Università
Ca'Foscari
Venezia

Master's Degree programme –
Second Cycle (*D.M. 270/2004*)
in Economics and Finance

Final Thesis

On the unbiased efficient frontier: some analytical and computational remarks

Supervisor
Ch. Prof. Marco Corazza
Ch. Prof. Claudio Pizzi

Graduand
Sara Cocchi
Matriculation Number 853774

Academic Year
2015 / 2016

Abstract

One of the most important and influential economic theories dealing with finance and investment is Modern Portfolio Theory. MPT has its roots in the work developed by Markowitz, resumed by Merton, and then studied and revisited by many authors, which still try to deal with the weaknesses of the theory. Starting from the approach proposed by Bodnar and Bodnar, which tries to correct the bias witnessed in the classical formulation of the sample efficient frontier, the scope of this work will be that of analyzing whether the unbiased estimator of the efficient frontier respects the properties which have been attributed to it by Corazza and Pizzi, if it is operationally effective and if it is robust to small departures from the assumption of independence. In order to do this, we will make exploratory analysis on our dataset and we will construct the different frontiers and confidence intervals with the use of the programming language and statistical software R. We will also apply differential sensitivity analysis to formulate a new proposition related to the unbiased estimator proposed by Bodnar and Bodnar. What we obtain is that the model is able to correct the bias of the classical sample efficient frontier, but also that it is not robust when changes to the degree of independence are introduced.

"If I have seen further it is by standing on the shoulders of giants."

Isaac Newton

To my family.

Contents

Introduction

In 1952 Harry Markowitz established a new approach for optimal asset allocation and portfolio selection, known as the mean-variance analysis. Since this seminal work, the mean-variance efficient frontier has been given considerable attention and has played a key role in a variety of financial fields, from investment analysis and asset pricing to topics in corporate finance. As a matter of fact Markowitz (1952) proposed to determine the frontier of efficient portfolios by maximizing the expected return for a given level of risk, or equivalently, by minimizing the risk given an average portfolio return.

> "The main innovation introduced by Markowitz was to measure the risk of a portfolio via the joint (multivariate) distribution of returns of all assets. Multivariate distributions are characterized by the statistical (marginal) properties of all component random variables and by their dependence structure. Markowitz described the former by the first two moments of the univariate distributions – the asset returns – and the latter via the linear (Pearson) correlation coefficient between each pair of random returns [. . .]"[1]

Merton (1972) showed that the set of all these optimal portfolios is lying on a parabola in the mean-variance space, called the efficient frontier. This parabola is uniquely characterized by the following parameters, namely:

- the expected return, and
- the variance of the global minimum variance portfolio[2].

From a theoretical point of view, this approach to portfolio management seems simple, but actually the investor is faced with a number of difficulties when implementing it in practice. As a matter of fact, the parameters characterizing the efficient frontier are unknown quantities. We thus generally estimate them using historical data. But, when replacing the parameters by their counterpart estimators we obviously obtain an estimate of the efficient frontier, the sample efficient frontier. As such, this sample efficient frontier, will be characterized by an estimation error.

From this problem it stems the need to determine the unbiased efficient frontier and to construct confidence intervals for the sample one. This problem in the financial literature is known as the uncertainty about the parameters of the asset returns.

[1] Szegö, G.: Measures of risk, European Journal of Operational Research 163 (2005) 5–19

[2] The global minimum variance portfolio (GMV) is the portfolio with the highest return and the least risk. It takes into account all securities available and uses their returns, variances, and correlations to reduce as much of the non-systematic risk as possible.

Many are the papers concerned in the statistical inference procedures for the mean-variance efficient frontier.

Jobson (1991) starting from the fact that an estimator of the efficient frontier is composed of three mutually independent statistics whose distributions are known, developed two alternative approaches for developing a confidence region for the mean-variance efficient set.

Siegel and Woodgate (2007) explain the poor out-of-sample performance of mean-variance efficient portfolios, developing theoretical bias adjustments for estimation risk by asymptotically expanding future returns of portfolios formed with estimated weights.

Kan and Smith (2008) show that the sample mean-variance efficient frontier is a highly biased estimator of the population frontier, and propose an improved estimator of the population frontier. They also provide the exact distribution of the out-of-sample mean and variance of sample mean-variance efficient portfolios, allowing us to understand the impact of estimation error on the performance of in-sample optimal portfolios.

Bodnar and Schmid (2008) proposed a different parameterization of the efficient frontier.

Bodnar and Schmid (2009) derive tests and confidence sets for all possible arrangements of the three parameters determining the mean-variance efficient frontier and they also determine a confidence region of the whole efficient frontier in the mean-variance space.

Bodnar and Bodnar (2009) suggest finite sample tests for the location of the efficient frontier with the estimated parameters in mean-variance space.

Bodnar and Bodnar (2010) derive an unbiased estimator of the efficient frontier correcting for the overoptimism of the sample efficient frontier documented in Siegel and Woodgate (2007), and present an exact F-test on the efficient frontier.

If the financial literature has extensively treated this argument, why should we worry about it? Analyzing more deeply all the above-mentioned papers, there are two assumptions which are common to all of them. As a matter of fact, all these articles, papers, works and studies, assume that the portfolio asset returns are multivariate normally distributed and independent.

As far as the normality assumption is concerned, it has been evidenced that it can work in the mean-variance framework. For what concerns, instead, the assumption of independent returns, it seems too stringent.

The aim of this work is thus that of investigating which impact does it have on the model predictive ability, a departure from its assumption of independence of logarithmic asset returns. In order to study this impact we will make our analyses on the unbiased efficient frontier proposed by Bodnar and Bodnar (2010).

In this paper we will proceed as follows: in Chapter 1 we will introduce the Modern Portfolio Theory, we will analyze the main results of Markowitz (1952) and Merton (1972).

In Chapter 2 we will deal with a review of the literature regarding inference procedures of the efficient frontier, and we will present the steps by which Bodnar and Bodnar derive their unbiased estimator of the efficient frontier, but before doing this we will show the results of Tu and Zhou (2004), Basak, Jagannathan and Ma (2005), Siegel and Woodgate (2007), Kan and Smith (2008); all works which are somehow related to that of Bodnar and Bodnar (2010).

In Chapter 3 we will deepen our the analysis regarding the estimator proposed by Bodnar and Bodnar, we will present the result of Corazza and Pizzi (2014) and further develop these results. In this chapter, we finally present our datasets and their peculiarities.

In Chapter 4 we practically apply the notion learned in the previous chapter on the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar and we will see whether it works, whether it shows asymptotic behavior and whether it is operationally effective. The chapter also presents a discussion about the slope of the unbiased estimator.

In Chapter 5 we try to test the robustness of the model and to do that we progressively try to reintroduce some degree of independence in our datasets.

Chapter 1

Portfolio selection theory

1.1 Markowitz portfolio selection problem

> "A good portfolio is more than a long list of good stocks and bonds. It is a balanced whole, providing the investor with protections and opportunities with respect to a wide range of contingencies. The investor should build toward an integrated portfolio which best suits his needs."[1]

In these few lines contained in the introduction to his seminal work firstly published in 1959, Markowitz is able to concentrate all the important matters, changes, and innovations that this book contains. In a clear-cut way he disentangles all these topics taking the reader from the basic assumptions of its model to the very results of his work, commonly known as Modern Portfolio Theory (MPT).

Modern portfolio theory is based on four behavioral assumptions.

- All investors evaluate each investment opportunity as being represented by a probability distribution of returns, measured over the same holding period.
- All investors are price-taker individuals, meaning that their actions do not influence market prices.
- Investors' risk estimates are based on and are proportional to the variability of returns, whether it is measured by the standard deviation, or by the variance of returns.
- Investors' utility function, $U(r)$, is a function of the variability of return (σ) and the expected return [$E(r)$] only.
- For any risk level, investors prefer higher returns to lower returns. Conversely, for any given level of rate of return, investors prefer less risk over more risk. This assumption implies that investors are rational and risk averse.

Modern portfolio theory does also take into account some other technical assumptions.

[1] Markowitz, H.: Portfolio selection: efficient diversification of investments, 2. ed Oxford Basil Blackwell (1991)

- Financial markets are frictionless, meaning that we do not have transaction costs, taxes, and the like.
- Individual assets on which investors invest are infinitely divisible.

Thus, according to all of these assumptions the steps that Markowitz recognizes for the choice of the optimal portfolio are the following:

1. Identify appropriate measures for risk and return.
2. Separate efficient portfolios from inefficient ones, based on the mean-variance criterion.
3. Select a proper portfolio for the investor, according to his or her risk aversion.

1.1.1 Measures for risk and return

Concerning the first step of this selection problem, how does Markowitz defines the above measures of risk and return?

In its classical formulation, Markowitz chooses the expected value, or the mean, of the rate of return as a proxy of the expected return and the variance of return as a measure of uncertainty, risk, and instability.

We start from the formulae of expected value and variance of individual securities. If we let $X = x_1, x_2, ..., x_n$ be a discrete random variable describing the possible returns from a given asset, and $p = p_1, p_2, ..., p_n$ the associated probability of realization, we know that they can be defined as follows:

$$E(X) = \sum_{i=1}^{n} x_i p_i \tag{1.1}$$

$$Var(X) = \sum_{i=1}^{n} (x_i - E(X))^2 p_i \tag{1.2}$$

Now that we have recalled how to compute mean and variance of individual securities, by keeping in mind that a portfolio can be defined as a weighted average of individual securities, we can compute the mean and the variance also for a portfolio. By letting R_p be the random rate of return of the portfolio, r_i be the expected rate of return of the i^{th} asset which has variance equal to σ_i^2, with $i = 1, 2, ..., n$ and w_i be the percentage of wealth invested in the same i^{th} asset, we have:

$$E(R_p) = \sum_{i=1}^{n} r_i w_i = r_P \tag{1.3}$$

$$Var(R_p) = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} = \sum_{i=1}^{n} w_i \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{i,j} = \sigma_P^2 \tag{1.4}$$

where $\sigma_{i,j}$ is the covariance between the i^{th} and the j^{th} asset, and where $\rho_{i,j}$ is the Pearson correlation coefficient.

We can also briefly recall the concepts of covariance and correlation that we have just mentioned. The covariance between two random variables can be considered a statistical measure of the linear dependence of the two random variables, with the sign of the covariance defining its direction; on the one hand, positive covariance refers to variables which tend to move in the same direction while, on the other hand, we witness negative covariance in cases in which they tend to move in opposite directions. If we have two securities, i and j, we can use the covariance to analyze the case in which a price movement is associated with that of the other security. Covariance is calculated as follows:

$$\sigma_{ij} = E\left\{[r_i - E(r_i)]\,[r_i - E(r_j)]\right\} \tag{1.5}$$

We can also recall that the covariance of some variable with itself equals the variance of that variable, namely:

$$\sigma_{ii} = E[r_i - E(r_i)]^2 = \sigma_i^2$$

The correlation coefficient is, instead, a statistical measure of the association between two random variables and it is derived from the covariance. It can be shown that the correlation coefficient is obtained by dividing the covariance by the product of the two variables' standard deviation:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \tag{1.6}$$

Within the context of portfolio analysis, the knowledge of the correlation coefficients between the available securities can be exploited in order to perform diversification.

1.1.2 Efficient versus inefficient portfolios

With reference to the second step, Markowitz distinguishes between efficient and inefficient portfolios by defining efficient portfolios as those from which it is impossible to obtain a greater average return without incurring greater standard deviation; or similarly, those portfolios for which it is impossible to obtain smaller standard deviation without giving up return. In line with this description, a portfolio is, instead, inefficient if it is possible to obtain higher expected (or average) return with no greater variability of return, or obtain greater certainty of return with no less average or expected return. Summing up, based on the mean-variance criterion an efficient portfolio is a portfolio characterized either by:

- the minimum expected risk at any given expected rate of return, or
- the maximum expected rate of return at any given level of expected risk.

The mean-variance criterion can also be stated by using a more formal mathematical notation. Let X and Y be two variables with mean equal to μ_X and μ_Y respectively, and variance equal to σ_X^2 and σ_Y^2. We say that X dominates Y in mean-variance sense if and only if the following conditions hold together:

- $\mu_X \geq \mu_Y$,
- $\sigma_X^2 \leq \sigma_Y^2$, and

and at least one of these inequalities holds in strict sense.

In his Portfolio Selection (1952), Markowitz does not derive the efficient set of portfolios, from now on called efficient frontier, from a mathematical stand point, but rather on a graphical ground, and he proceeds as follows.

Firstly, he considers the case of three securities. In this case the model reduces to:

$$E = \sum_{i=1}^{3} X_i \mu_i \tag{1.7}$$

$$V = \sum_{i=1}^{3} \sum_{j=1}^{3} X_i X_j \sigma_{ij} \tag{1.8}$$

$$\sum_{i=1}^{3} X_i = 1 \tag{1.9}$$

$$X_i \geq 0 \text{ for } i = 1, 2, 3. \tag{1.10}$$

From Equation 1.9 we get:

$$X_3 = 1 - X_1 - X_2 \tag{1.11}$$

and if we substitute it in Equations 1.7 and 1.8 we get E and V as two functions of X_1 and X_2 only, so that we can work with two dimensional geometry.

The exact formulae are not too important here (but are given below for seek of completeness)[2] so that we can simply write:

$$E = E(X_1, X_2), \tag{1.12}$$

$$V = V(X_1, X_2), \tag{1.13}$$

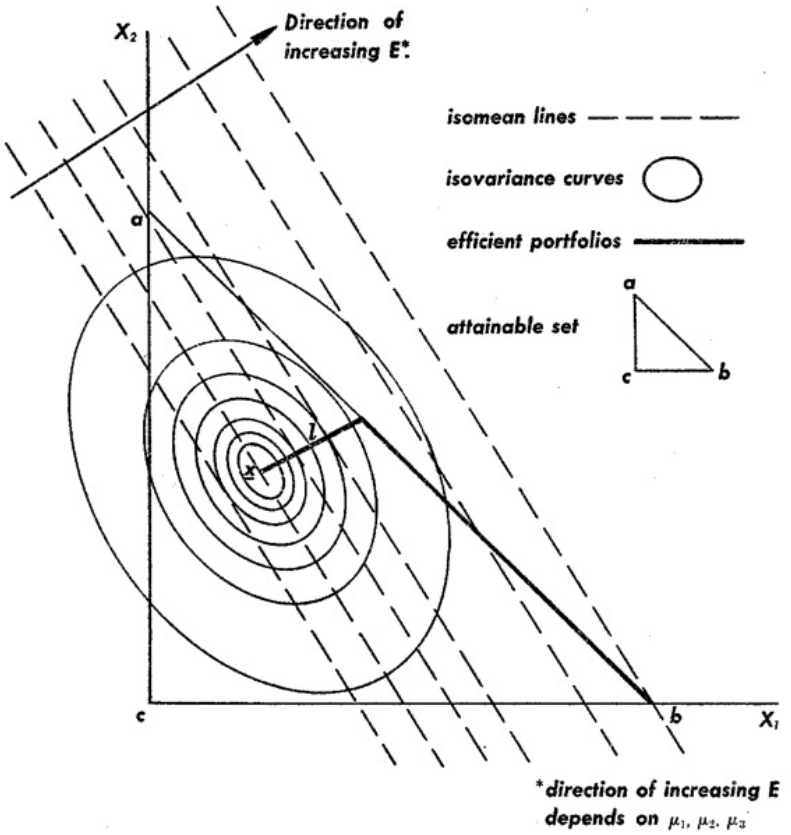
$$X_1 \geq 0, X_2 \geq 0, 1 - X_1 - X_2 \geq 0. \tag{1.14}$$

The attainable set of portfolios consists of all portfolios which satisfy constraints in Equation 1.14 and Equation 1.11, or equivalently constraints in Equations 1.9 and 1.10. The set of possible portfolios is represented in Figure 1.1. Any point to the left of the X_2 axis is not attainable because

[2] $E = \mu_3 + X_1(\mu_1 - \mu_3) + X_2(\mu_2 - \mu_3)$,
$V = X_i^2(\sigma_{11} - 2\sigma_{13} + \sigma_{33}) + X_2^2(\sigma_{22} - 2\sigma_{23} + \sigma_{33}) + 2X_1X_2(\sigma_{12} - \sigma_{13} - \sigma_{23} + \sigma_{33}) + 2X_1(\sigma_{13} - \sigma_{33}) + 2X_2(\sigma_{23} - \sigma_{33}) + \sigma_{33}$.

it violates the condition that $X_1 \geq 0$. Any point below the X_1 axis is not attainable because it violates the condition that $X_2 \geq 0$. Any point above the line $(1 - X_1 - X_2 = 0)$ is not attainable because it violates the condition that $X_3 = 1 - X_1 - X_2 \geq 0$.

Figure 1.1: Set of attainable efficient portfolios with X inside.



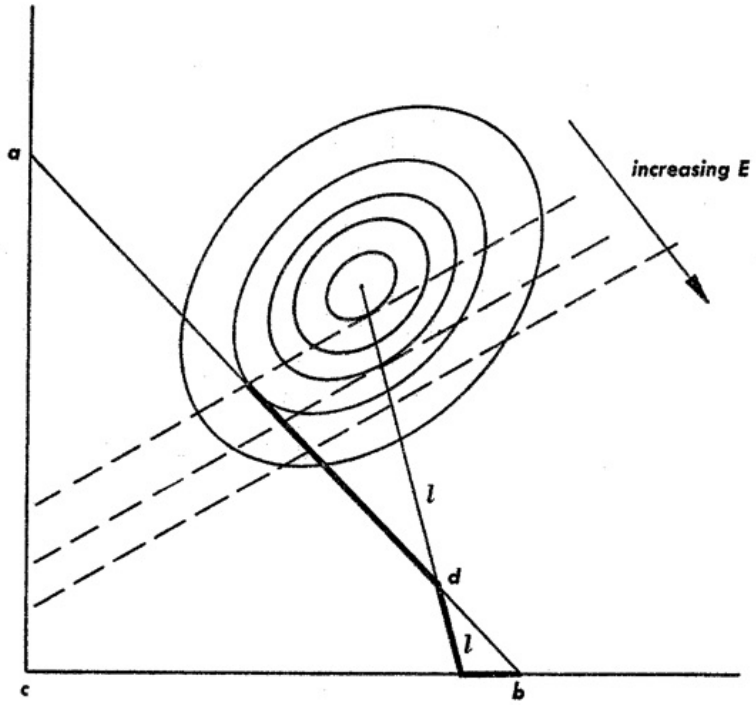
Source: Markowitz, H.: Portfolio Selection, The Journal of Finance 7 (1952) p. 84

Markowitz defines an isomean curve as the set of all portfolios having the same expected return. Similarly, he defines an isovariance line as the set of all portfolios with a given variance of return. By examining the formulae for E and V we can immediately get an insight on the shape of these lines. Specifically, they tell us that the isomean curves are a system of parallel straight lines, while the isovariance curves are a system of concentric ellipses. The center of this isovariance curves, which we will label X, is the point of the system minimizing the variance, V. In other words, if one isovariance curve lies closer to X than another, this means that the former curve is associated with a smaller variance than the latter. It is important to underline that as you move away from the center X, variance will increase. This center of the system of isovariance curves may fall either inside or outside the attainable set. The case in which X falls inside the attainable set, it is efficient because there is no other portfolio characterized by a V as low as X. No portfolio can have either smaller V with the same or greater E or greater E with the same or smaller V.

Now, consider all points with a given E, namely all points on the same isomean line associated with a given expected return. The point on the isomean line at which V takes on its least value is the point at which the isomean line is tangent to an isovariance curve. We will call this point $\hat{X}(E)$. If we let E vary, $\hat{X}(E)$ traces out a curve, which is a straight line in our three securities

case, and it is labeled in Figure 1.1 as l, the critical line. This line passes through X for this point minimizes V for all points with $E(X_1, X_2) = E$. The segment of the critical line from X to the point where the critical line crosses the boundary of the attainable set is part of the efficient set. The rest of the efficient set is the segment of the $\bar{ab}$ line from d to b. b is the point of maximum attainable E.

Figure 1.2: Set of attainable efficient portfolios with X outside.



