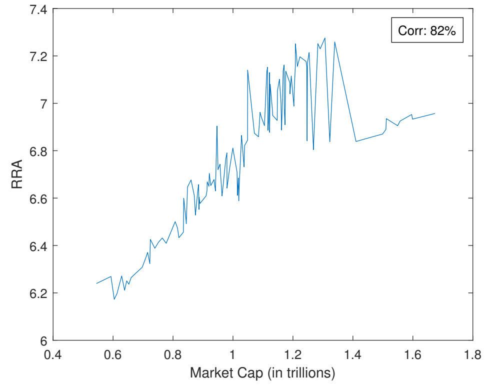


(c) STOXX 600

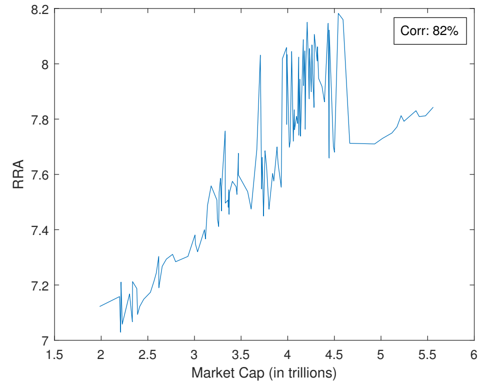


(d) S&P 500

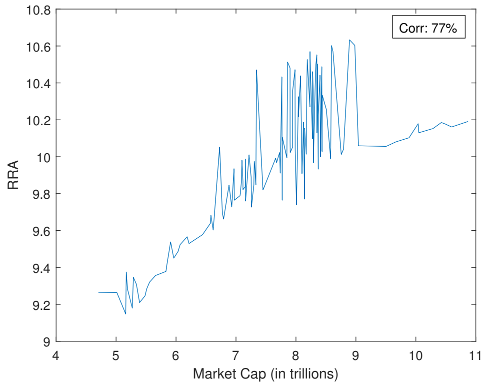
Figure 4.5: ARA with sorted wealth



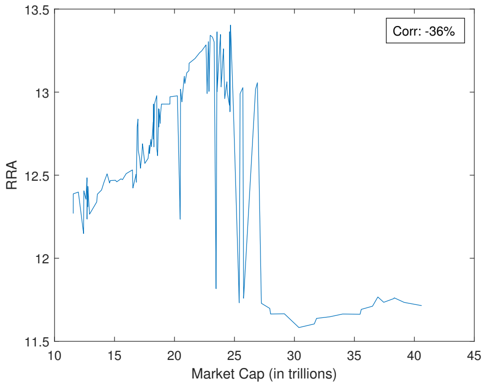
(a) CAC 40



(b) EURO

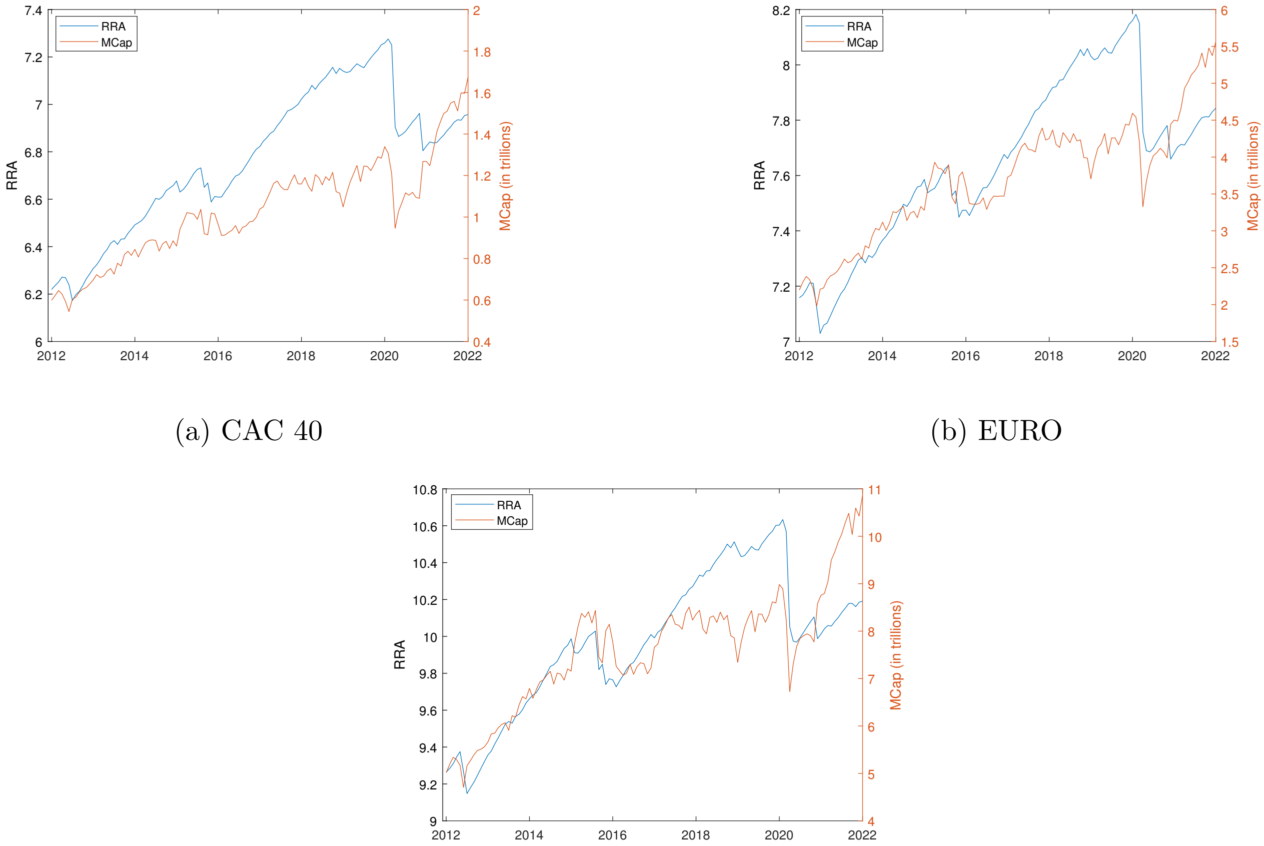


(c) STOXX 600



(d) S&P 500

Figure 4.6: RRA with sorted wealth

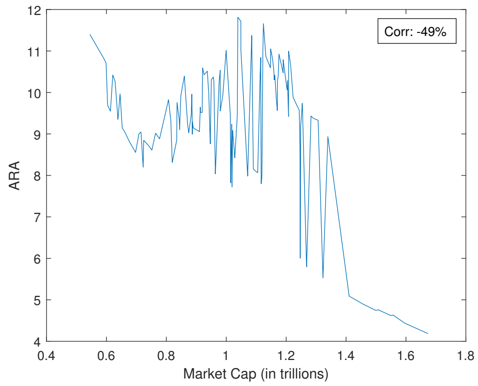


(a) CAC 40

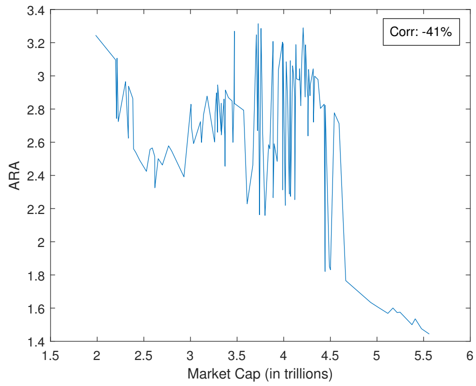
(b) EURO

(c) STOXX 600

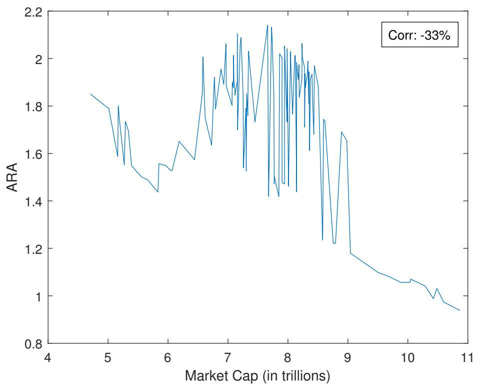
Figure 4.7: RRA in time



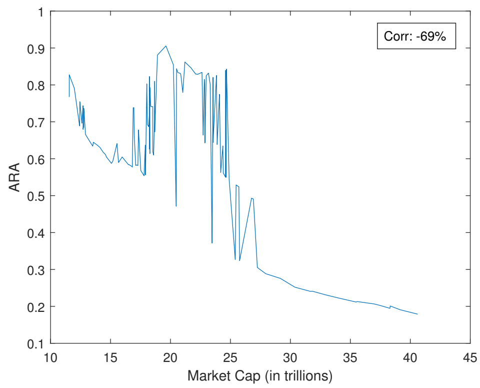
(a) CAC 40



(b) EURO

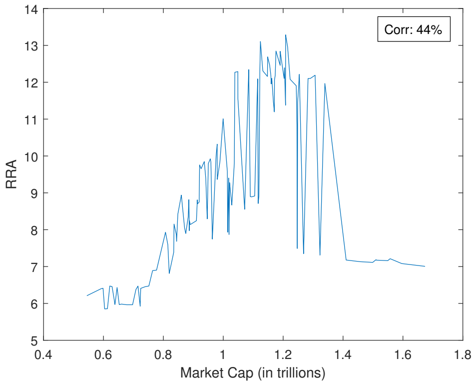


(c) STOXX 600

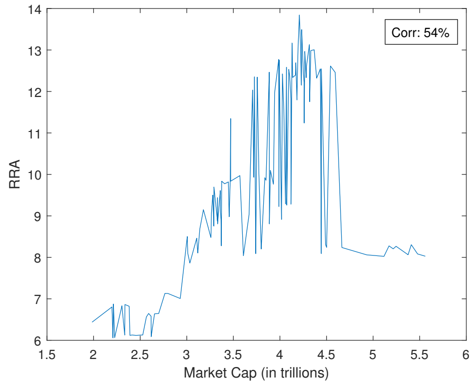


(d) S&P 500

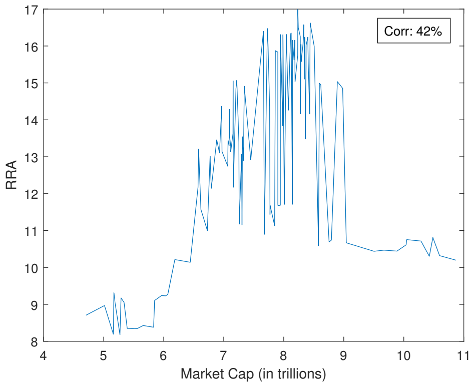
Figure 4.8: ARA with $M = 60$



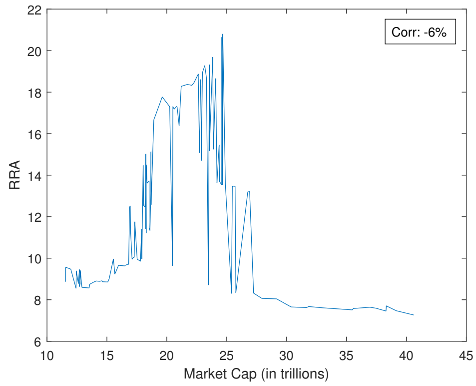
(a) CAC 40



(b) EURO

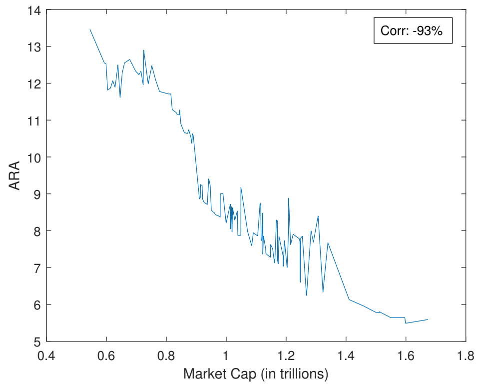


(c) STOXX 600

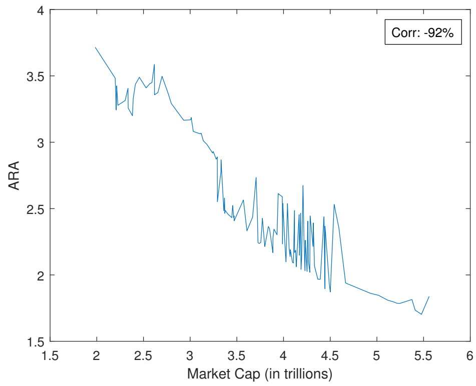


(d) S&P 500

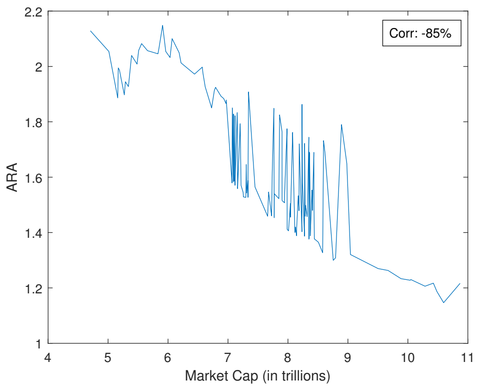
Figure 4.9: RRA with $M = 60$



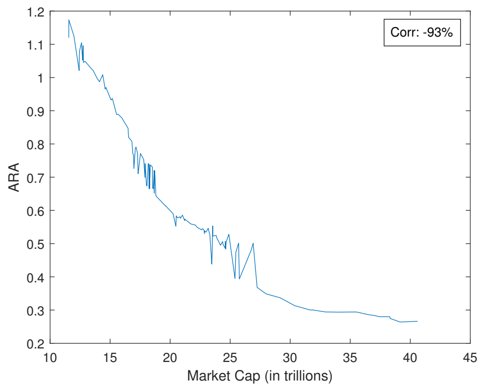
(a) CAC 40



(b) EURO

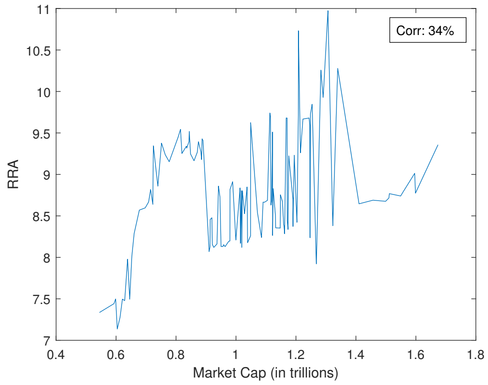


(c) STOXX 600

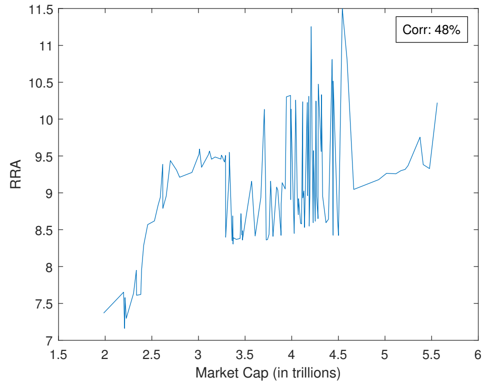


(d) S&P 500