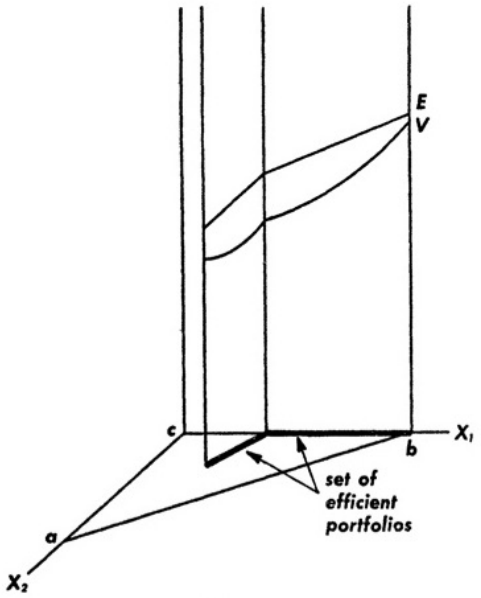
Source: Markowitz, H.: Portfolio Selection, The Journal of Finance 7 (1952) p. 85

In Figure 1.2, X lies outside the admissible area but the critical line cuts the admissible area. The efficient line begins at the attainable point with minimum variance, which in this case is the $\bar{ab}$ line. It moves toward b until it intersects the critical line, moves along the critical line until it intersects a boundary and finally moves along the boundary to b.

If we consider the four securities case, as well as the general case in which we have $N \geq 2$ securities, the efficient set will be a series of connected line segments. At one end of the efficient set there will be the point of minimum variance and, at the other end, the point of maximum expected return.

After having seen the nature of the set of the efficient portfolios, it is not difficult to see the nature of the set of efficient (E, V) combinations. In the three securities case E is a plane and V is a paraboloid. As shown in Figure 1.3, the section of the E-plane over the efficient portfolio set is a series of connected line segments. The section of the V-paraboloid over the efficient portfolio is a series of connected parabola segments.

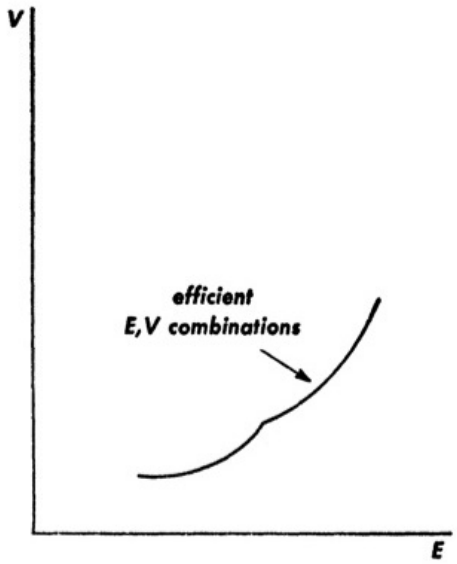
Figure 1.3: Set of efficient portfolios.



Source: Markowitz, H.: Portfolio Selection, The Journal of Finance 7 (1952) p. 86

If we plot V against E for efficient portfolios we will again get a series of connected parabola segments, as shown in Figure 1.4. The results just mentioned will be the same for any number of securities considered.

Figure 1.4: Set of efficient portfolios in the $E - V$ space.



Source: Markowitz, H.: Portfolio Selection, The Journal of Finance 7 (1952) p. 88

1.1.3 Portfolio selection

The third and last step of the optimal portfolio selection problem by Markowitz involves the selection of the proper portfolio in accordance with the investor's preferences regarding risk and expected return.

In order to do this, we still have two steps two be accomplished.

1.1.3.1 Analytical derivation of the efficient frontier

Markowitz derives the mean-variance efficient portfolio frontier mainly from a geometrical stand point in his original paper of 1952, and dedicates less effort to the analytical derivation of the efficient frontier and the efficient portfolio. It was only later, in its publication of 1959 and in the work of Merton (1972) that we can find its explicit and extensive mathematical derivation. They were also able to consider different portfolio compositions, namely portfolios made up of different number of securities, riskless and risky assets in different combinations. In his paper Merton (1972) also shows that under certain conditions, the classic graphical technique for deriving the efficient portfolio frontier is incorrect. In order to proceed to the analytical derivation of the efficient frontier, Merton (1990) formalizes a set of assumptions which are needed in order to solve the general problem of choosing this best investment mix. The following assumptions are made.

1. Frictionless market assumption. There are no transactions costs or taxes, and all securities are perfectly divisible.

2. Price-taker assumption. The investor believes that his actions cannot affect the probability distribution of returns on the available securities. Hence, if w_i is the fraction of the investor's initial wealth W_0 allocated to security i, then $\{w_1, ..., w_n\}$ uniquely determines the probability distribution of his terminal wealth. A riskless security is defined to be a security, or feasible portfolio of securities, whose return over the period is known with certainty.

3. No-arbitrage opportunities assumption. All riskless securities must have the same return. This common return will be denoted by R.

4. No-institutional restrictions assumption. Short-sales of all securities, with full use of proceeds, are allowed without restriction. If there exists a riskless security, then the borrowing rate equals the lending rate.

Merton then starts from the analysis of particular portfolios, in order to come up with the formulation of the mean-variance optimization problem in the general case of N assets, which we will now analyze.

Our formulation starts by assuming that we have N available assets corresponding to the set $X_i = X_1, X_2, ..., X_n$ with (random) rates of return $R_i = R_1, R_2, ..., R_n$. These assets are characterized respectively by expected rate of return $r_i = r_1, r_2, ..., r_n$ and covariances $\sigma_{i,j} = \rho_{i,j}\sigma_i\sigma_j$ for $i, j = 1, 2, ..., n$. As we have stated throughout this paper, a portfolio is defined as a set of weights, $w_i = w_1, w_2, ..., w_n$ which have to sum up to 1, $\sum_{i=1}^{n} w_i = 1$.

The frontier of all feasible portfolios which can be constructed from these N securities is defined as the locus of all feasible portfolios that have the smallest variance for a prescribed expected return.

We can do this by fixing a desired target return for our portfolio, so that $E(R_p) = r_p = \pi$, which will of course depend on the investor's risk aversion. From this condition we start the solution of the portfolio selection problem. As a matter of fact, the frontier can be described as the set of portfolios satisfying the following constrained minimization problem:

$$\begin{aligned} & min_w \sigma_P^2 \\ \text{subject to } & \begin{cases} r_p = \pi \\ \sum_i w_i = 1. \end{cases} \end{aligned} \tag{1.15}$$

Here, it is important to underline the fact that we do not have the $w_i > 0$ for $\forall i$ constraint as we had in the previous formulation of the maximization problem presented by Markowitz (1952). This means that in Merton (1972) we do not have restrictions on short selling, which is thus allowed.

We can also write the above-mentioned maximization problem in matrix form as follows:

$$\begin{aligned} & min_w w' Vw \\ \text{subject to } & \begin{cases} w'r = \pi \\ w'e = 1, \end{cases} \end{aligned} \tag{1.16}$$

where w is the vector of the portfolio weights, V is the variance-covariance matrix and e is a column vector of ones.

Since

> "technically, we minimize a convex function subject to linear constraints. Observe that $w'Vw$ is convex because V is positive definite and also note that the two linear constraints define a convex set. Therefore, the problem has a unique solution and we only need to obtain the first–order conditions"[3]

together with the fact that if all the securities are assumed to be risky and if there is no security that can be represented as a linear combination of the other securities, providing us a non-singular variance-covariance matrix of returns, $\Omega = [\sigma_{i,j}]$ we can proceed by solving this quadratic linear problem by the use of the following theorem.

Theorem. Let V be an $N \times N$–matrix of variances and covariances and let r be an N–vector of means. If V is nonsingular and positive definite and if $r_i \neq r_j$ for some$i, j = 1, ..., N$ then the (basic) portfolio selection problem has the unique following solution:

$$w^* = \frac{\left(\gamma V^{-1}\mathrm{r} - \beta V^{-1}\mathrm{e}\right)\pi + \left(\alpha V^{-1} e - \beta V^{-1} r\right)}{\alpha\gamma - \beta^2} \tag{1.17}$$

where

$\alpha = r' V^{-1} r,$

$\beta = r' Ve = e' V^{-1} r,$

$\gamma = e' V^{-1} e.$

[3]Jarrow, Maksimovic and Ziemba (1995) Finance, page 4

We are now able to compute efficient portfolios for each value of the target return π and to draw the efficient frontier. Merton then proves that we can rearrange Equation (1.16) and express the efficient frontier in terms of variance, thus getting the following:

$$wVw' = \sigma_{\mathrm{P}}^2 = \frac{\gamma\pi^2 - 2\beta\pi + \alpha}{\alpha\gamma - \beta^2} \tag{1.18}$$

From this formula, we can immediately see that our efficient frontier will be characterized by a bullet, parabolic shape when we consider the variance of returns, or a hyperbolic shape in case we consider the standard deviation of return as our measure of risk.

In order to solve the quadratic linear problem in Equation 1.17 we can differentiate the following Lagrangian with respect to each variable and to each Lagrangian multiplier, and by setting each equation result equal to zero:

$$\mathrm{L} = w'\,Vw - \lambda_1\,(w'r - \pi) - \lambda_2(w'e - 1). \tag{1.19}$$

from which we obtain the following first order conditions:

$$\begin{cases} \frac{\partial \mathrm{L}}{\partial \mathrm{w}} = 2w'\,V - \lambda_1\,r' - \lambda_2\,e' = 0_N \\ \frac{\partial \mathrm{L}}{\partial \lambda_1} = -w'r + \pi = 0 \\ \frac{\partial \mathrm{L}}{\partial \lambda_2} = -w'e + 1 = 0 \end{cases} . \tag{1.20}$$

This leads to the following system of equations:

$$w' = \frac{1}{2}\lambda_1\,r'\,V^{-1} + \frac{1}{2}\lambda_2\,e'\,V^{-1}\,w'r = \pi\,w'e = 1. \tag{1.21}$$

After we have substituted w' with its formula in each equation of the system and rearranged the results, we obtain the following system of equations

$$\begin{cases} w' = \frac{1}{2}\lambda_1\,r'\,V^{-1} + \frac{1}{2}\lambda_2\,e'\,V^{-1} \\ \frac{1}{2}\lambda_1 = \frac{\pi\gamma - \beta}{\alpha\gamma - \beta^2} \\ \frac{1}{2}\lambda_2 = \frac{\alpha - \pi\beta}{\alpha\gamma - \beta^2} \end{cases} . \tag{1.22}$$

where α, β, and γ take on the values listed above. By simple substitution, we will come up with the Equation 1.17.

1.1.3.2 Expected utility and the efficient portfolio

Once we have derived analytically our efficient frontier, we can obtain, through maximization of the expected utility, our efficient portfolio.

According to the expected utility framework, we know that each individual investor should act as if:

- he attaches numbers, called their utilities, to each possible outcome, and

- when faced with chance alternatives, he selects the one with the greatest expected value of utility.

Utility is a measure of derived happiness. Investment activity affects utility through its effect on wealth. Because the rate of return is a measure of the rate at which wealth is accumulated, the utility from investment activity can be restated as a function of the rates of return from a person's wealth. In a world of certainty where all outcomes are known in advance, a utility maximizer would simply invest his or her wealth in the one asset with the highest expected rate of return. However, in our uncertain world, investors can only maximize what they expect utility to be, not what it will actually turn out to be, because the future is unknown, and by introducing uncertainty, also considerations of risk enter the picture.

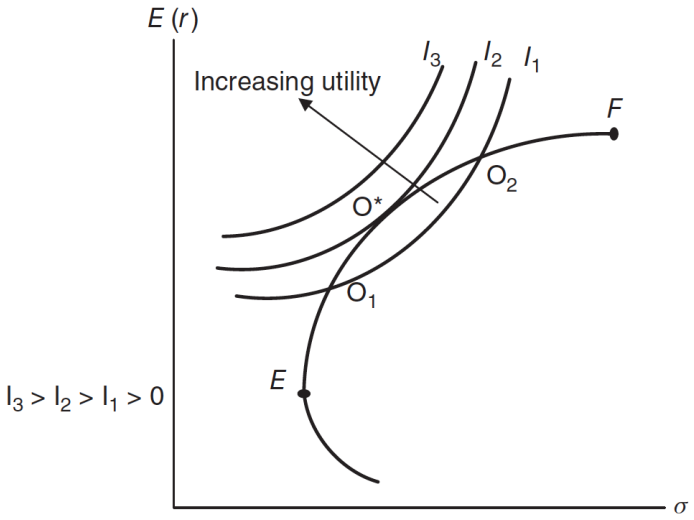
In this framework, we call indifference curves those curves plotted in the risk-return space, that are used to represent investors' preferred trade-off between risk, σ, and return, $E(r)$. Indifference curves are drawn so that an investor's satisfaction is the same throughout their length, and they are also called utility isoquants. If we assume that investors dislike risk, meaning that they get disutility from it, while they like positive expected returns, meaning that they receive utility from them, the indifference curves will be characterized by a positive slope in the mean-variance plane. The slope will then depend on the investor's particular preferences in terms of risk-return trade-off. These indifference curves grow more vertical as they rise, reflecting a diminishing willingness to assume risk. Indifference curves for a risk-averse investor are convex away from the $E(r)$ axis, while the efficient frontier is concave toward the $E(R)$ axis, due to the covariance effect. In Figure 1.5 we have the efficient frontier, bounded by the points E, which is typically referred to as the global minimum variance (GMV) point because no other portfolio exists that has lower variance, and F, which is typically referred to as the maximum return portfolio because no other portfolio exists that has a higher level of expected return. It is of particular importance, however, pointing out that F does only exist in the Markowitz formulation of the efficient frontier problem, where we have the constraint according to which no short-selling is allowed, namely where $w_i \geq 0, \forall i$. F does not exist in Merton formulation of the MPT, where we do not have such a constraint.

Establishing the convexity of a risk-averse investor's indifference curves and the concavity of the efficient frontier allows us to state the efficient frontier theorem, according to which the optimal portfolio for a risk-averse investor will be located on the efficient frontier. This theorem can be proven intuitively by noting that the risk-averse investor's convex indifference curves represent increasing expected utility as one moves from one utility isoquant to another in northwesterly direction. Because:

- the investor seeks a portfolio on the most northwestern indifference curve possible,
- the opportunity set is concave on its northwest boundary, and
- the northwest boundary of the opportunity set is by definition the efficient frontier, the investor's optimal portfolio will be an efficient portfolio.

Figure 1.5 shows the identification of a risk-averse investor's optimal portfolio O^*, which is a tangency point between the efficient frontier and the highest feasible indifference curve.

Figure 1.5: Identifying the optimal portfolio.



Source: Francis, J.C. and Kim, D.: Modern Portfolio Theory. Foundation, Analysis, and New Developments, Wiley Finance (20123) p. 126

In the end, the choice of the optimal portfolio depends on an investor's degree of risk aversion. If the investor is highly risk averse, the slope of the indifference curve will be steeper. In this case, the investor's optimal portfolio will be located on the lower part of the efficient frontier. Highly risk-averse investors generally prefer portfolios with low risk, and consequently low return.

The investor's utility function which is coherent with the mean-variance framework is a quadratic one. This can be proved in the following way.

If we assume to have a generic quadratic utility function $U(W) = W - \frac{b}{2}W^2$ with $b > 0$ being a strictly positive coefficient that reflects the risk aversion of the investor, and with W being the random wealth stemming from a certain portfolio, we will have that:

$$\begin{aligned} E[U(W)] &= E\left[W - \frac{b}{2}W^2\right] \\ &= E[W] - \frac{b}{2}E\left[W^2\right] \\ &= E[W] - \frac{b}{2}(E[W])^2 - \frac{b}{2}Var(W). \end{aligned}$$

It is now easy to see that this expression only depends on the first two moments of W. It is know important to underline the fact that we should use this kind of utility function only for $w < \frac{1}{b}$, because otherwise we will not get an increasing utility function in w.

For a generic utility function $F(W)$ that is infinitely derivable in w, mean and variance are not sufficient to pursue the utility maximization, because its expected value may depend also upon higher moments of the returns distribution. We can prove it by using Taylor expansions series, which allows for the approximation of a function as an infinite sum of terms, computed using the values of the function derivatives at a given point w.

$$
\begin{aligned}
E\left[F\left(W\right)\right] &= E\left[\sum_{i=0}^{\infty}\frac{F^{(i)}(w)}{i!}\left(W-w\right)^{i}\right] \\
&= \sum_{i=0}^{\infty}E\left[\frac{F^{(i)}(w)}{i!}\left(W-w\right)^{i}\right] \\
&= \sum_{i=0}^{\infty}\left[\frac{F^{(i)}(w)}{i!}E\left(W-w\right)^{i}\right].
\end{aligned}
$$

Also in Levy and Markowitz (1979) it is established that a well selected point from the mean-variance efficient set is "expected to yield almost maximum expected utility if the investor's utility function is approximately quadratic, or if his a priori beliefs are approximately normal"[4]. However, as it is noted in Szegö (2005) applying a mean-variance approach when the distribution of the returns differ significantly from an elliptical one, may lead to "sever underestimate extreme events which cause the most severe the losses"[5].

Now that we have established that from the many types of utility functions available, only the quadratic one works in the mean-variance framework, we can state that our utility function should be of the following form:

$$U(w,\mu,V,\gamma) = w'\mu - \frac{\gamma}{2}w'Vw. \tag{1.23}$$

Since we have established above that the optimal portfolio will thus be O^*, the tangency point between the efficient frontier and the highest feasible indifference curve, we will have to derive this tangency point. After having established that the investor's utility function will be of the form in Equation 1.23

What we actually do, is solving the following maximization problem:

$$max_w E(U) = max_w w'\mu - \frac{\gamma}{2}w'Vw. \tag{1.24}$$

$$\text{subject to } w'e = 1.$$

To proceed we just have to remember that maximizing a function is equivalent to minimizing its negative. We differentiate the below Lagrangian:

$$L(w,\lambda) = -r'w + \gamma w'Vw - \lambda(w'e - 1) \tag{1.25}$$

We derive its solution by applying the same procedure we used in Equation 1.20 and in Equation 1.21 for the classical mean-variance optimization problem, thus obtaining:

$$w^u = \frac{1}{\gamma}V^{-1}\left(r - \frac{e'V^{-1}r - \gamma}{e'V^{-1}e}e\right). \tag{1.26}$$

[4]Levy, H. and Markowitz, H.M.: Approximating expected utility by a function of mean and variance, The American Economic Review 69, 308-317 (1979).

[5]Szegö, G.: Measures of risk, European Journal of Operational Research, 163 , 5–19 (2005).

There is, however, one drawback in using a direct utility maximization procedure. This problem lies in the choice of the risk-aversion parameter γ. The size of this quantity is, in fact, still an empirical issue, and its estimates seem to vary with respect to the specific environment considered.

1.2 The effect of Markowitz diversification

Markowitz in his work

> "had the brilliant insight that, while diversification would reduce risk, it would not generally eliminate it (...). Probably the most important aspect of Markowitz's work was to show that it is not a security's own risk that is important to an investor, but rather the contribution the security makes to the variance of the entire portfolio, and that this was primarily a question of its covariance with all the other securities in his portfolio." [6]

As a matter of fact, Markowitz introduced a new kind of diversification, different from the diversification which had been used since then in financial economics, which is named naïve diversification. Naïve diversification ignores the covariance between securities and simply follows the maxim according to which an investor "must not put all its eggs into the same basket", resulting in superfluous diversification. With the naïve diversification rule we refer to a strategy according to which one invests equally across N assets of interest by splitting the wealth available for investment equally upon them, without the aid of any theory nor any data.

Markowitz diversification, instead, involves something more, namely the combination of assets with less-than-perfect positive correlations in order to reduce risk in the portfolio as a whole without sacrificing any of the portfolio's return. In general, the lower the correlation between the assets in a portfolio, the less risky the portfolio. This is true regardless of how risky the assets of the portfolio are when analyzed on an individual basis. Markowitz explains this approach to diversification as follows:

> "Not only does [portfolio analysis] imply diversification, it implies the "right kind" of diversification for the "right reason." The adequacy of diversification is not thought by investors to depend on the number of different securities held. A portfolio with sixty different railway securities, for example, would not be as well diversified as the same size portfolio with some railroad, some public utility, mining, various sorts of manufacturing, etc. The reason is that it is generally more likely for firms within the same industry to do poorly at the same time than for firms in dissimilar industries. Similarly, in trying to make variance [of returns] small it is not enough to invest in many securities. It is necessary to avoid investing in securities with high covariances [or correlations] among themselves."[7]

[6] Rubinstein, M.: Markowitz's "Portfolio selection": a fifty-year retrospective, The Journal of Finance 57, 1041-1045 (2002).