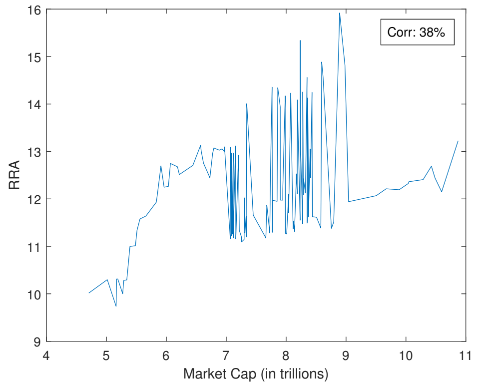
Figure 4.10: ARA with $M = 120$



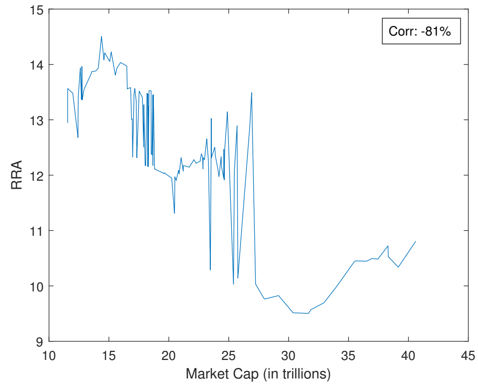
(a) CAC 40



(b) EURO

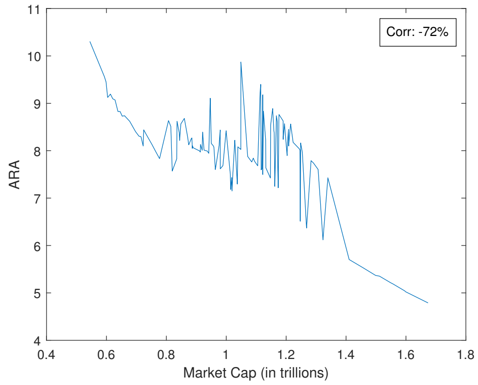


(c) STOXX 600

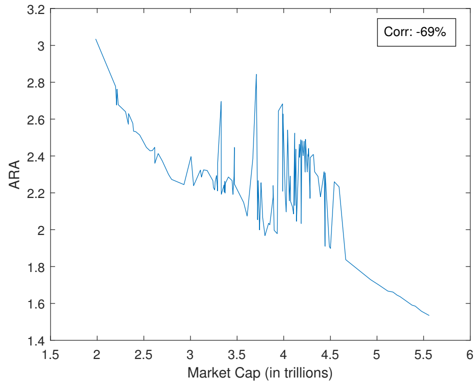


(d) S&P 500

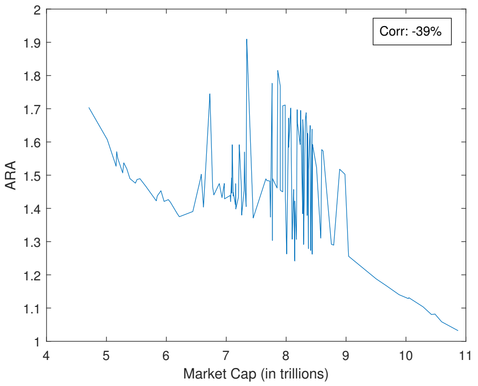
Figure 4.11: RRA with $M = 120$



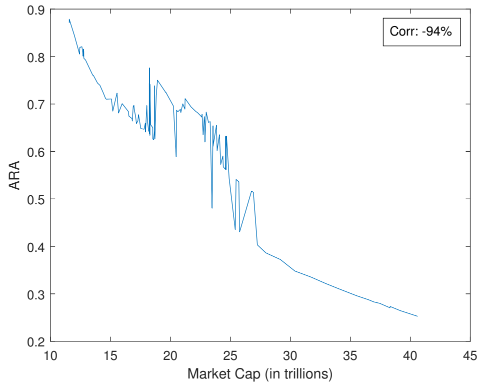
(a) CAC 40



(b) EURO

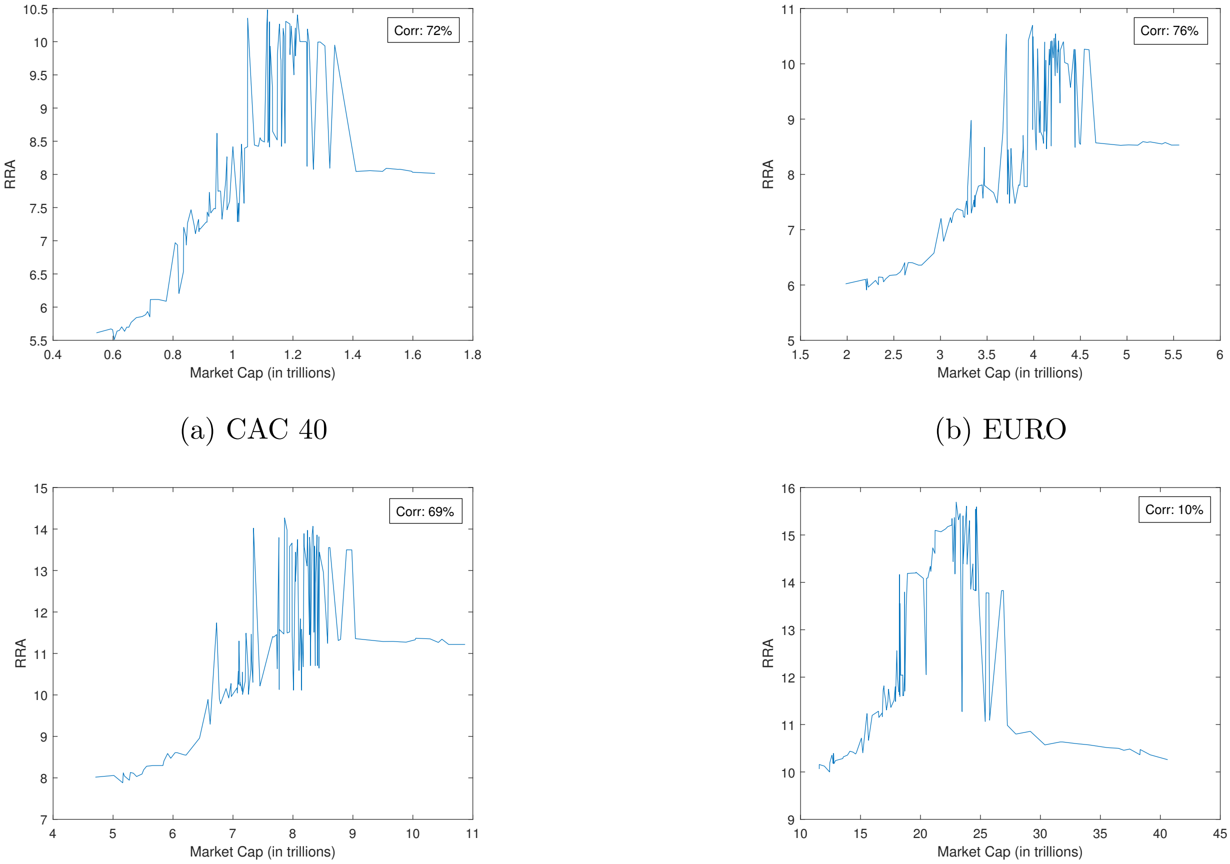


(c) STOXX 600



(d) S&P 500

Figure 4.12: ARA with $M = 180$



(a) CAC 40

(b) EURO

(c) STOXX 600

(d) S&P 500

Figure 4.13: RRA with $M = 180$

Bibliography

[1] Agnew, R. A. “Counter-examples to an Assertion Concerning the Normal Distribution and a New Stochastic Price Fluctuation Model”. In: *The Review of Economic Studies* 38.3 (1971), pp. 381–383.

[2] Aït-Sahalia, Yacine and Lo, Andrew W. “Nonparametric risk management and implied risk aversion”. In: *Journal of Econometrics* 94.1-2 (2000), pp. 9–51.

[3] Andrews, Donald W. K. “Tests for Parameter Instability and Structural Change With Unknown Change Point”. In: *Econometrica* 61.4 (1993), pp. 821–856.

[4] Arrow, Kenneth J. *Aspects of The Theory of Risk-Bearing*. Helsinki: Yrjo Jahnsson Foundation, 1965.

[5] Bai, B Y Jushan and Perron, Pierre. “Estimating and Testing Linear Models with Multiple Structural Changes”. In: *Econometrica* 66.1 (1998), pp. 47–78.

[6] Bai, Jushan and Perron, Pierre. “Computation and analysis of multiple structural change models”. In: *Journal of Applied Econometrics* 18.1 (2002), pp. 1–22.

[7] Banerjee, Anindya and Urga, Giovanni. “Modelling structural breaks, long memory and stock market volatility: An overview”. In: *Journal of Econometrics* 129.1-2 (2005), pp. 1–34.

[8] Bar-Shira, Z., Just, R. E., and Zilberman, D. “Estimation of farmers’ risk attitude: An econometric approach”. In: *Agricultural Economics* 17.2-3 (1997), pp. 211–222.