[7] Markowitz, H. : Portfolio selection: efficient diversification of investments, 2. ed Oxford : Basil Blackwell (1991).

1.3 Modern portfolio theory's fallacies

Whilst Markowitz optimal portfolio selection has been widely accepted in financial and economic theory, it has also been criticized, especially when dealing with its practical application. From the very first assumptions of the model, we can see that they are way too unrealistic.

The fact that it assumes a market without frictions, without institutional restrictions, where assets which are infinite divisible are sold, does not seem particularly realistic.

If we analyze more in depth the model we come up with other inconsistencies, such as the fact that it assumes that returns are normally distributed. As a matter of fact, usually, financial instruments have returns which presents skewness and fat tails in their distributions. However, we should now mention the fact that Fama (1976) showed that the normal assumption is appropriate for describing returns of monthly data. Moreover, according to Tu and Zhou (2004) the normal approach works well for a mean-variance investor, showing that the impact of returns' fat tails is not large. The other main assumption on which the theory is based on, the one of independent asset returns, seems unrealistic too, and we will try to deal with it in the next chapters.

The use of the variance as a measure of risk has been criticized as well, since it is a symmetric risk measure that puts equal weight below and above the mean when, instead, several empirical studies such as Knetsch and Borges (1998) have shown that investors treat differently positive and negative returns.

Critics have been made also to Markowitz diversification, because it is true that adding assets to a portfolio reduces its risk, but this effect tends to decrease as additional assets are included complicating even more the portfolio selection problem, due to the size of the variance-covariance matrix. Moreover, as stated in Statman and Scheid (2008) correlation is not a good measure for two reasons. The first is that we tend to misperceive correlations. Relatively high correlations bring relatively low diversification benefits. The second is that the benefits of diversification depend not only upon correlation between investment returns but also n the standard deviations of each investment's returns.

Important is also the problem of Markowitz's theory that does not assume that investors should use historical data for computing expected return. Markowitz (1952) states that investors' expectations are based on "relevant beliefs about future performances" . The problem is that it does not suggest how to build these informed beliefs. As a matter of fact Markowitz says:

> "Portfolio selection based solely on past performances assumes, in effect, that past averages and standard deviations are reasonable measures of the uncertainty of return in the future. (...) Rather than using past performances per se, we could use the "probability beliefs" of experts as inputs to a portfolio analysis."

This estimation error problem, does not only affect the out-of-sample performance of a mean-variance strategy giving solutions to the optimization problem which are not really optimal, but also cause problems to portfolio weights which become very unstable. As a matter of fact, instability of mean-variance solutions is another issue of modern portfolio theory. This is due to the variance of the estimation error itself where a greater variance of this error term leads to heavier changes in the weights of optimal portfolios, but it is also due to the inherent structure of the mean-variance

approach related to the asset structure of the efficient frontier. As noted by Michaud (1989) even when taking a neighborhood of the optimal portfolio lying on the true efficient frontier, we can find many statistically equivalent portfolios, with really close expected return and standard deviation, which may have a completely different asset weights allocation. In subsequent periods then, even for very small changes in the inputs and in absence of estimation error, the usual mean-variance model may suggest us to heavily revise our portfolio, causing large transaction costs and making our investment strategy less appealing both in terms of net average return and in terms of transaction costs-adjusted Sharpe ratio.

Chapter 2

Inference procedure under Modern Portfolio Theory

2.1 Estimation for Markowitz efficient portfolios

Even if the mean-variance analysis presented so far seems theoretically simple, the investor is faced with a number of difficulties when trying to implement it practically. As a matter of fact, the first two moments of the asset returns distribution which, as we already pointed out, completely specify the set of optimal portfolios in the mean-variance space, are unknown quantities that have to be estimated using historical data.

This problem, known in the financial literature as the uncertainty about the parameters of the asset returns, has been deeply analyzed.

Dickinson (1974) finds that the estimates of the weights and variance of the global mean-variance portfolio are highly unreliable.

Michaud (1989) discusses various problems of using the Markowitz method of selecting optimal portfolio.

Broadie (1993) demonstrates that sample frontiers are overly optimistic while the out-of-sample frontiers are inferior to both the true and sample frontiers.

Smith and Kan (2008) derive the finite sample distribution and moments of the sample minimum-variance frontier when returns are independent multivariate normal random variables, and find that the sample mean-variance frontier is a heavily biased estimator of the population frontier, even when the length of the estimation window is very long. In order to correct for this bias, they also present a new adjusted estimator of the population frontier that has a significantly smaller bias than the traditional sample estimator.

Important results in this field have been obtained in Bodnar and Schmid (2008) where the exact distributions of the estimated parameters of the efficient frontier are obtained. In particular, Bodnar and Schmid (2009) derived exact tests for the parameters of the efficient frontier and constructed the confidence region for optimal portfolios in the mean-variance space. But, since in this last paper the authors ignore the uncertainty about the estimated parameters of the efficient

frontier in the first stage, Bodnar and Bodnar (2009) derive statistical inference procedures for the expected return, μ, the variance, σ^2, and the slope coefficient, s, of the efficient frontiers. They also suggest finite sample tests for the location of the efficient frontier with the estimated parameters in $E-V$ space and they derive the exact densities of the test statistics. While Bodnar and Bodnar (2010) derive an unbiased estimator of the efficient frontier, which corrects the overoptimism of the sample efficient frontier documented in Siegel and Woodgate (2007); they also present an exact F-test on the efficient frontier.

> "Optimal portfolio selection requires knowledge of each security's expected return, variance, and covariance with other security returns. In practice, each security's expected return, variance, and covariance with other security returns are unknown and must be estimated from available historical or subjective information. When portfolio optimization is implemented using the historical characteristics of security returns, estimation error can degrade the desirable properties of the investment portfolio that is selected." [1]

Many are then the authors that studied the uncertainty about the parameters of the asset returns. Since the papers dedicated to this matters are many, it is natural to have many different approaches in which this uncertainty problem has been investigated. In the following sections, thus, we would like to have a look at the more recent and relevant ones, even if our presentation makes no claim to be exhaustive.

The ultimate goal of this work is that of investigating the impact of a change in the assumptions of the unbiased efficient estimator of the efficient frontier proposed by Bodnar and Bodnar (2010) in order to investigate if it is possible to extend their results to a larger family of asset returns than the one considered so far in the literature, namely for a family of asset returns are not independent. Since we have decide to base our further analysis on the efficient frontier estimation approach of Bodnar and Bodnar we will present their results, together with the results of other authors which deal with this same argument but approaching it from a different standpoint. We will namely presents the results of Tu and Zhou (2004), Basak, Jagannathan and Ma (2005), Siegel and Woodgate (2007), Kan and Smith (2008), and at last the results of Bodnar and Bodnar (2010).

2.2 Tu and Zhou Bayesian approach to data generating process uncertainty

In order to derive their estimated mean and variance, Tu and Zhou (2004) follow the framework of Pastor and Stambaugh (2000). They start by considering an investment universe which contains cash plus n spread positions. The spread position i, constructed at the end of the period $t-1$, is a purchase of one asset coupled with an offsetting position in which we have the short sale of another asset. These two assets are denoted respectively L_i and S_i, and their rates of return in period t are denoted as $R_{L_i t}$ and $R_{S_i t}$. Thus, a spread position of size X_i has a dollar payoff $X_i\left(R_{L_i,t} - R_{S_i,t}\right)$. Since regulation T requires the use of margins for risky investments, a constant $c > 0$ is used to characterize the degree of margin requirements. The spread position involves at least one risky

[1] Frost and Savarino (1986), "An Empirical Bayes Approach to Efficient Portfolio Selection", Journal of Financial and Quantitative Analysis, Vol.21, No.3.

asset which, without loss of generality, is defined as L_i. If the other asset of position i, S_i of size X_i, is risky as well, then $(2/c)\,|X_i|$ dollars of capital are required; otherwise $(1/c)\,|X_i|$ dollars of capital are required. The total capital required to establish the spread positions must be less than or equal to the investor's wealth, W_{t-1}

$$\sum_{i\in\Lambda}(2/c)\,|X_i| + \sum_{i\notin\Lambda}(1/c)\,|X_i| \leq W_{t-1} \tag{2.1}$$

where Λ denotes the set of positions in which S_i is risky, or alternatively

$$\sum_{i\in\Lambda} 2\,|w_i| + \sum_{i\notin\Lambda}|w_i| \leq c \tag{2.2}$$

where $w_i = X_i/W_{t-1}$.

They continue their analysis by letting r_t denote an n-vector such that its i^{th} element $r_{i,t}(= R_{L_i,t} - R_{S_i,t})$ represents the return of the i^{th} risky position at time t. If there is a riskless asset with a rate of return $R_{f,t}$ then the excess return of this portfolio is:

$$R_{p,t} - R_{f,t} = \sum_{i=1}^{n} w_i r_{i,t}. \tag{2.3}$$

The investor is assumed to choose w as to maximize the mean-variance objective function

$$U = E\,(R_{p,t}) - \frac{1}{2}AVar\,(R_{p,t}) \tag{2.4}$$

subject to the wealth constraint in Equation 2.2 where A is interpreted as the coefficient of relative risk aversion. They denote respectively the mean vector and covariance matrix of r_t as E and V, then the investor's optimal portfolio choice problem can be rewritten ad the solution to the following maximization problem:

$$\begin{array}{c} max_w\left(w'E - \frac{1}{2}Aw'Vw\right) \\ \text{subject to } \sum_{i\in\Lambda} 2\,|w_i| + \sum_{i\notin\Lambda}|w_i| \leq c \end{array}. \tag{2.5}$$

Assume that r_t follows a well-defined probability distribution

$$r_t \sim P(E, V) \tag{2.6}$$

where P is the DGP of the data.

In the classical framework, the investor solves the maximization problem in Equation 2.5 with E and V replaced by their estimators, in light of the assumed distribution P. The density function of P is usually available, and E and V are then estimated using maximum likelihood method. However, estimation errors are often ignored.

In the Bayesian[2] framework, on the other hand, both parameter E and V are viewed as random variables, and the parameter uncertainty is captured by their posterior distributions in light of

[2]For a recap on Bayesian inference, see Appendix A.

the data. Incorporating the parameter uncertainty amounts to integrating over the posterior distributions. Which is mathematically equivalent to replacing the mean and the covariance E and V in Equation 2.5 by the predictive mean vector and covariance matrix of r_t conditionally on the available data.

The question is now how to model the uncertainty implicit in the data generating process. This uncertainty stems from the fact that investors do not have clues on the distribution of the DGP. In order to address this problem, Tu and Zhou assume that investors know a set of probability distributions that are likely candidates for the true DGP. In Tu and Zhou, this set consists of 31 multivariate t distributions with degrees of freedom varying from $2.1, 3, 4, 5, ..., 31$, to 32. The smaller the degrees of freedom, the greater the deviation from the normal distribution. The DGP uncertainty is captured by v, the degrees of freedom of the t distribution. On a theoretical ground, a normal distribution has an infinite number of degrees of freedom, but it is well modeled by a t with $v = 32$. If investors are not sure about the degree of non-normality in the data, they may place an equal probability on all possible values of v.

They specify a set of 31 multivariate t distributions, with degrees of freedom contained in the set $S_v = 2.1, 3, 4, ...31, 32$, as likely candidates for the DGP. The returns thus will follow

$$r_t \sim DGP\,(E, V, v) \tag{2.7}$$

where DGP is a t distribution with unknown parameter v.

In this Bayesian framework, v is viewed as a random variable which takes values from the set S_v. The diffuse prior[3] on v can be written as

$$p_o\,(v) = \frac{1}{|S_v|'} \tag{2.8}$$

where $|S_v|$ is the number of elements in the set S_v.

Mathematically, the DGP can be any distribution, and v can then simply be an index parameter. However, as we restrict the candidate distributions to the t class, we can interpret the parameter v as an indicator of data's kurtosis. Since the kurtosis is a decreasing function of v, the smaller its value, the more leptokurtic the data.

For a more intuitive understanding of the priors, it is useful to cast the DGP into a regression setting. They let $r_t = (y_t, x_t)$, where y_t contains the excess returns of m non-benchmark positions and x_t contains the excess returns of $k\,(= n - m)$ benchmark positions. They consider the following multivariate regression

$$y_t = \alpha + Bx_t + u_t \tag{2.9}$$

where u_t is an $m \times 1$ vector with zero means and non-singular covariance matrix.

To relate α and B to the earlier parameters E and V, consider the corresponding partition:

[3] A diffuse prior, or noninformative prior, or vague prior, is one that expresses ignorance as to the value of θ. In general, a diffuse prior is dominated by the likelihood functions. In other words, such a prior:

- does not change much over the region in which the likelihood is appreciable, and
- does not assume large values outside that region.

$$E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix}, V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \tag{2.10}$$

Under the usual multivariate normal distribution, it is clear that the distribution of y_t conditional on x_t is also normal, and the conditional mean is linear function of x_t.

Hence,

$$E\left(y_t \mid x_t\right) = E_1 + V_{12}V_{22}^{-1}\left(x_t - E_2\right) \tag{2.11}$$

$$Var\left(y_t \mid x_t\right) = V_{11} - V_{12}V_{22}^{-1}V_{21}. \tag{2.12}$$

Therefore, the parameters α and B obey the following relationship with the earlier parameters E and V:

$$\alpha = E_1 - BE_2, B = V_{12}V_{22}^{-1}V_{21} \tag{2.13}$$

and

$$\Sigma = V_{11} - BV_{22}B'. \tag{2.14}$$

Here, the Σ matrix is the familiar notation for the covariance matrix of u_t under the usual normality assumption. However, in this case the DGP is in the class of the t distributions. While the mean of $y_t$conditional onx_t is still a linear function of x_t as above, its conditional covariance matrix is no longer independent of x_t

$$Var\left(y_t \mid x_t\right) = k\left(x_t, v\right)\left(V_{11} - V_{12}V_{22}^{-1}V_{21}\right) \tag{2.15}$$

where $k\left(x_t, v\right)$ is a function of both the quadratic form $\left(x_t - E_2\right)' V_{22}^{-1}\left(x_t - E_2\right)$ and the degrees of freedom v. Correspondingly, the covariance matrix of the residual term u_t is $k\left(x_t, v\right)\Sigma$. But since Σ no longer represents the covariance matrix of u_t, unless we have $v = \infty$. That is why they affirm that we must take special care in drawing the parameters from their posterior distributions even though the mapping of the parameters can be done in exactly the same way as the normal case.

We turn our attention to priors on α, B, and Σ by allowing for mispricing uncertainty. They do this by using the Pastor and Stambaugh's prior on α which follows the normal distribution:

$$\alpha \mid \Sigma \sim N\left(-BE_2, \sigma_\alpha^2 I_m\right). \tag{2.16}$$

They also derive the remaining priors, but their derivation would be beyond the scope of this work.

We, thus, only report the complete prior on all parameters, which can be written as follows:

$$p_o(\theta) = p_0(\alpha \mid \Sigma)\, p_0(\Sigma)\, p_0(B)\, p_0(E_2)\, p_0(V_{22})\, p_0(v). \tag{2.17}$$

It is important to underline that they assume independence between the DGP and the mispricing uncertainties. This may be true in general, since a given DGP may not allow the validity of an asset pricing model and hence can disallow a dogmatic belief about the model.

With the priors given above, the investor forms her posterior belief $p(0 \mid R)$ in light of the data $\{R : r_t, t = 1, ..., T\}$. The predictive distribution of r_{T+1} can then be computed as:

$$p(r_{T+1} \mid R) = \int_\theta p(r_{T+1} \mid R, \theta)\, p(\theta \mid R)\, \mathrm{d}\theta. \tag{2.18}$$

When the investor solves Equation 2.5 by integrating over the parameter uncertainty, it is easy to see that this is equivalent to replacing E and V with E^* and V^*, the mean and the covariance matrix of the predictive distribution. By the law of iterated expectations, it is clear that the predictive mean obeys the following relation:

$$E^* = E(r_{T+1} \mid R) = E(E(r_{T+1} \mid \theta, R) \mid R). \tag{2.19}$$

This says that the predictive mean is identical to the posterior mean. This means that the classical approach of plugging in the estimates for the mean should have little effect on the valuation of the quadratic utility function.

However, the predictive covariance matrix obeys the following relation:

$$V^* = Var(r_{T+1} \mid R) = E(Var(r_{T+1} \mid \theta, R) \mid R) + Var(E(r_{T+1} \mid \theta, R) \mid R). \tag{2.20}$$

This means that the predictive variance is a sum of two components, the posterior mean of the variance of a given model, which is the posterior mean of V under normality and a diffuse prior, and the variance of E due to its uncertainty. Since the classical estimation usually provides only an estimate for the first component, simply plugging-in this estimate is likely to over-value the quadratic utility function. Finally, as the DGP is of the class of t distributions, the implementation of both posterior and the predictive evaluations will be more complex than in the normal case.

2.3 Basak, Jagannathan, and Ma jackknife-type approach

In their work, Basak, Jagannathan, and Ma (2005) show that in-sample estimate of the variance of a global minimum risk portfolio constructed using an estimate covariance matrix of returns, will on average be strictly smaller than its true variance.

They proceed by illustrating other different estimation approaches widely used in the literature, how they are used, how they are proved, and mainly how they are deemed to be inferior when compared to their jackknife-type estimation[4], by using simulation.

The estimation approaches they compare throughout the paper are the following:

[4]For more details on the jackknife technique see Appendix B.

- in-sample estimate approach,
- degree-of-freedom adjusted in-sample estimate approach,
- Bayesian estimate under diffuse priors approach, and
- jackknife-type estimate approach.

We will give more space to the presentation of their original contribution, namely to the jackknife-type estimator.

2.3.1 In-sample estimate approach

As it has also been observed in the literature, they show that the in-sample value of the variance of an optimal portfolio constructed using historical return data is an optimistic estimate of the portfolio's true variance. They state that this bias, which is commonly referred to as in-sample optimism, increases with the number of assets used to construct mean-variance efficient portfolios. They continue by identifying the reason for this in-sample optimism and by suggesting method for correcting for it. However, since the focus of their work is in assessing the risk of an efficient portfolio, and not its mean, they focus on the global minimum variance and minimum benchmark tracking-error variance portfolios constructed using an estimated covariance matrix. They refer in the paper to this portfolios as SMRPs, i.e. sample minimum risk portfolios. They also show through simulation that this in-sample optimism can be substantial in certain situations.

2.3.2 Degree-of-freedom adjusted in-sample estimate approach

Since it may be argued that by scaling up the in-sample variance by a factor that is related to the degree-of-freedom of the distribution of the estimated covariance matrix, we may be able to correct for this in-sample optimism. But they promptly show that when returns are drawn from an independent and identically distributed multivariate normal distribution, scaling up the in-sample variance by a degrees-of-freedom related factor provides an unbiased estimate of the variance of the true global minimum variance portfolio. Since the out-of-sample variance of the sample minimum risk portfolio is strictly larger on average, this procedure does not adequately correct for the in-sample optimism. They show that it would be necessary to add twice the amount of the degrees-of-freedom correction to the in-sample risk to arrive at a good estimate of the out-of-sample risk. They finally underline that, despite being inadequate, degrees-of-freedom based correction are difficult to derive in general, and are also unavailable when efficient portfolios are constructed subject to portfolio weight constraints.

2.3.3 Bayesian estimate under diffuse priors approach

Basak, Jagannathan, and Ma continue their analysis by noting that it would appear to be convenient that in order to avoid the in-sample optimism, one could use the Bayesian approach, since it would take into account the randomness of the means and covariances used in optimization procedure. However, they argue that with the standard diffuse priors commonly used in empirical works, the variance of the efficient portfolio computed using the predictive distribution can also be

substantially below its variance. They conclude by showing that the variance of the global minimum variance portfolio computed using the predictive distribution with standard diffuse priors will be almost the same as the degree-of-freedom adjusted in-sample under the classical method.

2.3.4 Jackknife-type estimate approach

Since there is general consensus that variances of returns and covariances among returns vary in a systematic stochastic fashion over time, to account for this persistence, they opt to consider two approaches. The first is a variation of the jackknife-type estimator for the out-of-sample variance of SMRPs that weights recent observations more heavily than those in the distant past. The second approach they propose is, instead, based on the Dynamic Conditional Correlation Model of Engle (2002), which provides an accurate estimate of risks in the SMRPs constructed using one- and three-factor models. However, since the first approach provides a more precise estimate of the out-of-sample risk of the SMRPs we will only present this former model.

They firstly consider the special case where the portfolio manager estimates covariance matrices and computes optimal portfolio weights at the end of each period, and the manager's investment horizon is also one period. They assume that the portfolio manager has time series data for periods T in the immediate past, and estimates the return covariance matrix using these T observations. They let S_T denote this estimate and w_T denote the portfolio weights formed using S_T. They also let $w_T' R_{T+1}$ be the return on this portfolio during period $T+1$, and they consider the following estimator of $Var(w_T' R_{T+1})$.

They first drop the i'^{th} return observation, for an arbitrary i in the interval $[1, T]$, and estimate the return covariance matrix. They let $S_{(T-1,i)}$ be the covariance matrix estimate obtained by using the data $\{R_1, ..., R_{i-1}, ..., R_{i+1}, ..., R_T\}$. They then construct the global minimum variance portfolio using $S_{(T-1,i)}$. They denote the resulting portfolio as $w_{(T-1,i)}$. They note that the portfolio returns, namely $w_{(T-1,i)}' R_i$ and $w_T' R_{T+1}$, have approximately the same distribution for large enough T, if the time series of returns on the primitive assets are i.i.d. Hence, a natural estimator of the out-of-sample variance, $Var(w_t' R_{T+1})$, would be the sample variance of the T portfolio returns, $\left[w_{(T-1,1)}' R_1, w_{(T-1,2)}' R_2, ..., w_{(T-1,T)}' R_T\right]$. With daily data, they say that they can ignore the effect of sample mean return on the calculation of the sample variance, and hence the jackknife estimator is:

$$\hat{q}^{JK} = \frac{1}{T} \sum_{i=1}^{T} \left(w_{(T-1,i)}' R_i \right)^2 . \tag{2.21}$$

While the validity of the jackknife also requires T to be large, it need not be as large as what may be required for the standard central limit theorem. This is because we only need $w_{(T-1,i)}' R_i$ and $w_T' R_{T+1}$ to have approximately the same distribution.

If it is assumed that the returns have the exchangeability property[5], the i.i.d. assumption can be relaxed.

They then turn to the case where returns are measured more frequently, say once a day, but the portfolio manager revises the holdings only once a month, with there being l days in a month.

[5]For exchangeability property we mean that for any fixed positive integer k, the joint density of $\{R_1, ..., R_k\}$, say $f(r_1, ..., r_k)$, is the same as $f\left(r_{\sigma(1)}, ..., r_{\sigma(k)}\right)$ for all permutations $\sigma : \{1, ..., k\} \to \{1, ..., k\}$.

They let T be the number of days of daily return data used to estimate covariance matrix S_T, and the efficient portfolio weights w_T, with $m = T/l$ the number of months of observation. They let $q_{T,l}$ denote the sample variance of the l post-formation daily returns on the efficient portfolio, i.e. the sample variance of $(w'_T R_{T+1}, ..., w'_T R_{T+l})$. The objective of Basak, Jagannathan and Ma, will be that of estimating $q_{T,l}$.

They thus assume that monthly returns satisfy the exchangeability property. Then, for each month $1 \leq i \leq m$, we estimate the covariance matrix $S_{(T-l,i)}$ by deleting the return data for the l days in month i and construct the global minimum variance portfolio using $S_{(T-l,i)}$. They let the resulting portfolio be $w_{(T-l,i)}$. They compute the sample variance of the sequence of l returns on the efficient portfolio, $\left(w'_{T-l,i} R_{(i-1)l+1}, ..., w'_{T-l,i} R_{il}\right)$. They let $q_{T,l}(i)$ denote this sample variance.

They note that $S_{(T-l,i)}$ will have approximately the same distribution as S_T, for moderately large T, and so each $w_{(T-1,i)}$ will have approximately the same distribution as the sequence $(w' R_{T+1}, ..., w'_T R_{T+l})$. Since $q_{T,l}(i)$ will have approximately the same distribution as $q_{T,l}$, the average of the m such variances, $q_{T,l}(i), i = 1, 2, ..., m$ provides a consistent estimate of $q_{T,l}$. This thus gives the following jackknife estimator:

$$\widehat{q_{T,l}}^{JK} = \frac{1}{m} \sum_{i=1}^{m} q_{T.l}(i) \tag{2.22}$$

They then modify the above estimator to take into account the persistence in the covariance structure. They start by recalling that when the interval over which returns are measured and the holding period match, the jackknife estimator for i.i.d. data is given by Equation 2.21. They say that if returns do not satisfy the exchangeability condition, specifically, if returns have persistent covariance structure, it would make intuitive sense to use a weighted average instead of straight average in Equation 2.21, giving more weight to the recent observations. One such estimator is given by:

$$\hat{q}^{JK} = \frac{\sum_{i=1}^{T} e^{\alpha i} \left(w'_{(T-1,i)} R_i\right)^2}{\sum_{i=1}^{T} e^{\alpha i}} \tag{2.23}$$

for some positive α.

When the efficient portfolio weights are recomputed once every month using historical returns, the jackknife estimator under the i.i.d. assumption is given by Equation 2.22. When variance and covariances vary over time in a systematic stochastic fashion, we modify the average in Equation 2.22 into a weighted average, yielding the weighted jackknife estimate of the out-of-sample variance:

$$\widehat{q_{T,l}}^{JK} = \frac{\sum_{i=1}^{m} e^{\alpha i} q_{T,l(i)}}{\sum_{i=1}^{m} e^{\alpha i}} \tag{2.24}$$

for some positive α.

They evaluate the performance of the jackknife and of the other estimators by examining the mean and the standard deviation of the forecast errors as well as the mean absolute forecast error over a large number of out-o-sample holding periods.

In the end, through this simulations they confirm empirically that this last method, the jackknife-type method, gives more precise estimates of the sample minimum risk portfolio's risk than in-sample estimate, the degree-of-freedom adjusted in-sample estimate, and the Bayesian estimate under diffuse priors.

2.4 Siegel and Woodgate closed-form asymptotic approach to inference in MPT

Siegel and Woodgate (2007) provide the first theoretical guidance regarding the exact size of the impact of estimation error. They are able to quantify the impact of the estimation error without relying upon prior assumptions on the unknown parameters, and to give an adjustment for statistical noise that will asymptotically reflect the actual performance within the Markowitz mean-variance framework. They provide asymptotic non-Bayesian closed-form formulae for portfolio performance. It is important to underline that in their analysis, they make the following assumption: they suppose to have an investor who forms classical sample estimates of asset means, variances, and covariances and uses them as if they were the true assets' distribution parameters to form mean-variance efficient portfolios and to judge their anticipated performance. Siegel and Woodgate refer to these classical sample performance as "naïve" estimates because they do not take into account the estimation error stemming from using a sample rather than the population in determining these parameters. In particular, the naïve portfolio mean estimates tend to be biased upwards and the naïve variance estimates tend to be biased downwards, resulting in nominally efficient portfolios that are over-optimistic, in the sense that such an investor will believe she can achieve a higher mean and lower variance than it actually is available from the performance of his or her portfolio. They then proceed in quantifying this over-optimism bias in closed-form asymptotic formulae.

In order to derive their non-Bayesian closed-form formulas for portfolio performance they proceed as follows. They consider $n \geq 2$ assets where R_{it} denotes the observed rate of return on asset i at time t. Asset return vectors $R_t = (R_{1t}, ..., R_{nt})'$ are assumed to be independent and identically normally distributed with unknown true mean vector $\mu = E(R_t)$ estimated unbiasedly at time T as $\hat{\mu} = \sum_{t=1}^{T} R_t/T$, and unknown true covariance matrix $\mathrm{V} = Cov(R_t)$ estimated unbiasedly as $\hat{\mathrm{V}} = (1/(T-1)) \sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'$. They also assume the non-singularity of the variance-covariance matrix V and that the elements of μ are not all equal. They denote the estimation errors as $\delta = \hat{\mu} - \mu$ and $\varepsilon = \hat{\mathrm{V}} - \mathrm{V}$.

They then use these estimates, based on observations at times $t = 1, ..., T$ to estimate the classical mean-variance optimal portfolio for a given target portfolio mean, denoted μ_0, using weights:

$$\hat{w} = \hat{\mathrm{V}}^{-1} \begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix} \left[\begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix}' \hat{V}^{-1} \begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \tag{2.25}$$

$$= \hat{\mathrm{V}}^{-1} \begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix} \hat{B} \begin{pmatrix} 1 \\ \mu_0 \end{pmatrix} \tag{2.26}$$

$$\tag{2.27}$$

where they define $\mathbf{1} = (1, 1, ..., 1)$ an $n \times 1$ vector and $\hat{B} = \left[\begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix}' \hat{\mathrm{V}}^{-1} \begin{pmatrix} \mathbf{1} & \hat{\mu} \end{pmatrix} \right]^{-1}$.

They then define the naïve portfolio mean at the target mean $\mu_0 = \hat{w}'\hat{\mu}$, while the naïve portfolio risk is the standard deviation $\hat{\sigma}_0 = \left(\hat{w}'\hat{\mathrm{V}}\hat{w}\right)^{1/2}$. They finally note that μ_0 and $\hat{\sigma}_0$ would have been respectively the portfolio mean and the portfolio standard deviation if asset returns at time $T+1$ were chosen from a distribution with mean $\hat{\mu}$ and covariance matrix $\hat{\mathrm{V}}$, instead of the true parameters μ and V.

They let $\hat{\sigma}_0 = \left(\hat{w}'\hat{\mathrm{V}}\hat{w}\right)^{1/2}$ be the true portfolio with weights w standard deviation and they use the method of statistical differentials, also called delta method, which substitutes the expected value of a second-order Taylor expansion, expanded about the unknown true values μ and V , in place of the function itself they derive. In this way they are able with the following two theorems to show how an investor should adjust for the bias inherent in the use of estimated values of mean and covariance matrix.

Theorem 1:

The target mean μ_0 is systematically biased as a measure of expected future performance because the delta-method expectation of future performance is:

$$E\left(\hat{w}'R_{T+1}\right) = E\left(\hat{w}'\mu\right) \cong E_{\Delta}\left(\hat{w}'\mu\right) = \mu_0 - \frac{n-3}{T} B_{22}\left(\mu_0 - \mu_*\right) \tag{2.28}$$

where $\mu_* = -B_{12}/B_{22}$ is the mean of the globally minimum variance portfolio. If we define the adjusted mean to be the estimated right-hand side of Equation 2.28, so that

$$\hat{\mu}_{adjusted} = \mu_0 - \frac{n-3}{T} \hat{B}_{22}\left(\mu_0 - \hat{\mu}_*\right) \tag{2.29}$$

where $\hat{\mu}_* = -\hat{B}_{12}/\hat{B}_{22}$, then $\hat{\mu}_{adjusted}$ is a second-order unbiased estimator of the expected future portfolio performance, eliminating the $O\left(1/T\right)$ term from the bias in the sense that

$$E\left(\hat{\mu}_{adjusted}\right) \cong E_{\Delta}\left(\hat{w}'R_{T+1}\right) + O\left(\frac{1}{T^2}\right). \tag{2.30}$$

Theorem 2:

The naïve variance $\hat{\sigma}_0^2$ is systematically biased because its delta-method expectation is:

$$\mathrm{E}(\hat{\sigma}_0^2) \cong E_{\Delta}(\hat{\sigma}_0^2) = \sigma_0^2 + \frac{B_{22}}{T}\sigma_0^2 - \frac{n-4}{T} B_{22}^2\left(\mu_0 - \mu_*\right)^2 - \frac{n-2}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right) \tag{2.31}$$

while the actual portfolio variance, evaluated using the delta-method is:

$$Var(\hat{w}'R_{T+1}) = E\left[\left(\hat{w}'R_{T+1}\right)^2\right] - \left[E\left(\hat{w}'RT+1\right)\right]^2 \tag{2.32}$$

$$= E\left[\hat{w}'V\hat{w} + \hat{w}'\mu\mu'\hat{w}\right] - \left[E\left(\hat{w}'\mu\right)\right]^2 \tag{2.33}$$

$$\cong E_{\Delta}\left[\hat{w}'V\hat{w} + \hat{w}\mu\mu'\hat{w}\right] - \left[E_{\Delta}\left(\hat{w}'\mu\right)\right]^2 \tag{2.34}$$

$$= \sigma_0^2 + \frac{B_{22}}{T}\sigma_0^2 - \frac{n-4}{T}B_{22}^2\left(\mu_0 - \mu_*\right)^2 + \frac{n-1}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right) \tag{2.35}$$

$$= E_{\Delta}(\hat{\sigma}_0^2) + \frac{2n-3}{T}\sigma_0^2 + O\left(\frac{1}{T^2}\right). \tag{2.36}$$

If we define the adjusted standard deviation to be:

$$\hat{\sigma}_{adjusted} = \left(1 + \frac{n-1.5}{T}\right)\hat{\sigma}_0 \tag{2.37}$$

then $\hat{\sigma}^2_{adjusted}$ is a second-order unbiased estimator of actual portfolio variance, eliminating the $O(1/T)$ term from the bias in the sense that

$$E(\hat{\sigma}^2_{adjusted}) \cong E_{\Delta}(\hat{\sigma}^2_{adjusted})$$

$$= Var_{\Delta}\left(\hat{w}'R_{T+1}\right) + O\left(\frac{1}{T^2}\right). \tag{2.38}$$

As expected, the adjusted expected return $\hat{\mu}_{adjusted}$ from Equation 2.29 is less than the target mean μ_0 when the target mean exceeds the estimated mean return $\hat{\mu}_*$ of the globally minimum-risk portfolio and $n > 3$, and the adjusted standard deviation adjusted from Equation 2.37 always exceeds the naïve standard deviation $\hat{\sigma}_0$ that would prevail if the true parameters (μ, V) were equal to the estimated parameters $\left(\hat{\mu}, \hat{V}\right)$. Note that only n and T are used to find the adjusted risk $\hat{\sigma}_{adjusted}$ from $\hat{\sigma}_0$, but that in finding the adjusted mean $\hat{\mu}_{adjusted}$ from μ_0, we also use the estimated values $\hat{\mu}$ and $\hat{V}$ to obtain $\hat{B}_{22}$ and $\hat{\mu}_* = -\hat{B}_{12}/\hat{B}_{22}$.

For both the mean and the standard deviation, the bias adjustment is greater when there are more assets n and when there are fewer time periods T . Adjustments are inversely proportional to the number T of observations, which is reasonable because, with more data, we expect the estimates to be more reliable. Adjustments are directly proportional to a linear function of the number n of assets, which is reasonable because, with more assets, there is more flexibility available to mislead while optimizing over the wrong distribution.

Finally Siegel and Woodgate derive in the following theorem the adjusted efficient frontier, by adjusting both the mean and the standard deviation with the findings of Theorem1 and Theorem2.

Theorem 3:

The adjusted efficient frontier, expressing $\hat{\sigma}_{adjusted}$ as a function of $\hat{\mu}_{adjusted}$, is:

$$\hat{\sigma}_{adjusted} = \left(1 + \frac{n-1.5}{T}\right)\sqrt{\hat{\sigma}_*^2 + \frac{\hat{B_{22}}}{\left[1 - (n-3)\,\hat{B}_{22}/T\right]^2}\left(\hat{\mu}_{adjusted} - \hat{\mu}_*\right)^2} \tag{2.39}$$

where $\hat{\sigma}_*^2 = \sqrt{\hat{B}_{11} - \hat{B}_{12}^2/\hat{B}_{22}}$ is the estimated standard deviation of the global minimum-risk portfolio.

2.5 Estimation of the efficient frontier à la Kan and Smith

2.5.1 Introduction

In their paper, Kan and Smith (2008) contribute to the literature on minimum-variance frontier by deriving the finite sample distribution and moments of the sample minimum-variance frontier when returns are independent multivariate normal random variables. They find that the sample minimum-variance frontier is a heavily biased estimator of the population frontier, even when the length of the estimation window is very long. In order to correct for this bias, they propose a new adjusted estimator of the population frontier that has a significantly lower bias than the traditional sample estimator.

2.5.2 Analysis of sample efficiency set constants

They first study the joint distribution of the three random variables that together determine the sample minimum-variance frontier. They let R_t be the vector of returns on $N \geq 2$ risky assets at time t for $t = 1, ..., T$, where T denotes the length of the time series. For finite sample inference, they assume R_t to be independent and identically distributed as multivariate normal with mean μ and covariance matrix V, making two further assumptions. Namely, they assume that μ is not proportional to 1_N and V is nonsingular. They finally assume that, since in their empirical applications they mainly deal with monthly stock returns, on which evidence of predictability and non-normality is much weaker than in daily returns, they will treat the i.i.d normality assumption as a reasonable working approximation.

For any target expected return μ_p, the minimum-variance portfolio solves the following minimization problem:

$$\begin{gathered} min_{w \in \Re^N} w'Vw \\ \text{subject to } w'\mu = \mu_p \text{ and } w'1_N = 1 \end{gathered} \tag{2.40}$$

where 1_N denotes a N-vector of ones. The minimum-variance frontier traces out the variance of the minimum-variance portfolios for a range of target expected returns μ_p, and it plays a central role in portfolio theory and asset pricing.

They continue by saying that the minimum-variance frontier can be characterized by using the three following efficiency set constants:

$$a = \mu' V^{-1} \mu, \quad b = \mu' V^{-1} 1_N, \quad c = 1_N' V^{-1} 1_N \ . \tag{2.41}$$

These three constants may be remapped to obtain:

$$\psi^2 = a - \tfrac{b^2}{c}, \quad \mu_g = \tfrac{b}{c}, \quad \sigma_g^2 = \tfrac{1}{c} \tag{2.42}$$

where ψ^2 is the square of the slope of the asymptote to the minimum-variance frontier in the (σ, μ) space, and μ_g and σ_g^2 are the mean and tha variance of the global minimum-variance portfolio.

We can also determine (a, b, c) using the following relations:

$$a = \psi^2 + \frac{\mu_g^2}{\sigma_g^2}, \quad b = \frac{\mu_g}{\sigma_g^2}, \quad c = \frac{1}{\sigma_g^2} \ . \tag{2.43}$$

They suppose to have T observations on R_t. The maximum likelihood estimates of μ and V are given by:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} R_t, \tag{2.44}$$

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} \left(R_t - \hat{\mu}\right) \left(R_t - \hat{\mu}\right)' . \tag{2.45}$$

Since it is well-known that under i.i.d. normality assumption, $\hat{\mu}$ and $\hat{V}$ are independent of each other and have the following exact distributions:

$$\hat{\mu} \sim N\left(\mu, V/T\right), \tag{2.46}$$

$$\hat{V} \sim W_N\left(T-1, V/T\right), \tag{2.47}$$

where $W_N\left(T-1, V/T\right)$ is an N-dimensional Wishart distribution[6] with $T-1$ degrees of freedom and covariance matrix V/T. They assume that $T \geq N$ so that $\hat{V}$ is invertible.

They then define the sample efficiency set constants as:

$$\hat{a} = \hat{\mu}' \hat{V}^{-1} \hat{\mu}, \quad \hat{b} = \hat{\mu}' \hat{V}^{-1} 1_N, \quad \hat{c} = 1_N' \hat{V}^{-1} 1_N \tag{2.48}$$

and, equivalently, define the sample counterparts of ψ^2, μ_g, and σ_g^2 as:

$$\hat{\psi}^2 = \hat{a} - \frac{\hat{b}^2}{\hat{c}}, \quad \hat{\mu}_g = \frac{\hat{b}}{\hat{c}}, \quad \hat{\sigma}_g^2 = \frac{1}{\hat{c}} \ . \tag{2.49}$$

Through the following proposition, they define the finite sample joint distribution of $\left(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2\right)$.

Proposition 1:

Let $r \sim (N-1) F_{N-1,T-N+1}\left(T\psi^2\right) / (T-N+1)$, where $F_{m,n}(\delta)$ stands for a non-central F-distribution with m and n degrees of freedom and non-centrality parameter of δ, $q \sim \chi^2_{T-N}$, and

[6] A Wishart distribution is a generalization to multiple dimensions of the chi-squared distrbution, or, in the case of non-integer degress of freedom, of the gamma distribtuion. Suppose X is an $n \times p$ matrix, each row of which is independently drawn from a p-variate normal distribution with zero mean $X_{(i)} = \left(x_i^1, ..., x_i^p\right) \sim N_p\left(0, V\right)$. Then the Wishart distribution is the probability distribution of the $p \times p$ random matrix $S = X^T X$ known as the scatter matrix. One indicates that S has that probability distribution by writing $S \sim W_p(V, n)$. The positive integer n is the number of degrees of freedom. Sometimes this is written $W(V, p, n)$. For $n \geq p$ the matrix S is invertible with probability 1 if V is invertible. If $p = V = 1$ then this distribution is a chi-squared distribution with n degrees of freedom.