[9] Baron P., David. “On The Utility Theoretic Foundations Of Mean-Variance Analysis”. In: *The Journal of Finance* XXXII.5 (1977), pp. 1683–1697.

[10] Barry, Christopher B. "Portfolio Analysis Under Uncertain Means, Variances, and Covariances". In: *The Journal of Finance* 29.2 (1974), p. 515.

[11] Barsky, B. Robert et al. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study". In: *The Quarterly Journal of Economics* 112.2 (1997), pp. 537–579.

[12] Bawa, Vijay S. and Klein, Roger W. "The effect of estimation risk on optimal portfolio choice". In: *Journal of Financial Economics* 3.3 (1976), pp. 215–231.

[13] Best, Michael J and Grauer, Robert R. "On the Sensitivity of Mean- Variance-Efficient Portfolios to Changes in Asset Means : Some Analytical and Computational Results". In: *The Review of Financial Studies* 4.2 (1991), pp. 315–342.

[14] Broadie, Mark. "Computing efficient frontiers using estimated parameters". In: *Annals of Operations Research* 45.1 (1993), pp. 21–58.

[15] Brown, R. L., Durbin, J., and Evans, J. M. "Techniques for Testing the Constancy of Regression Relationships Over Time". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 37.2 (1975), pp. 149–163.

[16] Brunnermeier, Markus K. and Nagel, Stefan. "Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals' asset allocation". In: *The American Economic Review* 98.3 (2008), pp. 713–736.

[17] Chamberlain, Gary. "A Characterization of The Distributions that Imply Mean-Variance Utility Functions". In: *Journal of Economic Theory* 29.1 (1983), pp. 185–201.

[18] Chiappori, Pierre André and Paiella, Monica. "Relative risk aversion is constant: Evidence from panel data". In: *Journal of the European Economic Association* 9.6 (2011), pp. 1021–1052.

[19] Chopra, K. Vijay and Ziemba, T. William. "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Theory". In: *The Journal of Portfolio Management* 111.479 (1993), pp. 6–11.

[20] Chopra, Vijay K., Hensel, Chris R., and Turner, Andrew L. "Massaging Mean-Variance Inputs: Returns from Alternative Global Investment Strategies in the 1980s". In: *Management Science* 39.7 (1993), pp. 845–855.

[21] Chow, Gregory C. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions Author(s): Gregory C. Chow Source:" in: *Econometrica* 28.3 (1960), pp. 591–605.

[22] Chu, Chia Shang James, Santoni, Gary J., and Liu, Tung. "Stock market volatility and regime shifts in returns". In: *Information Sciences* 94.1-4 (1996), pp. 179–190.

[23] Cohn, Richard A et al. "Individual Investor Risk Aversion and Investment Portfolio Composition". In: *The Journal of Finance* 30.2 (1975), pp. 605–620.

[24] DeMiguel, Victor, Garlappi, Lorenzo, and Uppal, Raman. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *The Review of Financial Studies* 22.5 (2009), pp. 1915–1953.

[25] Dickinson, J. P. "The Reliability of Estimation Procedures in Portfolio Analysis". In: *The Journal of Financial and Quantitative Analysis* 9.3 (1974), pp. 447–462.

[26] Eisenhauer, Joseph G. and Ventura, Luigi. "Survey measures of risk aversion and prudence". In: *Applied Economics* 35.13 (2003), pp. 1477–1484.

[27] Feldstein, M. S. "Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection". In: *The Review of Economic Studies* 36.1 (1969), pp. 5–12.

[28] Frankfurter, George M., Phillips, Herbert E., and Seagle, John P. "Portfolio Selection: The effects of Uncertain Means, Variances and Covariances". In: *The Journal of Financial and Quantitative Analysis* 6.2 (1971), pp. 1251–1262.

[29] Friend, Irwin and Blume, Marshall E. "The Demand for Risky Assets". In: *The American Economic Review* 65.5 (1975), pp. 900–922.

[30] Gandhi K., Devinder and Saunders, David. "The Superiority of Stochastic Dominance Over Mean Variance Efficiency Criteria: Some Clarifications". In: *Journal of Business Finance & Accounting* 1 (1981), pp. 51–60.

[31] Granger, Clive W.J. and Hyung, Namwon. “Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns”. In: *Journal of Empirical Finance* 11.3 (2004), pp. 399–421.

[32] Guiso, Luigi, Jappelli, Tullio, and Terlizzese, Daniele. “Income Risk, Borrowing Constraints, and Portfolio Choice”. In: *The American Economic Review* 86.1 (1996), pp. 158–172.

[33] Hadar, Josef and Russell, William R. “American Economic Association Rules for Ordering Uncertain Prospects”. In: *The American Economic Review* 59.1 (1969), pp. 25–34.

[34] Hanoch, Giora and Levy, Haim. “Efficient Portfolio Selection with Quadratic and Cubic Utility”. In: *The Journal of Business* 43.2 (1970), pp. 181–189.