$x \sim N(0,1)$, independently of each other. $\hat{\psi}^2, \hat{\mu}_g$, and $\hat{\sigma}_g^2$ can be written as functions of r, q and x as follows:

$$\hat{\psi}^2 = r, \quad \hat{\mu}_g = \mu_g + \left(\tfrac{1+r}{T}\right)^{1/2} \sigma_g x, \quad \hat{\sigma}^2 = \tfrac{\sigma_g^2 q}{T} \ . \tag{2.50}$$

Conditional on $\hat{\psi}^2, \hat{\mu}_g \sim N\left(\mu_g, \sigma_g^2\left(1+\hat{\psi}^2\right)/T\right)$. The joint density function of $\left(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2\right)$ is given by:

$$\begin{array}{c} f\left(\hat{\psi}^2, \hat{\mu}_g, \hat{\sigma}_g^2\right) = f\left(\hat{\psi}^2, \hat{\mu}_g\right) f\left(\hat{\sigma}_g^2\right) \\ = f\left(\hat{\psi}^2\right) f\left(\hat{\mu}_g \mid \hat{\psi}^2\right) f\left(\hat{\sigma}_g^2\right) \end{array} . \tag{2.51}$$

The three sample efficiency set constants are given:

$$\hat{a} = \hat{\psi}^2 + \tfrac{\hat{\mu}_g^2}{\hat{\sigma}_g^2}, \quad \hat{b} = \tfrac{\hat{\mu}_g}{\hat{\sigma}_g^2}, \quad \hat{c} = \tfrac{1}{\hat{\sigma}_g^2} \ .$$

Equipped with Proposition1, they are able to obtain the joint moments of $\left(\hat{a}, \hat{b}, \hat{c}\right)$. Through the following lemma, they present the mean and covariance matrix of this estimators as well as their unbiased ones.

Lemma 1:

The expected values of $\hat{a}, \hat{b}$, and $\hat{c}$ are given by:

$$E[\hat{a}] = \tfrac{N+Ta}{T-N-2}, \quad E\left[\hat{b}\right] = \tfrac{Tb}{T-N-2}, \quad E[\hat{c}] = \tfrac{Tc}{T-N-2} \ . \tag{2.52}$$

The unbiased estimators of a, b and c are given by:

$$\hat{a}_u = \tfrac{(T-N-2)\hat{a}-N}{T}, \quad \hat{b}_u = \tfrac{(T-N-2)\hat{b}}{T}, \quad \hat{c}_u = \tfrac{(T-N-2)\hat{c}}{T} \ . \tag{2.53}$$

The covariance matrix of $\hat{a}, \hat{b}$ and $\hat{c}$ is given by:

$$Var[\hat{a}] = \frac{2T^2a^2 + 2(T-2)(N+2Ta)}{(T-N-1)(T-N-2)(T-N-4)}, \tag{2.54}$$

$$Var\left[\hat{b}\right] = \frac{T^2\left[ac + ((T-2)/T)c + ((T-N)/(T-N-2))b^2\right]}{(T-N-1)(T-N-2)(T-N-4)}, \tag{2.55}$$

$$Var[\hat{c}] = \frac{2T^2c^2}{(T-N-2)^2(T-N-4)}, \tag{2.56}$$

$$Cov\left[\hat{a}, \hat{b}\right] = \frac{2T^2ab + 2(T-2)Tb}{(T-N-2)^2(T-N-4)}, \tag{2.57}$$

$$Cov[\hat{a}, \hat{c}] = \frac{2T^2\left[ac + ((T-2)/T)c + (T-N-2)b^2\right]}{(T-N-1)(T-N-2)^2(T-N-4)}, \tag{2.58}$$

$$Cov\left[\hat{b},\hat{c}\right]=\frac{2T^2bc}{\left(T-N-2\right)^2\left(T-N-4\right)}. \tag{2.59}$$

In the following lemma, they also obtain the joint moments of $\hat{\psi}^2,\hat{\mu}_g$, and $\hat{\sigma}_g^2$.

Lemma 2:

The expected values of $\hat{\psi}^2,\hat{\mu}_g$, and $\hat{\sigma}_g^2$ are given by:

$$E\left[\hat{\psi}^2\right]=\tfrac{N-1+T\psi^2}{T-N-1},\quad E\left[\hat{\mu}\right]=\mu_g,\quad E\left[\hat{\sigma}_g^2\right]=\tfrac{(T-N)\sigma_g^2}{T}\ . \tag{2.60}$$

The unbiased estimators of $\hat{\psi}^2,\hat{\mu}_g$, and $\hat{\sigma}_g^2$ are given by:

$$\hat{\psi}_u^2=\tfrac{(T-N-1)\hat{\psi}^2-(N-1)}{T},\quad \hat{\mu}_{gu}=\hat{\mu}_g,\quad \hat{\sigma}_{gu}^2=\tfrac{T\hat{\sigma}_g^2}{T-N}\ . \tag{2.61}$$

The covariance matrix of $\hat{\psi}^2,\hat{\mu}_g$, and $\hat{\sigma}_g^2$ is given by:

$$Var\left[\hat{\psi}^2\right]=\frac{2T^2\psi^4+2\left(T-2\right)\left(N-1+2T\psi^2\right)}{\left(T-N-1\right)^2\left(T-N-3\right)}, \tag{2.62}$$

$$Var\left[\hat{\mu}_g\right]=\frac{\left[T\left(1+\psi^2\right)-2\right]\sigma_g^2}{T\left(T-N-1\right)}, \tag{2.63}$$

$$Var\left[\hat{\sigma}_g^2\right]=\frac{2\left(T-N\right)\sigma_g^4}{T^2}, \tag{2.64}$$

$$Cov\left[\hat{\psi}^2,\hat{\mu}_g\right]=Cov\left[\hat{\psi}^2,\hat{\sigma}_g^2\right]=Cov\left[\hat{\mu}_g,\hat{\sigma}_g^2\right]=0. \tag{2.65}$$

2.5.3 In-sample performance of sample minimum-variance portfolios

According to Kan and Smith, if μ and V are known, then the minimum-variance portfolio of the N risky assets with targeted expected return of μ_p is given by:

$$w=V^{-1}\left[\mu,1_N\right]A^{-1}\left[\mu_p,1\right]' \tag{2.66}$$

where

$$A=\begin{bmatrix} a & b\\ c & d\end{bmatrix}. \tag{2.67}$$

The variance of this optimal portfolio is given by $\sigma_p^2=w'Vw$. Where σ_p^2 and μ_p are related by the following equation:

$$\sigma_p^2=\frac{a-2b\mu_p+c\mu_p^2}{ac-b^2}. \tag{2.68}$$

By varying Equation 2.68 they are able to obtain the population minimum-variance frontier, represented by the following equation:

$$\sigma_p^2 = \sigma_g^2 + \frac{(\mu_p - \mu_g)^2}{\psi^2}. \tag{2.69}$$

If, instead, μ and V are not known, the minimum-variance portfolio is constructed using the sample moments. The weights of the sample minimum-variance portfolio are given by:

$$\hat{w} = \hat{V}^{-1} [\hat{\mu}, 1_N] \hat{A}^{-1} [\mu_p, 1]' \tag{2.70}$$

where

$$\hat{A} = \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{c} & \hat{d} \end{bmatrix}. \tag{2.71}$$

For a given value of μ_p, the in-sample variance of the sample minimum-variance portfolio is given by:

$$\hat{\sigma}_p^2 = \hat{w}' \hat{V} \hat{w} = \frac{\hat{a} - 2\hat{b}\mu_p + \hat{c}\mu_p^2}{\hat{a}\hat{c} - \hat{b}^2} = \hat{\sigma}_g^2 + \frac{(\mu_p - \hat{\mu}_g)^2}{\hat{\psi}^2}. \tag{2.72}$$

They call Equation 2.72 the sample minimum-variance frontier, and this is typically what researchers plot when they use historical data to estimate the population frontier.

Using Proposition1, we can write the distribution of $\hat{\sigma}_p^2$ as function of the three random variables r, q, x as we can see from the following proposition.

Proposition 2:

Let $u \sim \chi^2_{N-1}\left(T\psi^2\right), v \sim \chi^2_{T-N+1}$, and $\hat{y} \sim N\left(\sqrt{T}\delta, 1\right)$, independent of each other. For a sample minimum-variance portfolio with target expected return of μ_p, the distribution of the sample variance is given by:

$$\hat{\sigma}_p^2 = \frac{\check{\sigma}_p^2 v}{T} \tag{2.73}$$

where

$$\check{\sigma}_p^2 = \sigma_g^2 \left(1 + \frac{\tilde{y}^2}{u}\right). \tag{2.74}$$

Before to introduce the exact moments of $\hat{\sigma}_p^2$, they introduce an important integral, which will prove useful throughout their discussion:

$$\phi = \frac{T\psi^2}{2} \int_0^1 e^{T\psi^2(y-1)/2} y^{(N-3)/2} \mathrm{d}y \quad \text{for} \quad N \geq 2 \ . \tag{2.75}$$

It is also possible to write ϕ using the confluent hypergeometric function as:

$$\phi = \frac{T\psi^2 e^{-T\psi^2/2}}{N-1}\,{}_1F_1\left(\frac{N-1}{2}, \frac{N+1}{2}; \frac{T\psi^2}{2}\right) = \frac{T\psi^2}{N-1}\,{}_1F_1\left(1, \frac{N+1}{2}; -\frac{T\psi^2}{2}\right). \tag{2.76}$$

Let P be an $N \times (N-1)$ orthogonal matrix such that its columns are orthogonal to $V^{-1/2}1_N$, i.e., $P'V^{-1/2}1_N = 0_{N-1}$. Define $z = \sqrt{T}P'V^{-1/2}\hat{\mu} \sim N(\mu_z, I_{N-1})$, where $\mu_Z = \sqrt{T}P'V^{-1/2}\mu$. It is easy to verify that $\mu_z'\mu_z = T\psi^2$, so $u = z'z \sim \chi^2_{N-1}(T\psi^2)$. The following lemma expresses the expectation of various ratios of $z'\mu_z$ and $z'z$ in terms of ϕ.

Lemma 3:

Suppose $z \sim N(\mu_z, I_{N-1})$, and $\mu_z'\mu_z = T\psi^2$. We have:

$$E\left[\frac{1}{z'z}\right] = \frac{1-\phi}{N-3} \quad \text{for} \quad N > 3\ , \tag{2.77}$$

$$E\left[\frac{1}{(z'z)^2}\right] = \frac{(N-5)\phi - T\psi^2(1-\phi) + 2}{2(N-3)(N-5)} \quad \text{for} \quad N > 5\ , \tag{2.78}$$

$$E\left[\frac{z'\mu_z}{z'z}\right] = \phi \quad \text{for} \quad N > 2\ , \tag{2.79}$$

$$E\left[\frac{z'\mu_z}{(z'z)^2}\right] = \frac{T\psi^2(1-\phi)}{2(N-3)} - \frac{\phi}{2} \quad \text{for} \quad N > 3\ , \tag{2.80}$$

$$E\left[\frac{(z'\mu_z)^2}{z'z}\right] = T\psi^2 - (N-2)\phi \quad \text{for} \quad N > 1\ , \tag{2.81}$$

$$E\left[\frac{(z'\mu_z)^2}{(z'z)^2}\right] = \frac{(N-2)\phi}{2} - \frac{T\psi^2(N-4)(1-\phi)}{2(N-3)} \quad \text{for} \quad N > 3\ . \tag{2.82}$$

Thanks to the above Lemma3, they are able to state the following proposition which presents the expected value and variance of $\hat{\sigma}_p^2$.

Proposition 3:

For a sample minimum-variance portfolio with target expected return of μ_p, its expected in-sample variance exists if and only if $N > 3$, the expected value of $\hat{\sigma}_p^2$ is given by:

$$\bar{\sigma}_p^2 = E\left[\hat{\sigma}_p^2\right] = \left(\frac{T-N+1}{T}\right) E\left[\check{\sigma}_p^2\right] \tag{2.83}$$

where $E\left[\check{\sigma}_p^2\right] = \sigma_g^2\left(1 + hE\left[u^{-1}\right]\right)$ and $h = T\delta^2 + 1$. In addition, the variance of $\hat{\sigma}_p^2$ exists if and only if $N > 5$. When $N > 5$, the variance of $\hat{\sigma}_p^2$ is given by:

$$Var\left[\hat{\sigma}_p^2\right] = \frac{(T-N+1)\left[(T-N+3)\,Var\left[\check{\sigma}_p^2\right] + E\left[\check{\sigma}_p^2\right]^2\right]}{T^2} \tag{2.84}$$

where $Var\left[\check{\sigma}_p^2\right] = \sigma_g^4\left[\left(h^2+4h-2\right)E\left[u^{-2}\right] - \left(hE\left[u^{-1}\right]\right)^2\right]$, with $E\left[u^{-1}\right]$ and $E\left[u^{-2}\right]$ are given in Equation 2.77 and Equation 2.78.

Using simulation, they readily show that the sample minimum-variance frontier is hugely biased and the bias is most pronounced for portfolios that are far from the global minimum-variance portfolio. This bias remain quite pronounced even when using moments estimated with 50 years of monthly data. From their empirical experiment, it is evident that the sample minimum-variance provides an overly optimistic assessment of the population frontier.

They identify as the main reason for this huge bias in the sample minimum-variance frontier, the use of $1/\hat{\psi}^2$ to estimate $1/\psi^2$. From Proposition1, it is known that $\hat{\psi}^2 \sim \chi^2_{N-1}\left(T\psi^2\right)/\chi^2_{T-N+1}$. Therefore, using the results in Lemma3, the expected value of $1/\hat{\psi}^2$ is given by:

$$E\left[\frac{1}{\hat{\psi}^2}\right] = \frac{(T-N+1)(1-\phi)}{N-3}. \tag{2.85}$$

Although the functional form of $1/\hat{\psi}^2$, correcting for its bias is not an easy task because the bias is determined by N, T and ψ^2, through the function ψ which, as we have seen previously, is a confluent hypergeometric function, a special function very difficult to approximate by polynomials. That is why standard bias reduction methods, i.e. Taylor series expansions or resampling techniques, jackknife method, do not provide a significant bias reduction.

If they were able to find an adjusted estimator $\hat{\psi}^2_a$, then we can construct an unbiased estimator of σ^2_p as follows:

$$\hat{\sigma}^2_{pu} = \frac{T\hat{\sigma}^2_g}{T-N} + \frac{1}{\hat{\psi}^2_a}\left[(\mu_p - \hat{\mu}_g)^2 - \frac{\hat{\sigma}^2_g\left(1+\hat{\psi}^2\right)}{T-N}\right]. \tag{2.86}$$

But, since they cannot find a way to construct an unbiased estimator of σ^2_p , they thus try to find an improved estimator of $1/\psi^2$.

After considering many different estimators, they settle for the following:

$$\frac{1}{\hat{\psi}^2} = \frac{T\int_0^z u^{(T-N-1)/2}(1-u)^{(N-5)/2}\,\mathrm{d}u}{2z^{(T-N-1)/2}(1-z)^{(N-3)/2}} \tag{2.87}$$

where $z = 1/\left(1+\hat{\psi}^2\right)$. As stated above, this estimator is not unbiased; but its bias is significantly less than before. In the following lemma, Kan and Smith present the expectation of this adjusted estimator.

Lemma 4:

When $N > 3$, the expectation of $1/\hat{\psi}^2_a$ exists and it is given by:

$$E\left[\frac{1}{\hat{\psi}^2}\right] = \frac{1-e^{-T\psi^2/2}}{\psi^2}. \tag{2.88}$$

This adjusted estimator can now be plugged into Equation 2.86 to obtain an approximately unbiased estimator of σ^2_p. Since $\hat{\sigma}^2_{pu}$ can be negative and hence inadmissible, the adjusted estimator of σ^2_p becomes:

$$\hat{\sigma}_{pa}^2 = \frac{T\hat{\sigma}_g^2}{T-N} + \frac{1}{\hat{\psi}_a^2}\max\left[(\mu_p - \hat{\mu}_g)^2 - \frac{\hat{\sigma}_g^2\left(1+\hat{\psi}^2\right)}{T-N}, 0\right]. \tag{2.89}$$

While there is still some bias when T is small, the bias becomes almost negligible when $T > 120$ months.

2.5.4 Out-of-sample performance of sample minimum-variance portfolios

Finally, Kan and Smith analyze the out-of-sample mean and variance of a sample minimum-variance portfolio. For the sample minimum-variance frontier, $\hat{w}$ is chosen such that $\hat{\mu}_p = \hat{w}'\hat{\mu} = \mu_p$, so there is no randomness in the in-sample mean. As a result, the only random variable is the in-sample variance $\hat{\sigma}_p^2 = \hat{w}'\hat{V}\hat{w}$. However, for the out-of-sample case, both the out-of-sample mean and variance, i.e., $\tilde{\mu}_p = \hat{w}'\mu$ and $\tilde{\sigma}_p^2 = \hat{w}'V\hat{w}$, are random variables. Meaning that there is one more random variable to study for the out-of-sample case. The following proposition presents the joint distribution of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$.

Proposition 5:

Let $\tilde{x} \sim N(0,1), \tilde{q} \sim \chi_{N-3}^2, \tilde{r} \sim \chi_{T-N+2}^2, \tilde{u} \sim \chi_{N-2}^2, \tilde{y} \sim N\left(\sqrt{T}\delta, 1\right)$, and $\tilde{z} \sim N\left(\sqrt{T}\psi, 1\right)$ and independent of each other. Let $u = \tilde{z}^2 + \tilde{u} \sim \chi_{N-1}^2\left(T\psi^2\right)$ and $\tilde{t} = \tilde{x}/\tilde{r}^{1/2}$. For a sample minimum-variance portfolio with target return of μ_p, its out-of-sample expected return and variance can be written as follows:

$$\tilde{\mu}_p = \mu_g + \psi\left[\sigma_g\frac{\tilde{z}}{u}\tilde{y} + \check{\sigma}_p\left(\frac{\tilde{u}}{u}\right)^{1/2}\tilde{t}\right], \tag{2.90}$$

$$\tilde{\sigma}_p^2 = \check{\sigma}_p^2\left(1 + \frac{\tilde{x}^2+\tilde{q}}{\tilde{r}}\right), \tag{2.91}$$

where $\check{\sigma}_p^2$ is defined in Proposition2.

There are now two sources of randomness in the weights of the sample minimum-variance portfolio; one is due to $\hat{\mu}$ and the other is due to $\hat{V}$. If V is known but $\mu$is not known, then the distribution of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ will be simplified to:

$$\tilde{\mu}_p = \mu_g + \psi\sigma_g\frac{\tilde{z}}{u}\tilde{y}, \tag{2.92}$$

$$\tilde{\sigma}_p^2 = \check{\sigma}_p^2. \tag{2.93}$$

The estimation risk of $\hat{V}$ adds to the uncertainty of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$. In particular, it will always increase $\tilde{\sigma}_p^2$ by a random factor of $1 + \left(\tilde{x}^2 + \tilde{q}\right)/\tilde{r}$ and this further degrades the out-of-sample performance of the sample minimum-variance portfolio. On the other hand, if μ is known but V is not known, $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ will become:

$$\tilde{\mu}_p = \mu_p, \tag{2.94}$$

$$\tilde{\sigma}_p^2 = \sigma_p^2 \left(1 + \frac{\tilde{x}^2 + \tilde{q}}{\tilde{r}}\right), \tag{2.95}$$

meaning that $\tilde{\mu}_p$ will always be equal to the target expected return μ_p and only $\tilde{\sigma}_p^2$ is random.

Because the out-of-sample mean and variance of a sample minimum-variance are random variables, we are interested in their expectations, which will give us an idea of the average out-of-sample return and risk that an investor can obtain by holding a sample minimum-variance portfolio. The following proposition presents the expected out-of-sample mean and variance of a sample minimum-variance portfolio, and it also presents the covariance matrix of $\left(\tilde{\mu}_p, \tilde{\sigma}_p^2, \hat{\sigma}_p^2\right)$.

Proposition 6:

For a sample minimum-variance portfolio with target expected return of μ_p, its expected out-of-sample mean exists if and only if $N > 2$. When $N > 2$, the expected value of $\tilde{\mu}_p$ is given by:

$$\underset{\sim}{\mu}_p = E\left[\bar{\mu}_p\right] = \mu_p - (1-\phi)\left(\mu_p - \mu_g\right) \tag{2.96}$$

where ϕ is defined in Equation 2.75. In addition, the expected out-of-sample minimum-variance portfolio exists if and only if $N > 3$. When $N > 3$, the expected value of $\tilde{\sigma}_p^2$ is given by:

$$\underset{\sim}{\sigma}_p^2 = E\left[\tilde{\sigma}_p^2\right] = \left(\frac{T-2}{T-N}\right) E\left[\check{\sigma}_p^2\right]. \tag{2.97}$$

For the covariance matrix of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$, $Var\left[\tilde{\mu}_p\right]$ and $Cov\left[\tilde{\mu}, \tilde{\sigma}_p^2\right]$ exist if and only if $N > 3$, and $Var\left[\tilde{\sigma}_p^2\right]$ exists if and only $N > 5$. When they exist, the variances and covariances are given by:

$$\begin{aligned} Var\left[\mu_p\right] = {} & \frac{\psi^2 E\left[\check{\sigma}_p^2\right] - \left(\sigma_g^2/T\right)\left(E\left[\left(z'\mu_z\right)^2/\left(z'z\right)\right] + hE\left[\left(z'\mu_z\right)^2/\left(z'z\right)^2\right]\right)}{T-N} + \\ & + \frac{\sigma_g^2 h}{T} E\left[\frac{\left(z'\mu_z\right)^2}{\left(z'z\right)^2}\right] - \phi^2\left(\mu_p - \mu_g\right)^2 \end{aligned}, \tag{2.98}$$

$$Var\left[\tilde{\sigma}_p^2\right] = \frac{(T-2)}{(T-N)(T-N-2)} \cdot \left[(T-4)\, Var\left[\check{\sigma}_p^2\right] + \frac{2(N-2)\, E\left[\check{\sigma}_p^2\right]^2}{(T-N)}\right], \tag{2.99}$$

$$Cov\left[\tilde{\mu}_p, \tilde{\sigma}_p^2\right] = \frac{(T-2)}{(T-N)}\left(\mu_p - \mu_g\right)\sigma_g^2 \cdot \left[(h+2)\, E\left[\frac{z'\mu_z}{\left(z'z\right)^2}\right] - \frac{h\phi(1-\phi)}{N-3}\right], \tag{2.100}$$

where $E\left[\left(z'\mu_z\right)^2/\left(z'z\right)\right], E\left[\left(z'\mu_z\right)/\left(z'z\right)^2\right]$, and $E\left[\left(z'\mu_z\right)^2/\left(z'z\right)^2\right]$ are given in Lemma3, and $h, E\left[\check{\sigma}_p^2\right]$, and $Var\left[\check{\sigma}_p^2\right]$ are defined in Proposition3. The covariances of the in-sample variance $\hat{\sigma}_p^2$ with $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ are given by:

$$Cov\left[\tilde{\mu}_p, \hat{\sigma}_p^2\right] = \left[\frac{(T-N+1)(T-N)}{T(T-2)}\right] Cov\left[\tilde{\mu}_p, \tilde{\sigma}_p^2\right], \tag{2.101}$$

$$Cov\left[\tilde{\sigma}_p^2, \hat{\sigma}^2\right] = \left[\frac{(T-N+1)(T-2)}{T(T-N)}\right] Var\left[\breve{\sigma}_p^2\right]. \tag{2.102}$$

They are thus able to check that the in-sample frontier is in general a downward biased estimate of the population frontier, which in turn heavily dominates the out-of-sample frontier. If an investor chooses to hold a sample minimum-variance portfolio, he is interested in the out-of-sample frontier because that represents the average true performance of his portfolio. Nevertheless, what he observes from the data is the sample frontier, and it is on average much better than the out-of-sample performance of his portfolio, he will be grossly disappointed. The sources of disappointment will be the following:

- From Equation 2.96 we can see that the average out-of-sample mean $\underline{\mu}_p$ shrinks toward the expected return of the global minimum-variance portfolio μ_g. As a result, $\underline{\mu}_p$ tends to be lower than the target expected return μ_p when $\mu_p > \mu_g$, so the investor will on average be disappointed with the out-of-sample mean of his portfolio.

- As for the out-of-sample variance of the portfolio, from Equation 2.83 and Equation 2.97 we obtain
$$\underline{\sigma}_p^2 = \left[\frac{(T-2)T}{(T-N)(T-N+1)}\right] \bar{\sigma}_p^2 \tag{2.103}$$
so the average out-of-sample variance of the sample minimum-variance portfolio $\underline{\sigma}_p^2$ tends to be much higher than its average in-sample variance $\bar{\sigma}_p^2$, especially when N/T is large. the investor will be thus disappointed by the larger-than-expected risk of his portfolio.

These results suggest that for a sample minimum-variance portfolio, the sample mean and the sample variance are too optimistic as forecast of its out-of-sample mean and variance. To come up with a reliable forecast of the true out-of-sample performance of a sample minimum-variance portfolio, we need to use a less optimistic forecast. The following proposition provides the unbiased forecast of $\underline{\mu}_p$and $\underline{\sigma}_p^2$.

Proposition 7:

The unbiased estimators of $\underline{\mu}_p$and $\underline{\sigma}_p^2$ are given by:

$$\hat{\mu}_{pa} = \mu_{pu} - \frac{N-3}{(T-N+1)\hat{\psi}^2}(\mu_p - \hat{\mu}_g) \text{ for } N > 3, \tag{2.104}$$

$$\hat{\sigma}_{pa}^2 = \left[\frac{(T-2)T}{(T-N)(T-N+1)}\right] \hat{\sigma}_p^2 \text{ for } N > 5. \tag{2.105}$$

Because $\underline{\mu}_p$ shrinks toward μ_g, our adjusted estimator also shrinks μ_p toward $\hat{\mu}_g$ to get an unbiased forecast of $\underline{\mu}_p$. In addition, the sample variance $\hat{\sigma}_p^2$ is too optimistic as a forecast of the out-of-sample variance. Proposition7 suggests that one should multiply $\hat{\sigma}_p^2$ by an adjustment factor of $(T-2)T/[(T-N)(T-N+1)]$ to get an unbiased forecast of the out-of-sample variance of the sample minimum-variance portfolio.

As a final remark, we should highlight that similar formulas are also presented in Siegel and Woodgate (2007), even if as we have seen before, their formulas are only approximately unbiased and not exact.

2.6 Estimation of the efficient frontier à la Bodnar and Bodnar

2.6.1 Introduction

In order to derive an unbiased estimator of the efficient frontier, Bodnar and Bodnar (2010) start by recalling Markowitz's mean-variance analysis formulation. They define an investor as someone holding a portfolio made up of k assets. Letting μ be the vector of the expected returns and Σ the variance-covariance matrix of asset returns, assuming this latter to be positive definite, they recall Merton's work, which proved that the efficient frontier is the upper part of a parabola given by:

$$g_p(R, V) = (R - R_{GMV})^2 - s(V - V_{GMV}), \tag{2.106}$$

where

$$s = \mu' R \mu \tag{2.107}$$

with

$$R = \Sigma^{-1} - \frac{\Sigma^{-1} 1 1' \Sigma^{-1}}{1' \Sigma^{-1} 1}$$

is the slope of the the efficient frontier.

The quantities

$$R_{GMV} = \frac{1' \Sigma^{-1} \mu}{1' \Sigma^{-1} 1} \tag{2.108}$$

and

$$V_{GMV} = \frac{1}{1' \Sigma^{-1} 1} \tag{2.109}$$

are the expected return and variance of the global minimum variance (GMV) portfolio. These parameters are those defining the vertex of the efficient frontier. The GMV portfolio is a special portfolio on the efficient frontier. It is the one characterized by the smallest possible variance and, moreover, it is the only portfolio which is independent of the market profitability that is given by the slope parameter of the parabola, s.

2.6.2 Sample efficient frontier

Statistically speaking, the sample efficient frontier is the estimation of the population efficient frontier and it is obtained by simply plugging the sample estimators of the expected returns and variance-covariance matrix, namely:

$$\hat{\mu} = \frac{1}{n} \sum_{J=1}^{n} X_j, \tag{2.110}$$

$$\Sigma = \frac{1}{n-1}\sum_{j=1}^{n}(\hat{X_j} - \mu)(\hat{X_j} - \mu)', \tag{2.111}$$

in Equation 2.106 instead of the unknown parameters μ and Σ, so that our efficient frontier becomes:

$$g_s(R, V) = (R - \hat{R}_{GMV})^2 - \hat{s}(V - \hat{V}_{GMV}) \tag{2.112}$$

where

$$\hat{R}_{GMV} = \frac{1'\hat{\Sigma}^{-1}\hat{\mu}}{1'\hat{\Sigma}^{-1}1}, \tag{2.113}$$

$$\hat{V}_{GMV} = \frac{1}{1'\hat{\Sigma}^{-1}1}, \tag{2.114}$$

$$\hat{s} = \hat{\mu}'\hat{R}\hat{\mu}, \tag{2.115}$$

with

$$\hat{R} = \frac{\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}11'\hat{\Sigma}^{-1}}{1'\hat{\Sigma}^{-1}1}.$$

Assuming that the asset returns are independently and normally distributed, Bodnar and Schmid (2008) and Kan and Smith (2008) independently derived the exact distributions of $\hat{R}_{GMV}$, $\hat{V}_{GMV}$, and $\hat{s}$, while Jobson (1991) derives them asymptotically.

Let χ^2_p denote a χ^2-distribution with p degrees of freedom and the symbol $N(\mu, \sigma^2)$ stands for the normal distribution with mean μ and variance σ^2. By $F_{a,b,k}$we denote the non-central F-distribution with a and b degrees of freedom and non-centrality parameter k. Bodnar and Schmid prove how to derive the exact distribution of $\hat{R}_{GMV}, \hat{V}_{GMV,}$ and $\hat{s}$, which we will only present through the following proposition.

Proposition 1:

Let $X_1, ..., X_n$ be independent random vectors and let $X_i \sim N_k(\mu, \Sigma)$ for $i = 1, ..., n$. Assume that Σ is positive definite. Let $k > 2$ and $n > k$. Then it holds that:

1. $\hat{V}_{GMV}$ is independent of $(\hat{R}_{GMV}, \hat{s})$.
2. $(n-1)\hat{V}_{GMV}/V_{GMV} \sim \chi^2_{n-k}$.
3. $\frac{n(n-k+1)}{(n-1)(k-1)}\hat{s} \sim F_{k-1,n-k+1,ns}$.
4. $\hat{R}_{GMV} \mid \hat{s} = y \sim N(R_{GMV}, \frac{1+\frac{n}{n-1}y}{n}V_{GMV})$.

Then, for proving the bias of the sample efficient frontier, we need the following lemma.

Lemma 1:

Let $X_1, ..., X_n$ be independent random vectors and let $X_i \sim N_k(\mu, \Sigma)$ for $i = 1, ..., n$. Assume that Σ is positive definite. Let $k > 2$ and $n > k$. Then it holds that:

1. $E(\hat{s}) = \frac{n-1}{n-k+1}s + \frac{(n-1)(k-1)}{n(n-k+1)}$.
2. $E((R - \hat{R}_{GMV})^2) = (R - R_{GMV})^2 + \left(\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1}s)\right) V$.
3. $E(\hat{V}_{GMV}) = \frac{n-k}{n-1}V_{GMV}$.

Using these results, we are able to show that $E(g_s(R,V)) \neq 0$ for $(R,V) \in \mathbb{R} \times \mathbb{R}_+$ such that $g_p(R,V) = 0$. It holds that:

$$\begin{aligned} E(g_s(R,V)) &= E((R - \hat{R}_{GMV})^2) - E(\hat{s}(V - \hat{V}_{GMV})) \\ &= E((R - \hat{R}_{GMV})^2) - E(\hat{s})(V - E(\hat{V}_{GMV})) \end{aligned}$$

where the last identity follows from point 2 of Proposition1.

Moreover, application of the Lemma above leads to:

$$\begin{aligned} E(g_s(R,V)) = (R - R_{GMV})^2 + \left(\frac{n-2}{n(n-k-1)} + \frac{1}{n-k-1}s\right) V_{GMV} + \\ -(\frac{n-1}{n-k-1}s + \frac{(n-1)(k-1)}{n(n-k-1)})(V - \frac{n-k}{n-1}V_{GMV}) \end{aligned} .$$

Hence,

$$\begin{aligned} E(g_s(R,V)) = g_p(R,V) - (\frac{k}{n-k-1}s + \frac{(n-1)(k-1)}{n(n-k-1)})V + \\ +(\frac{2}{n-k-1}s + \frac{(n-k)(k-1)+n-2}{n(n-k-1)})V_{GMV} \end{aligned} .$$

The last equality shows that the sample efficient frontier provides an unbiased estimator only for one portfolio, namely the portfolio with $V = \gamma V_{GMV}$ where

$$\gamma = \frac{\frac{2}{n-k-1}s + \frac{(n-k)(k-1)+n-2}{n(n-k-1)}}{\frac{k}{n-k-1}s + \frac{(n-1)(k-1)}{n(n-k-1)}} = \frac{2ns + (n-k)(k-1) + n - 2}{kns + (n-k)(k-1)}.$$

If $V > \gamma V_{GMV}$ then the sample efficient frontier overestimates the population efficient frontier.

If, instead, $V < \gamma V_{GMV}$, we witness underestimation.

If $1 + \frac{-ns+\sqrt{(ns+2)^2+4n-12}}{2} < k < n$, we obtain that $\gamma < 1$ and hence the sample efficient frontier overestimates the population efficient frontier for all $V \geqslant V_{GMV}$ in this case. The last finding shows that the mean-variance efficient portfolios are overestimated when the number of the assets is large enough when compared to the sample size.

2.6.3 Unbiased estimator of the efficient frontier

After having shown that the sample efficient frontier is a biased estimator of the population efficient frontier, Bodnar and Bodnar (2010) suggest a new estimator, correcting the overoptimism of the sample efficient frontier. In the next theorem it is proved that this estimator is an unbiased estimator of the efficient frontier.

Theorem:

Let $X_1, ..., X_n$ be independent random vectors and let $X_i \sim N_k(\mu, \Sigma)$ for $i = 1, ..., n$. Assume that Σ is positive definite. Let $k > 2$ and $n > k$. Then the unbiased estimator of the efficient frontier is given by

$$g_u(R, V) = (R - \hat{R}_{GMV})^2 - \frac{(n-2)(n-1)}{n(n-k)(n-k-1)}\hat{V}_{GMV} \tag{2.116}$$

$$-\left(\frac{n-k-1}{n-1}\hat{s} - \frac{k-1}{n}\right)\left(V - \frac{(n-k-2)(n-1)}{(n-k-1)(n-k)}\hat{V}_{GMV}\right).$$

Proof:

For proving this theorem, we need to show that $E(g_u(R, V)) = 0$ for all points $(R, V) \in \mathbb{R} \times \mathbb{R}_+$ such that $g_p(R, V) = 0$. Let us consider:

$$E(g_u(R, V) = E((R - \hat{R}_{GMV})^2) - E\left(\frac{(n-2)(n-1)}{n(n-k)(n-k-1)}\hat{V}_{GMV}\right)$$

$$-E((\frac{n-k-1}{n-1}\hat{s} - \frac{k-1}{n})(V - \frac{(n-k-2)(n-1)}{(n-k-1)(n-k)}\hat{V}_{GMV})).$$

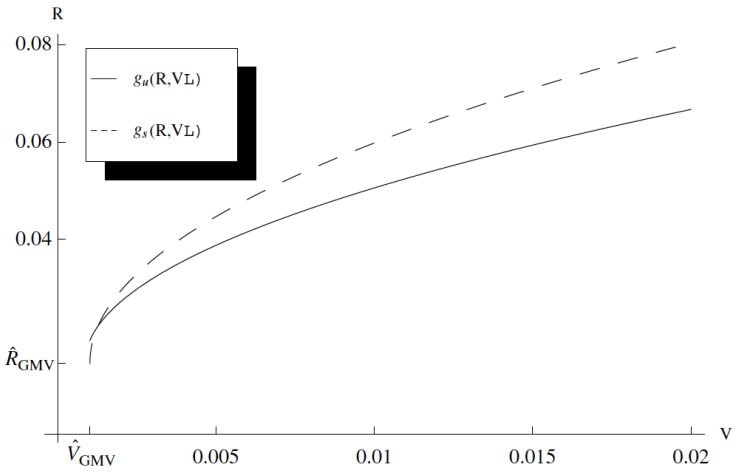
Application of the Proposition and the Lemma enunciated before, leads to:

$$E(g_u(R, V)) = (R - R_{GMV})^2 + \left(\tfrac{n-2}{n(n-k-1)} + \tfrac{1}{n-k-1}s\right)V_{GMV} +$$
$$-\tfrac{n-2}{n(n-k-1)}V_{GMV} - s\left(V - \tfrac{n-k-2}{n-k-1}\hat{V}_{GMV}\right) =$$

$=$(R-R$_{GMV}$)$^2 - s(V - V_{GMV}) = g_p(R, V) = 0.$

These results shows that the application of $\hat{V}_{GMV}$ and $\hat{s}$ in estimating the efficient frontier leads to a considerable bias in the estimator of the efficient frontier. We can even show these results graphically. In Figure 2.1 we see the comparison between the unbiased efficient frontier $g_u(R, V)$ and the sample estimator $g_s(R, V)$ given in Equation 2.112. We can see that $g_u(R, V)$ lie below $g_s(R, V)$ for larger values of V and about for low values. From this, stems the fact that the unbiased estimator corrects the overoptimism of the investor for larger values of V that is caused by the application of the sample efficient frontier.

Figure 2.1: Sample and unbiased efficient frontier.



Source:Bodnar, O. and Bodnar, T.:On the unbiased estimator of the efficient frontier, International Journal of Theoretical and Applied Finance 7 (2010) p.1071

2.6.4 Overall F-test on the efficient frontier

Bodnar and Bodnar conclude their study by deriving an F-test on the efficient frontier. In this framework the slope parameter plays an important role. As a matter of fact, the slope shows how profitable the market is, or how large the increase in the portfolio's profit is relative to a unit increase of the portfolio's variance.

When $s = 0$ in Equation 2.106, the population efficient frontier degenerates into a straight line. In such a case, there is only one optimal portfolio, namely the global minimum, variance portfolio.

Bodnar and Bodnar (2010) thus propose a test for the efficiency of any portfolio from the efficient frontier with respect to this global minimum variance portfolio. The test hypothesis is given by:

$$\begin{cases} H_0 : s = 0 \\ H_1 : s > 0 \end{cases}. \tag{2.117}$$

The non-rejection of the null hypothesis means that the slope coefficient of the efficient frontier is equal to zero, meaning that there is no reason for the investor to continue to hold a risky portfolio. As an alternative investment possibility Bodnar and Bodnar consider the GMV portfolio which is independent of s. Such a test can be performed by making use of Proposition (3). Because the non-central F-distribution with $s = 0$ is a central F-distribution, it holds that the null hypothesis is rejected if

$$T_S = \frac{n(n-k+1)}{(n-1)(k-1)}\hat{s} > F_{k-1,n-k+1;\alpha}. \tag{2.118}$$

They also underline that the same test statistics is obtained if the hypothesis $H_0 = \mu_1 = \mu_2 = ... = \mu_k$ is tested. This is due to the fact that the hypothesis are equivalent.

in both cases, there is only one optimal portfolio: the global minimum variance portfolio.

The test statistic in Equation 2.118 has an important application. In the Theorem we enunciated before, an unbiased estimator of the population efficient frontier is derived. Because this estimator is a function of the sample it may happen that $\hat{\hat{s}} < 0$ which is the case if $\frac{n-k-1}{n-1}\hat{s} - \frac{k-1}{n} < 0$ or, equivalently, if $T_S < \frac{n-k+1}{n-k-1}$. The last inequality is true in the most of cases if and only if the null hypothesis of the test in Equation 2.117 cannot be rejected. In this case, the efficient frontier consists only of the GMV portfolio and is independent of s.