[35] Hanoch, Giora and Levy, Haim. “The Efficiency Analysis of Choices Involving Risk”. In: *The Review of Economic Studies* 36.3 (1969), pp. 335–346.

[36] Harvey, Campbell R. et al. “Portfolio Selection with Higher Moments”. In: *Quantitative Finance* 10.5 (2004), pp. 1–50.

[37] Jackwerth, Jens Carsten. “Recovering Risk Aversion from Option Prices and Realized Returns”. In: *The Review of Financial Studies* 13.2 (2000), pp. 433–451.

[38] Jagannathan, Ravi and Ma, Tongshu. “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps”. In: *The Journal of Finance* 58.4 (2003), pp. 1651–1683.

[39] Jobson, J. D. and Korkie, B. “Estimation for Markowitz Efficient Portfolios”. In: *Journal of the American Statistical Association* 75.371 (1980), pp. 544–554.

[40] Jobson, J. D. and Korkie, B. “Putting Markowitz theory to work”. In: *The Journal of Portfolio Management* 41.5 (1981), pp. 1–5.

[41] Johnstone, D. J. and Lindley, D. V. “Elementary Proof that Mean-Variance Implies Quadratic Utility”. In: *Theory and Decision* 70.2 (2011), pp. 149–155.

[42] Jondeau, Eric and Rockinger, Michael. “On the importance of time variability in higher moments for asset allocation”. In: *Journal of Financial Econometrics* 10.1 (2012), pp. 84–123.

[43] Jondeau, Eric and Rockinger, Michael. "Optimal Portfolio Allocation under Higher Moments". In: *European Financial Management* 12.1 (2006), pp. 29–55.

[44] Kan, Raymond and Zhou, Guofu. "Optimal Portfolio Choice with Parameter Uncertainty". In: *The Journal of Financial and Quantitative Analysis* 42.3 (2007), pp. 621–656.

[45] Killick, R., Fearnhead, P., and Eckley, I. A. "Optimal detection of changepoints with a linear computational cost". In: *Journal of the American Statistical Association* 107.500 (2012), pp. 1590–1598.

[46] Kourtis, Apostolos, Dotsis, George, and Markellos, Raphael N. "Parameter Uncertainty in Portfolio Selection: Shrinking the Inverse Covariance Matrix". In: *Journal of Banking and Finance* 36.9 (2012), pp. 2522–2531.

[47] Kroll, Yoram, Levy, Haim, and Harry M., Markowitz. "Mean-Variance versus Direct Utility Maximization: A Comment". In: *The Journal of Finance* 49.1 (1984), pp. 47–61.

[48] Lai, Tze Leung, Xing, Haipeng, and Chen, Zehao. "Mean-Variance Portfolio Optimization when Means and Covariances are Unknown". In: *The Annals of Applied Statistics* 5.2A (2011), pp. 798–823.

[49] Lavielle, Marc and Moulines, Eric. "Least-Squares Estimation Of An Unknown Number Of Shifts In A Time Series". In: *Journal of Time Series Analysis* 21.1995 (2000), p. 662.

[50] Ledoit, Olivier and Wolf, Michael. "Honey, I Shrunk the Sample Covariance Matrix". In: *The Journal of Portfolio Management* (2004), pp. 1–21.

[51] Ledoit, Olivier and Wolf, Michael. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection". In: *Journal of Empirical Finance* (2001), pp. 1–21.

[52] Levy, Haim. "Absolute and Relative Risk Aversion: An Experimental Study". In: *Journal of Risk and Uncertainty* 8 (1994), pp. 289–307.

[53] Levy, Haim. *Stochastic Dominance Investment Decision Making under Uncertainty.* Vol. 30. 9. 2016, p. 845.

[54] Levy, Haim. "The Rationale of the Mean-Standard Deviation Analysis: Comment". In: *The American Economic Review* 64.3 (1974), pp. 434–441.

[55] Levy, Haim and Duchin, Ran. "Asset return distributions and the investment horizon". In: *The Journal of Portfolio Management* 30.3 (2004), pp. 47–62.

[56] Levy, Haim and Levy, Moshe. "Prospect theory and mean-variance analysis". In: *Review of Financial Studies* 17.4 (2004), pp. 1015–1041.

[57] Levy, Haim and Markowitz, Harry M. "Approximating Expected Utility by a Function of Mean and Variance". In: *The American Economic Review* 69.3 (1979), pp. 308–317.

[58] Levy, Moshe. "Market Efficiency, the Pareto Wealth Distribution, and the Lévy Distribution of Stock Returns". In: *The Economy as an Evolving Complex System, III: Current Perspectives and Future Directions* October (2005), pp. 1–52.

[59] Linden, Mikael. "A model for stock return distribution". In: *International Journal of Finance & Economics* 6.2 (2001), pp. 159–169.

[60] Malavasi, Matteo, Ortobelli Lozza, Sergio, and Trück, Stefan. "Second order of stochastic dominance efficiency vs mean variance efficiency". In: *European Journal of Operational Research* 290.3 (2020), pp. 1192–1206.