Chapter 3

The unbiased estimator à la Bodnar and Bodnar

In this chapter we will firstly recall the assumptions of the model on which our empirical analysis will be based on. We will also explain its drawbacks and fallacies and present some propositions stemming from the model. Once we have clear in mind our instruments, we will introduce our dataset and analyze it. We will then close this chapter with some important remarks concerning the slope of the unbiased efficient frontier.

3.1 The unbiased efficient frontier à la Bodnar and Bodnar

As we have already analyzed in the last section of Chapter 2, the unbiased estimator proposed by Bodnar and Bodnar is based on the following assumptions:

- k assets, with $k > 2$,
- n observations, with $n > k$,
- normally and independently distributed logarithmic returns.

From this assumptions, Bodnar and Bodnar (2010) are able to come up with the result that the sample efficient frontier in the following Equation 3.1 is biased.

$$g_s(R, V) = (R - \hat{R}_{GMV})^2 - \hat{s}(V - \hat{V}_{GMV}). \tag{3.1}$$

In mathematical terms, this can be stated as follows:

$$E(g_s(R, V)) \neq 0 \text{ for(R,V)} \in \mathbb{R} \times \mathbb{R}_+.$$

They also prove that the classical estimator of the efficient frontier underestimates risk.

They then propose the following unbiased estimator:

$$g_u\left(R,V\right)=\left(R-\hat{R}_{GMV}\right)^2-A\cdot\hat{V}_{GMV}-\left(B\cdot\hat{s}-C\right)\left(V-D\cdot\hat{V}_{GMV}\right)=0, \tag{3.2}$$

where:

$$\begin{aligned} A&=\tfrac{(n-2)(n-1)}{n(n-k)(n-k+1)},\\ B&=\tfrac{n-k-1}{n-1},\\ C&=\tfrac{k-1}{n},\\ D&=\tfrac{(n-k-2)(n-1)}{(n-k-1)(n-k)}. \end{aligned} \tag{3.3}$$

3.2 Limits and fallacies of the unbiased efficient frontier à la Bodnar and Bodnar

Although this approach seems interesting on a theoretical ground, it also presents some limits and drawbacks when applied practically to real data.

The first problem we encounter when faced with the use of the unbiased efficient frontier à la Bodnar and Bodnar, deals with the main hypothesis on which their work is based on. Namely, the hypothesis according to which the assets' returns must be multivariate normally distributed and independent, thus determining a diagonal matrix of the covariances of the following type:

$$\sum=\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 & 0\\ 0 & \sigma_2^2 & 0 & 0 & 0\\ 0 & 0 & \sigma_3^2 & 0 & 0\\ 0 & 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & 0 & \sigma_n^2 \end{pmatrix}. \tag{3.4}$$

The matrix in Equation 3.4, however, does not describe properly how things work in the real world, and we will also show this later in the chapter when analyzing our dataset.

A second problem which appears when dealing with the unbiased efficient frontier à la Bodnar and Bodnar, which has also been underlined by a study by Corazza and Pizzi (2014), is the strange shape that the unbiased efficient frontier takes on in some occasions. We will better explain this point in the next paragraph, when analyzing their results and propositions.

At last, we can say that the operational efficiency of working with this kind of unbiased estimator of the efficient frontier seems limited on a practical ground. As a matter of fact in order to check whether the unbiased efficient frontier à la Bodnar and Bodnar is worth the implementation of a more complicated model than the classical one, we can construct a confidence interval, starting from a given confidence level α, which will define an upper and a lower limit in which our estimator will fall into. Once we have defined such a confidence interval, how do we check whether the unbiased estimator à la Bodnar and Bodnar is truly an enhancement with respect to the classical sample efficient frontier? We only have to check where the unbiased efficient frontier falls in the confidence

interval just drawn, and compare its position with that of the classical sample efficient frontier in this very same interval. More precisely we can say that:

- if the unbiased efficient frontier falls within the sample efficient frontier's confidence interval, the two frontiers are statistically equivalent;
- otherwise, the unbiased efficient frontier is outside the confidence interval of the sample efficient frontier, the two frontiers are not equivalent and thus it will be preferable to use the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar.

3.3 Considerations on the unbiased efficient frontier à la Bodnar and Bodnar by Corazza and Pizzi

In their work on the unbiased efficient estimator of the efficient frontier proposed by Bodnar and Bodnar, Corazza and Pizzi (2014) study the asymptotic behavior of the unbiased efficient estimator of the efficient frontier, they analyze the operational effectiveness of the unbiased estimator of the efficient frontier by a bootstrap approach, they also investigate the strange behavior of the unbiased estimator of the efficient frontier.

By doing this, they are able to state the following two propositions[1]:

Proposition 1:

The unbiased efficient frontier à la Bodnar and Bodnar, as the number of observations n increases, tends asymptotically to the sample efficient frontier.

$$lim_{n\to\infty}\text{Unbiased Efficient Frontier} = \text{Sample Efficient Frontier}. \tag{3.5}$$

This can be seen from Equation 3.3. We can suppose that the following applies:

$$\begin{aligned} A &= \tfrac{(n-2)(n-1)}{n(n-k)(n-k+1)}, \text{ with } A \sim \tfrac{1}{n}, \\ B &= \tfrac{n-k-1}{n-1}, \text{ with } B \sim 1, \\ C &= \tfrac{k-1}{n}, \text{ with } C \sim \tfrac{1}{n}, \\ D &= \tfrac{(n-k-2)(n-1)}{(n-k-1)(n-k)}, \text{ with } D \sim 1. \end{aligned} \tag{3.6}$$

From Equation 3.6, we can see that if $n \to +\infty$ the following will asymptotically hold true:

[1]For proofs of the two propositions see MAF presentation by Corazza and Pizzi (2014).

$$\begin{aligned} lim_{n\to\infty}A &= 0, \\ lim_{n\to\infty}B &= 1, \\ lim_{n\to\infty}C &= 0, \\ lim_{n\to\infty}D &= 1. \end{aligned} \tag{3.7}$$

Proposition 2:

The unbiased efficient frontier à la Bodnar and Bodnar takes negative values of the variance when:

$$\hat{s} < \frac{(n-1)\,k}{n\,(n-k-2)}, \tag{3.8}$$

where:

$$\hat{s} = \sum_{i=1}^{k} \frac{\hat{\mu}_i^2}{\hat{\sigma}_i^2}\,(1-\hat{p}_i) - \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{\hat{\mu}_i \hat{\mu}_j}{\hat{\sigma}_j^2}\hat{p}_j, \tag{3.9}$$

with:

$$\hat{p}_i = \frac{\frac{1}{\hat{\sigma}_i^2}}{\sum_{j=1}^{k}\frac{1}{\hat{\sigma}_j^2}} \in (0,1)\,. \tag{3.10}$$

Sometimes the efficient frontier shows a specularity at the origin of the graph with respect to the normal shape it usually takes on, determining returns characterized by negative variances.

Having negative variances, however, is clearly a nonsense both from a mathematical-statistical point of view and from a financial one. These hold true respectively because the variance is a sum of squared distances so it is always greater or equal to zero, and because we cannot have portfolios of assets characterized by negative risk.

Corazza and Pizzi (2014) found that this anomalous behavior of the unbiased efficient frontier à la Bodnar and Bodnar appears especially in case of daily logarithmic returns, and when we are dealing with a number of observations n which is big. They, however, do not exclude that this behavior can show up also in case of monthly linear returns.

3.3.1 Some further considerations

Corazza and Pizzi (2014), derive the results in Proposition 1 by considering only the sensitivity of the unbiased efficient frontier à la Bodnar and Bodnar with respect to the number of observations, n.

However, by taking into consideration the formula of the unbiased estimator proposed by Bodnar and Bodnar as it is written in Equation 3.2, it is straightforward to see that, as we will analyze more

in depth later on, there is another fundamental quantity which as a central role in determining the shape of such an estimator, which is the slope of the unbiased efficient frontier, $\hat{s}$. The slope of the unbiased efficient frontier à la Bodnar and Bodnar is, in turn, a function of $\hat{\mu}_i$ and $\hat{\sigma}_i$.

That is the reason why we will conduct a sensitivity analysis[2] of the unbiased efficient frontier with respect to its slope, and not with respect to the number of observations.

In order to do this, we will apply differential sensitivity analysis so that to check the different impact of $\hat{\mu}_i$ and of $\hat{\sigma}_i$ on the results of the unbiased estimator by Bodnar and Bodnar.

3.3.1.1 $N = 2$ case

We start our analysis by considering the case in which $N = 2$. In this case, starting from Equation 3.9 we end up with the following Equation for $\hat{s}$:

$$\hat{s} = \left[\frac{\hat{\mu}_1^2}{\hat{\sigma}_1^2}(1-\hat{p}_1) + \frac{\hat{\mu}_2^2}{\hat{\sigma}_2^2}(1-\hat{p}_2)\right] - \left(\frac{\hat{\mu}_1\hat{\mu}_2}{\hat{\sigma}_2^2}\hat{p}^2 + \frac{\hat{\mu}_2\hat{\mu}_1}{\hat{\sigma}_1^2}\hat{p}_1\right). \tag{3.11}$$

From Equation 3.11 we start by differentiating with respect to $\hat{\mu}_1$. What we obtain is reported below:

$$\frac{\partial \hat{s}}{\partial \hat{\mu}_1} = \frac{2\hat{\mu}_1}{\hat{\sigma}_1^2}(1-\hat{p}_1) - \left(\frac{\hat{\mu}_2}{\hat{\sigma}_2^2}\hat{p}_2 + \frac{\hat{\mu}_2}{\hat{\sigma}_1^2}\hat{p}_1\right). \tag{3.12}$$

By imposing to the result obtained in Equation 3.12 the constraint of being greater than 0 and rearranging it, we come up with the following inequality:

$$\hat{\mu}_1\left[\frac{2}{\hat{\sigma}_1^2}(1-\hat{p}_1)\right] > \hat{\mu}_2\left(\frac{\hat{p}_2}{\hat{\sigma}_2^2} + \frac{\hat{p}_1}{\hat{\sigma}_1^2}\right) \tag{3.13}$$

$$\text{where} \begin{cases} \hat{p}_1 = \frac{\frac{1}{\hat{\sigma}_1^2}}{\frac{1}{\hat{\sigma}_1^2}+\frac{1}{\hat{\sigma}_2^2}} \\ \hat{p}_2 = \frac{\frac{1}{\hat{\sigma}_2^2}}{\frac{1}{\hat{\sigma}_1^2}+\frac{1}{\hat{\sigma}_2^2}}. \end{cases} \tag{3.14}$$

Even if in Equation 3.13 we know that:

- the variances $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are quantities which are always positive,
- the quantities $\hat{p}_1$ and $\hat{p}_2$ are always positive in that they are the ratio of the reciprocals of two positive quantities,

we are not able to determine the sign of the Equation 3.13.

We do not differentiate the Equation 3.11 with respect to $\hat{\sigma}_1$ because it is already clear from the fact that $\hat{p}_1$ and $\hat{p}_2$ in Equation 3.14 depend on both $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ that we will not be able to compute the sign of the differentiated equation.

What we decide to do is, instead, trying to find the general version of Equation 3.13.

[2]For a more detailed explanation of sensitivity analysis see Appendix C.

3.3.1.2 $N = 3$ case

In order to derive the generalized version of Equation 3.13 we must see what happens in the case of $N = 3$ and see if there is some recursive behavior in the equations that can suggest a clear pattern for the general case $N > 2$.

Starting from the general Equation 3.9 of the slope of the unbiased efficient frontier à la Bodnar and Bodnar, we are able to come up to the following Equation in the case in which we consider 3 securities:

$$\begin{aligned}\hat{s} = &\left[\frac{\hat{\mu}_1^2}{\hat{\sigma}_1^2}(1-\hat{p}_1) + \frac{\hat{\mu}_2^2}{\hat{\sigma}_2^2}(1-\hat{p}_2) + \frac{\hat{\mu}_3^2}{\hat{\sigma}_3^2}(1-\hat{p}_3)\right] \\ &- \left(\frac{\hat{\mu}_1\hat{\mu}_2}{\hat{\sigma}_2^2}\hat{p}_2 + \frac{\hat{\mu}_1\hat{\mu}_3}{\hat{\sigma}_3^2}\hat{p}_3 + \frac{\hat{\mu}_2\hat{\mu}_1}{\hat{\sigma}_1^2}\hat{p}_1 + \frac{\hat{\mu}_2\hat{\mu}_3}{\hat{\sigma}_3^2}\hat{p}_3 + \frac{\hat{\mu}_3\hat{\mu}_1}{\hat{\sigma}_1^2}\hat{p}_1 + \frac{\hat{\mu}_3\hat{\mu}_2}{\hat{\sigma}_2^2}\hat{p}_2\right).\end{aligned} \tag{3.15}$$

By differentiating Equation 3.15 , which represents $\hat{s}$ in the case of $N = 3$, with respect to $\hat{\mu}_1$ we get the following Equation:

$$\frac{\partial \hat{s}}{\partial \hat{\mu}_1} = \frac{2\hat{\mu}_1}{\hat{\sigma}_1^2}(1-\hat{p}_1) - \left(\frac{\hat{\mu}_2}{\hat{\sigma}_2^2}\hat{p}_2 + \frac{\hat{\mu}_3}{\hat{\sigma}_3^2}\hat{p}_3 + \frac{\hat{\mu}_2}{\hat{\sigma}_1^2}\hat{p}_1 + \frac{\hat{\mu}_3}{\hat{\sigma}_1^2}\hat{p}_1\right). \tag{3.16}$$

By rearranging Equation 3.16 and imposing that it should be greater than 0, what we obtain is reported in the Equation below:

$$\hat{\mu}_1\left[\frac{2}{\hat{\sigma}_1^2}(1-\hat{p}_1)\right] > \hat{\mu}_2\left(\frac{\hat{p}_2}{\hat{\sigma}_2^2} + \frac{\hat{p}_1}{\hat{\sigma}_1^2}\right) + \hat{\mu}_3\left(\frac{\hat{p}_3}{\hat{\sigma}_3^2} + \frac{\hat{p}_1}{\hat{\sigma}_1^2}\right). \tag{3.17}$$

Again we have that, even if we know that $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2, \hat{p}_1, \hat{p}_2, \hat{p}_3$ are quantities which are always greater than 0, we are not able to determine the overall sign of Equation 3.17.

3.3.1.3 $N > 2$ general case

What we can do is, instead, deriving the Equation for the general case in which $N > 2$. As a matter of fact, from Equation 3.13 and Equation 3.17 we can derive the following:

$$\hat{\mu}_1\left[\frac{2}{\hat{\sigma}_1^2}(1-\hat{p}_1)\right] > \sum_{i=2}^{k}\left(\frac{\hat{p}_i}{\hat{\sigma}_i^2} + \frac{\hat{p}_1}{\hat{\sigma}_1^2}\right). \tag{3.18}$$

Even in the general case in which $N > 2$, we cannot determine a priori the sign of Equation 3.18, meaning that

3.4 Description of the dataset

As stated above, we will proceed by checking for operational effectiveness of the unbiased estimator à la Bodnar and Bodnar. We will study its behavior and we will compare it with the sample efficient

frontier. We will also study the hypothesis of normality and independence, which are common to both the frontiers.

In order to perform our analysis we will use real data. We chose a basket of ten different stocks from the FTSE MIB. The FTSE MIB is the benchmark stock market index for Borsa Italiana and it is made up by the 40 most-traded stock classes on the exchange.

As of August 2016 its components are those listed in Table 3.1.

Table 3.1: Constituents of FTSE MIB, August 2016.

	NAME		NAME
1	A2a	21	Leonardo - Finmeccanica
2	Atlantia	22	Luxottica
3	Azimut Holding	23	Mediaset
4	Banca Mediolanum	24	Mediobanca
5	Banca Monte Paschi Siena	25	Moncler
6	Banca Popolare Emilia Romagna	26	Poste Italiane
7	Banca Popolare Milano	27	Prysmian
8	Banco Popolare	28	Recordati
9	Buzzi Unicem	29	Saipem
10	Campari	30	Salvatore Ferragamo
11	Cnh Industrial	31	Snam
12	Enel	32	Stmicroelectronics
13	Eni	33	Telecom Italia
14	Exor	34	Tenaris
15	Ferrari	35	Terna - Rete Elettrica Nazionale
16	Fiat Chrysler Automobiles	36	Ubi Banca
17	Finecobank	37	Unicredit
18	Generali	38	Unipol
19	Intesa Sanpaolo	39	Unipol Sai
20	Italcementi	40	Yoox Net-a-Porter Group

Source: http://www.borsaitaliana.it/borsa/azioni/ftse-mib/lista.html

In order to select from these forty available stocks the ten to be used in our analysis, we had to take into account two important criteria.The first is that each stock chosen must be quoted on the Italian stock exchange at least since January 2000, since we have decided to consider as the one ranging from January 2000 to June 2016. The second aspect we decide to take into consideration for this choice is that of making a pool of asset the most representative as possible, and in order to do this we took ten different stocks from ten different ICB[3] sectors. What we have obtained is listed in Table 3.2.

[3]The Industrial Classification Benchmark is an industry classification taxonomy used to segregate markets into sectors within the macroeconomy. The ICB is used globally (though not universally) to divide the market into increasingly specific categories, allowing investors to compare industry trends between well-defined subsectors.

Table 3.2: Stocks considered in the analysis.

	NAME	BLOOMBERG	ICB SECTOR
1	A2A	A2A IM	Utilities
2	Banca Monte Paschi Siena	BMPS IM	Banks
3	Buzzi Unicem	BZU IM	Construction & Materials
4	Exor	EXO IM	Financial Services
5	Fiat Chrysler Automobiles	FCA IM	Automobiles & Parts
6	Generali	G IM	Insurance
7	Mediaset	MS IM	Media
8	Saipem	SPM IM	Oil & Gas
9	Stmicroelectronics	STM IM	Technology
10	Telecom Italia	TIT IM	Telecommunications

Source: http://www.borsaitaliana.it

As already we have already pointed out above, we have decided to take into account a long time series of asset returns. This choice has been driven from the fact that we will need a significant number of observations even when dealing with the dataset of monthly asset returns. As a matter of fact, we come up with two datasets: the monthly asset returns one which is made up of 197 observations, and the daily asset returns one, which is instead made up of 4185 observations, for all the 10 stocks considered in our analysis.

These time series have been downloaded from the Bloomberg terminal, in the form of daily closing prices and monthly average closing prices.

Since in order to perform our analysis, we will need returns of the assets, and not closing prices, we chose to use logarithmic returns, which have been calculated as follows:

$$r_{ln} = ln\left(P_t\right) - ln\left(P_{t-1}\right) = ln\left(\frac{P_t}{P_{t-1}}\right). \tag{3.19}$$

Finally, the data was downloaded in spreadsheet form and imported to the language and environment for statistical computing and graphics, R.

3.5 Empirical analysis of the dataset

Since, as we have extensively pointed out, the main hypothesis on which the sample classical efficient frontier as well as the unbiased estimator à la Bodnar and Bodnar are based on, are those of multivariate normality and independence of returns, we will investigate for their presence in our dataset. We will then implement the model, by applying it to our datasets. We will finally check for the robustness of the estimator.

3.5.1 Normality

We have already stated that whether financial asset returns should be deemed to be normally distributed is a very debated argument. The first to contribute to this debate was Bachelier (1900) which stated that financial returns are normally distributed. Then, we have Mandelbrot (1963)

which pointed out that asset returns are not distributed normally, but are somehow leptokurtic[4]. But at last, Fama (1976) showed that the normal assumption is appropriate for describing returns of monthly data. Of course, as we have already said throughout this work, this is a very debated argument and the list of authors cited above is not meant to be exhaustive.

We will then try to grasp some other insight on this complicated matter, and we will analyze normality both for our monthly returns and daily returns.

In order to check for the normality of our returns, here denoted by $x_i = x_1, x_2, ..., x_n$, we will use the Shapiro-Wilk test of normality. It is a test of normality based on the correlation between sample quantiles and theoretical (normal) quantiles. Its test statistic is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

where

- $x_{(i)}$ is the i^{th} order statistic, the i^{th}-smallest value of the sample;
- $\bar{x} = (x_1 + ... + x_n)/n$ is the sample mean;
- the constant a is given by $(a_1, ..., a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$,
  where $m = (m_1, ..., m_n)^T$ and $(m_1, ..., m_n)$ is the vector of the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics.