[61] Mandelbrot, Benoit. "The Variation of Certain Speculative Prices". In: *The Journal of Business* 36.4 (1963), p. 394.

[62] Mantegna, Rosario N. and Stanley, H. Eugene. "Scaling behaviour in the dynamics of an economic index". In: *Nature* 376.6535 (1995), pp. 46–49.

[63] Markowitz, Harry M. "Mean-variance approximations to expected utility". In: *European Journal of Operational Research* 234.2 (2014), pp. 346–355.

[64] Markowitz, Harry M. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

[65] Markowitz, Harry M. *Portfolio Selection: Efficient Diversification of Investments.* Vol. 244. 21. 1959, pp. 1–356.

[66] Markowitz, Harry M. "Portfolio theory: As I still see it". In: *Annual Review of Financial Economics* 2 (2010), pp. 1–23.

[67] Michaud, Richard O. “The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal?” In: *Financial Analysts Journal* March (1989), pp. 1–12.

[68] Morin, Roger-A. and Suarez, A. Fernandez. “Risk Aversion Revisited”. In: *The Journal of Finance* 38.4 (1983), pp. 1201–1216.

[69] Ogaki, Masao and Zhang, Qiang. “Decreasing Relative Risk Aversion and Tests of Risk Sharing”. In: *Econometrica* 69.2 (2001), pp. 515–526.

[70] Owen, Joel and Rabinovitch, Ramon. “On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice”. In: *The Journal of Finance* 38.3 (1983), pp. 745–752.

[71] Pérignon, Christophe and Villa, Christophe. “Extracting information from options markets: Smiles, state–price densities and risk aversion”. In: *European Financial Management* 8.4 (2002), pp. 495–513.

[72] Pesaran, Hashem M., Pettenuzzo, Davide, and Timmermann, Allan. “Forecasting time series subject to multiple structural breaks”. In: *Review of Economic Studies* 73.4 (2006), pp. 1057–1084.

[73] Pesaran, M. Hasherm and Timmermann, Allan. “How costly is it to ignore breaks when forecasting the direction of a time series?” In: *International Journal of Forecasting* 20.3 (2004), pp. 411–425.

[74] Pettenuzzo, Davide and Timmermann, Allan. “Predictability of stock returns and asset allocation under structural breaks”. In: *Journal of Econometrics* 164.1 (2011), pp. 60–78.

[75] Pflug, Georg Ch, Pichler, Alois, and Wozabal, David. “The 1/N investment strategy is optimal under high model ambiguity”. In: *Journal of Banking and Finance* 36.2 (2012), pp. 410–417.

[76] Pratt, John W. “Risk Aversion in the Small and in the Large Author”. In: *Econometrica* 32.1 (1964), pp. 122–136.

[77] Quirk, James P. and Saposnik, Rubin. “Admissibility and Measurable Utility Functions”. In: *The Review of Economic Studies* 3 (1962), pp. 140–146.

[78] Rapach, David E. and Wohar, Mark E. “Structural breaks and predictive regression models of aggregate U.S. stock returns”. In: *Journal of Financial Econometrics* 4.2 (2006), pp. 238–274.

[79] Sahm, Claudia R. *How Much Does Risk Tolerance Change?* Vol. 2. 4. 2012, pp. 1–38.

[80] Schuhmacher, Frank, Kohrs, Hendrik, and Auer, Benjamin R. “Justifying Mean-Variance Portfolio Selection when Asset Returns Are Skewed”. In: *Management Science* 67.12 (2021), pp. 7812–7824.

[81] Sen, Ashish and Srinvastava, S. Muni. “On Tests for Detecting Change in Mean”. In: *The Annals of Statistics* 19.3 (1975), pp. 1403–1433.

[82] Sharpe, William F. “A Theory of Market Equilibrium under Conditions of Risk”. In: *The Journal of Finance* 19.3 (1964), pp. 425–442.

[83] Siegel, Frederick W. and Hoban, James P. “Relative Risk Aversion Revisited”. In: *The Review of Economics and Statistics* 28.8 (1982), pp. 875–886.

[84] Stein, Charles. “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution”. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (1956), pp. 1–11.

[85] Timmermann, Allan. “Structural Breaks, Incomplete Information, and Stock Prices”. In: *Journal of Business & Economic Statistics* 19.3 (2001), pp. 299–314.

[86] Tobin, James. “Liquidity preference as behavior towards risk”. In: *The Review of Economic Studies* 25.2 (1958), pp. 65–86.

[87] Von Neumann, John and Morgenstern, Oskar. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press. 3rd ed., 1944.

[88] Whitmore, G. A. “Third-Degree Stochastic Dominance”. In: *The American Economic Review* 60.3 (1970), pp. 457–459.

[89] Xiong, James X. and Idzorek, Thomas M. “The impact of skewness and fat tails on the asset allocation decision”. In: *Financial Analysts Journal* 67.2 (2011), pp. 23–35.