The null hypothesis of this test is that the population is normally distributed. Thus, if we obtain a p-value which is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population; in other words, the data are not normal. On the contrary, if the p-value is greater than the chosen alpha level, then the null hypothesis that the data came from a normally distributed population cannot be rejected.

In our case, we chose a significance level of $\alpha = 0.05$.

We start by analyzing our monthly returns. By applying to this first dataset the Shapiro-Wilk test we obtain the results in Table 3.3.

Table 3.3: Shapiro-Wilk normality test of monthly returns.

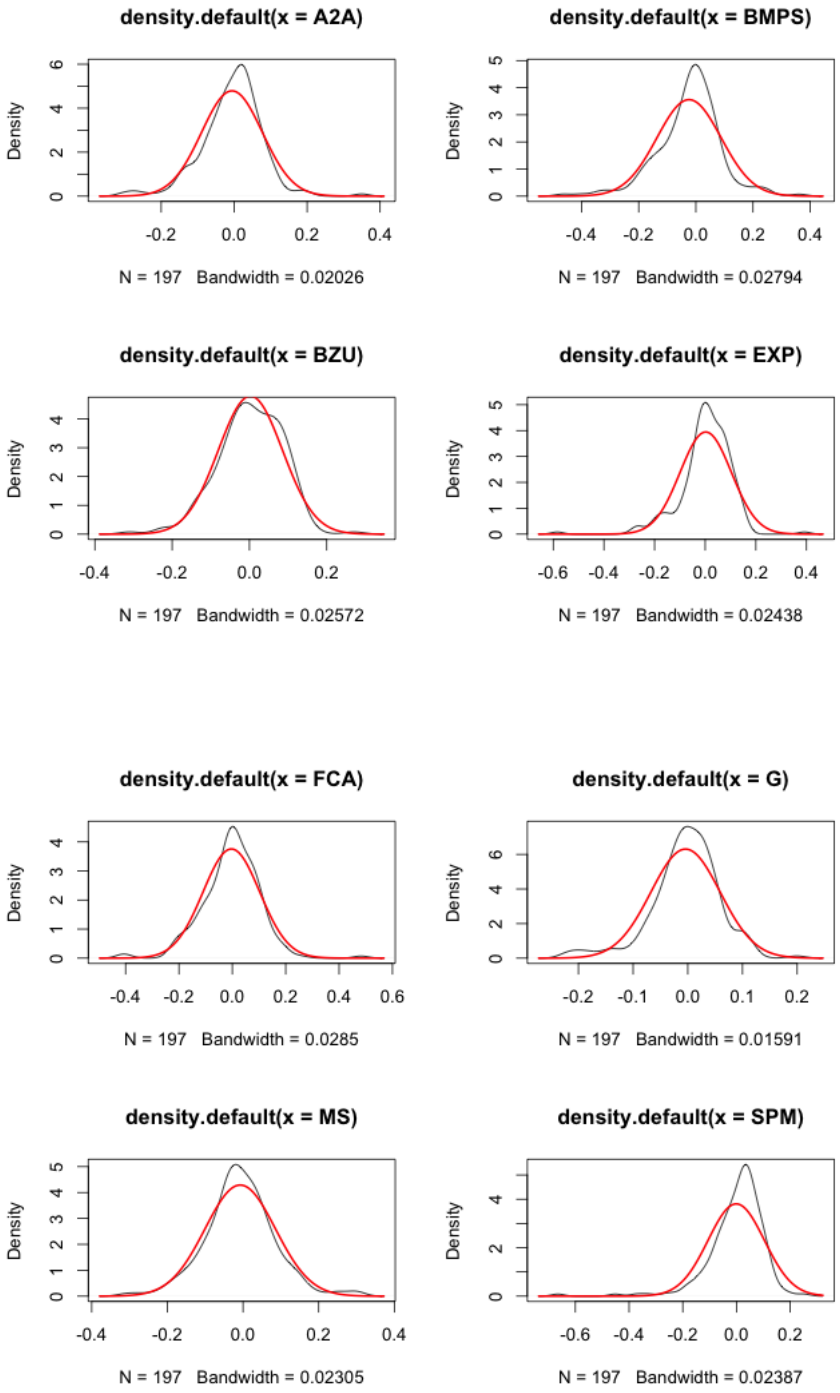
	STOCK	p-value
1	A2A	1.131e-05
2	Banca Monte Paschi Siena	1.17e-05
3	Buzzi Unicem	0.01638
4	Exor	8.15e-10
5	Fiat Chrysler Automobiles	2.485e-05
6	Generali	3.21e-06
7	Mediaset	0.01429
8	Saipem	9.256e-13
9	Stmicroelectronics	0.004498
10	Telecom Italia	9.772e-08

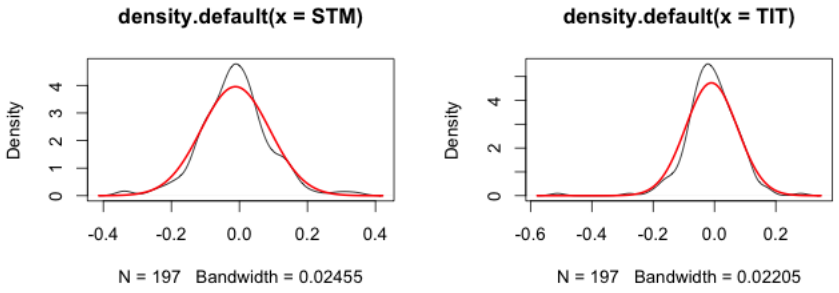
[4]A distribution is leptokurtic when the points along the X-axis are clustered, resulting in a higher peak, or more precisely in higher kurtosis, than the normal curvature characterizing a normal distribution. A leptokurtic distribution is also characterized by fat tails.

It is easy to see that since the p-value of each stock is lower, more or less consistently, then the chosen significance level for each of our ten stocks, we can conclude that this dataset is not normally distributed.

We can come up with the very same conclusion, as those shown in Table 3.3, also by taking a look at the graphical representation of the distribution of each stock against the normal distribution, as plotted in Figure 3.1.

Figure 3.1: Plot of each stock distribution against the normal distribution, monthly returns.





Here, we have plotted each density function of our stocks, against the normal density distribution in order to check if the two are reasonably close or not, if they exhibit skewness or kurtosis, namely if they are asymmetric of if they are characterized by fat tails.

In order to read this graphs, we must take into account that the red curve represents the normal density distribution, while the superimposed black one represents the Kernel density distribution referred to each particular asset return we have considered in our analysis. Then we have N, which represents the number of observations which make up our dataset, and bandwidth which determines the resolution and the smoothness of the estimator. The choice of the bandwidth is related to the characteristics of our data; more precisely it depends on the sample size and on the true density. In R we can use the command adjust to tune the value of the bandwidth, which is otherwise automatically computed by the software itself, in order to get more or less smoothed estimations.

In the majority of the cases, the densities of our asset returns are sharper in the center than the normal density, only BZU can be deemed to not show this peculiarity. At the same time, also skewness is a common feature in the above-plotted densities.

All of this information together clearly suggest non-normality.

In Table 3.4 we report the results of the Shapiro-Wilk test of normality computed in the case of daily logarithmic returns. Since we obtain the same infinitesimally small result for all of our asset returns, it is again clear that we have absence of normality in the distribution of our dataset.

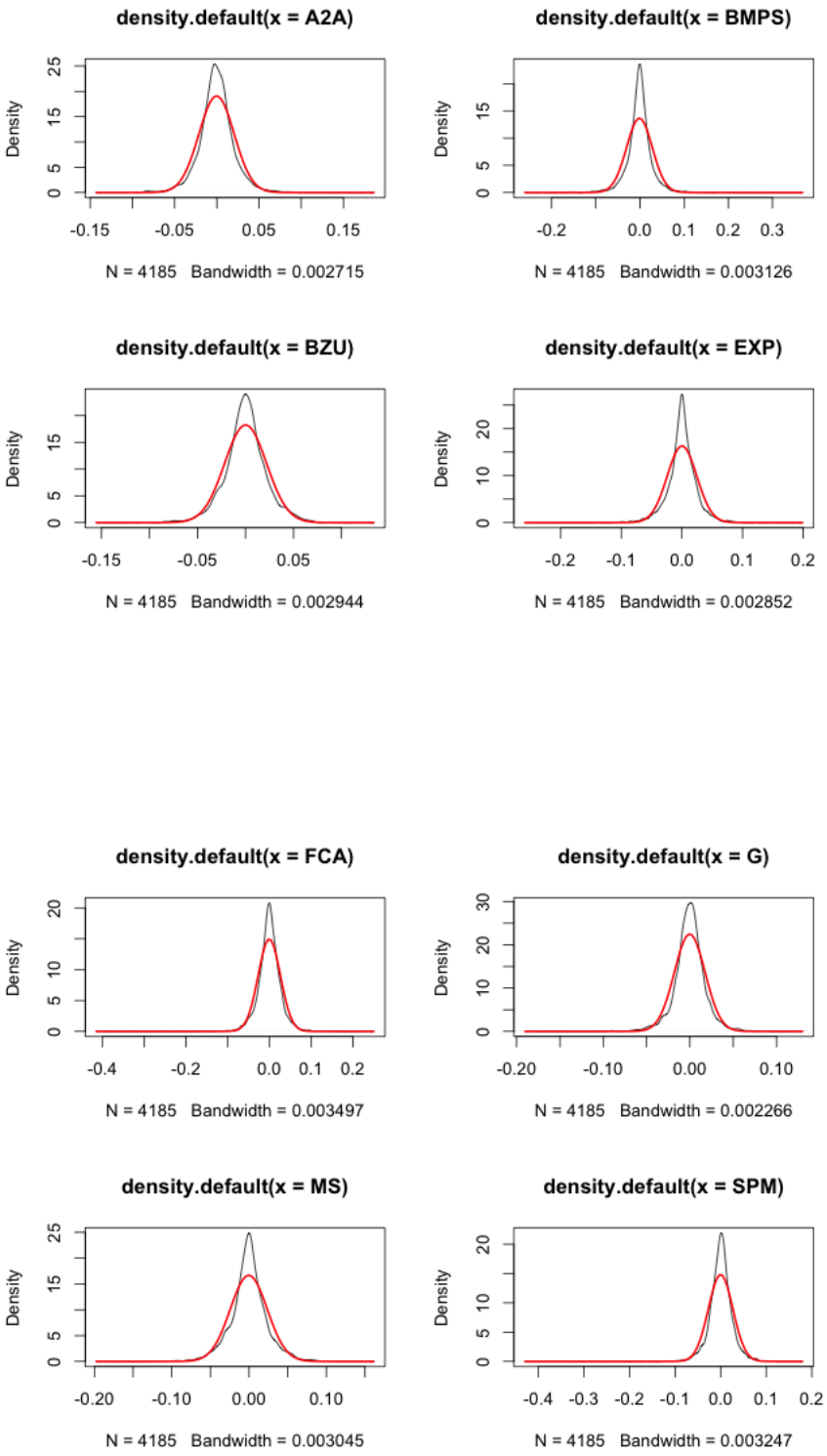
Table 3.4: Shapiro-Wilk test of normality of daily returns.

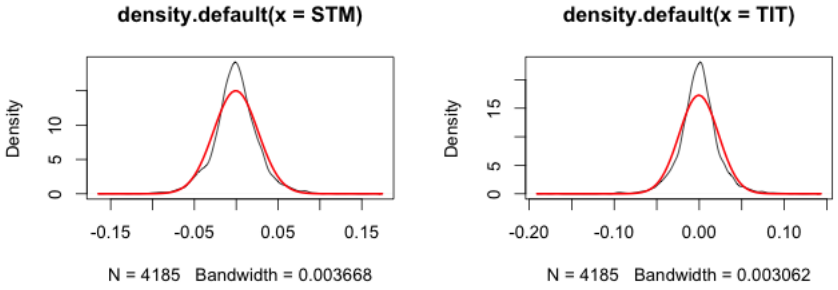
	STOCK	p-value
1	A2A	$<$ 2.2e-16
2	Banca Monte Paschi Siena	$<$ 2.2e-16
3	Buzzi Unicem	$<$ 2.2e-16
4	Exor	$<$ 2.2e-16
5	Fiat Chrysler Automobiles	$<$ 2.2e-16
6	Generali	$<$ 2.2e-16
7	Mediaset	$<$ 2.2e-16
8	Saipem	$<$ 2.2e-16
9	Stmicroelectronics	$<$ 2.2e-16
10	Telecom Italia	$<$ 2.2e-16

If we compute, as in the case of monthly logarithmic returns, the graphical plots of these daily returns we get the results in Figure 3.2. Again the red curve and the black one represent respectively

the normal density distribution and the distribution of each specific dataset, N is the number of observations, and the bandwidth is automatically chosen by our software and represents the resolution and smoothness of the estimator of the density distribution.

Figure 3.2: Plot of each stock distribution against the normal distribution, daily returns.





The plots in Figure 3.2 exhibit the same non-normal behavior witnessed before, when dealing with monthly returns, in that all of them exhibit a sharper peak in the center and all are skewed.

Both in the case of monthly returns and in the case of daily ones, we obtain the very same results also by applying the Jarque-Bera test of normality, which is a test based on the comparison between the sample cumulative distribution function and the theoretical (normal) cumulative distribution function. There is no clue according to which we can state that our returns are somehow normal.

For seek of completeness, we decide to implement another test, called the D'Agostino test of normality. Again, the null hypothesis of this test is that the sample is distributed following a normal distribution, we thus reject the hypothesis of normality when the p-value of the test is below our significance level, in this specific case below $\alpha = 5\%$.

By applying this D'Agostino test of normality, we can investigate more precisely whether returns are not normal but always why they depart from normality; we can check in which specific case non-normality is due to skewness, to kurtosis, or to both of them.

For what concerns monthly returns, we can see from Table 3.5 that since the p-value of observed skewness is greater than our significance level α for the returns of A2A, FCA, MS, and STM, these do not exhibit such departure from normality. They only exhibit kurtosis. For what concerns, instead, the values for kurtosis, the only p-value which is above α is the one for BZU, as a matter of fact we also see this feature of its distribution from the plot in Figure 3.1.

Table 3.5: D'Agostino test of normality, monthly returns.

	p-value Normality	p-value Skewness	p-value Kurtosis
A2A	4.978e-05	0.07231	4.649e-05
BMPS	2.101e-05	0.005573	0.0001972
BZU	0.008028	0.01533	0.05212
EXP	6.661e-16	2.333e-09	4.69e-09
FCA	4.641e-05	0.3401	1.276e-05
G	9.582e-07	7.889e-05	0.0004957
MS	0.01378	0.3525	0.005507
SPM	$<$ 2.2e-16	$<$ 2.2e-16	1.396e-11
STM	0.006422	0.4208	0.002114
TIT	7.427e-14	4.161e-07	3.575e-09

On the other hand, when considering daily returns we can see from Table 3.6 we have that all of them clearly show kurtosis, since all the p-values for the skewness are infinitesimally small, and

thus are well below the chosen significance level. Instead, the distribution of BMPS and STM do not seem to show any degree of skewness, due to their p-values above the chosen significance level.

Table 3.6: D'Agostino test of normality, daily returns.

	p-value Normality	p-value Skewness	p-value Kurtosis
A2A	$<$ 2.2e-16	1.262e-05	$<$ 2.2e-16
BMPS	$<$ 2.2e-16	0.2863	$<$ 2.2e-16
BZU	$<$ 2.2e-16	0.01094	$<$ 2.2e-16
EXP	$<$ 2.2e-16	5.269e-08	$<$ 2.2e-16
FCA	$<$ 2.2e-16	$<$ 2.2e-16	$<$ 2.2e-16
G	$<$ 2.2e-16	1.332e-15	$<$ 2.2e-16
MS	$<$ 2.2e-16	0.01909	$<$ 2.2e-16
SPM	$<$ 2.2e-16	$<$ 2.2e-16	$<$ 2.2e-16
STM	$<$ 2.2e-16	0.5742	$<$ 2.2e-16
TIT	$<$ 2.2e-16	6.994e-10	$<$ 2.2e-16

3.5.2 Independence

According to what we have extensively stated throughout this paper, the second fundamental pillar on which all the theories concerning estimators of the efficient frontier are based on is the independence between asset returns.

In order to check for independence of our returns, we can compare their covariance matrices with the matrix in Equation 3.4.

We have computed the covariance matrix of both of our datasets. We do not report here the results, but what we have obtained are two covariance matrices which clearly do not show all the off-diagonal elements equal to zero; actually, none of the covariances was equal to zero nor in the case of monthly or in the case of daily returns. Nor the covariance matrix of monthly or the covariance matrix of daily returns do resemble the empirical one, stated in Equation 3.4.

Even if the results appear quite clear, we perform a test in order to check the significance of the degree of departure of these matrices from the hypothesis of independence. We perform a linear test of independence by using the covariance test, `cov.test`, present in R software. This test has as null hypothesis, H_0, according to which the true correlation between two data is equal to zero, while the alternative, H_1, is that true correlation between the data is not equal to zero. This test only checks for linear dependence, as we have stated above since it tests the relation between a pair of time series at a time. Since by computing this matrix, all the p-value are well below the significance level $\alpha = 0.05$ there is no evidence that our data can be deemed to be linearly independent.

We perform the same test for daily returns, and again all the p-values suggest that we have to reject the hypothesis of independence.

Above, we have checked for linear dependence. But in a multivariate time series setting, we know that we may also witness non-linear dependence. Non-linear dependence is not relevant in the framework of the unbiased efficient estimator proposed by Bodnar and Bodnar, but we propose to use as a proxy of this linear dependence the Granger causality test[5].

[5]For more details on Granger causality test and for its results see Appendix D.

3.5.3 Robustness

In order to conclude our analysis on the classical sample efficient frontier we can analyze how a change, more precisely, an increment, of the number of the stocks considered or of the length of our time series, has an impact on the definition of our sample efficient frontier.

The increase in the number of stocks considered, causes a consequent decrease in the overall level of diversifiable, or systematic, risk of our portfolio. This should theoretically lead to a frontier characterized by a better risk-return profile.

For what concerns, instead, the length of the time series it is important to underline that since our dataset is a sample, its accuracy will be positively related to the number of observations considered. The greater the number of observations in the sample, the lower the sampling error.

In order to check the behavior listed above we decide to divide our dataset as follows:

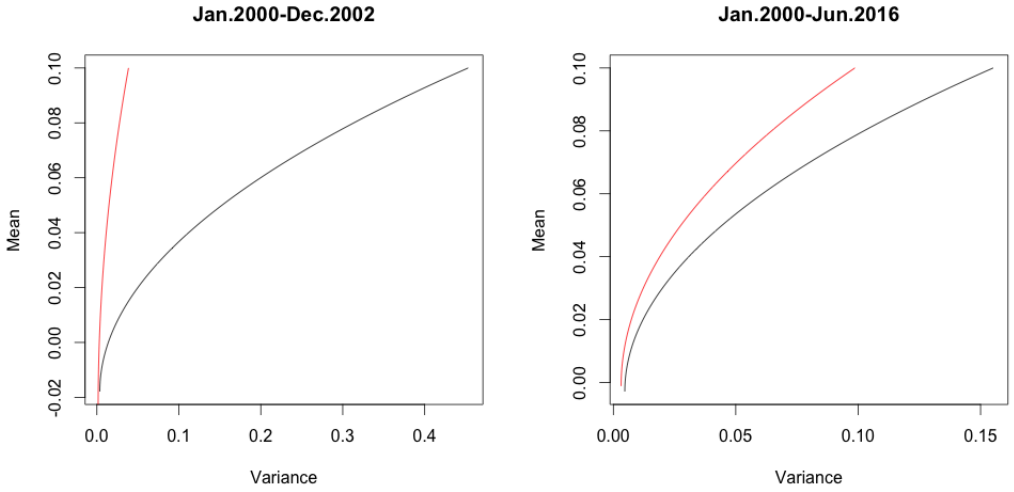
- firstly we analyze a restricted set of our original dataset, namely out of our ten initial stocks A2A, Banca Monte Paschi Siena, Buzzi Unicem, Exor, Fiat Chrysler Automobile, Generali, Mediaset, Saipem, Stmicroelectronics, and Telecom Italia, we consider a smaller sample made up of the first 3 stocks, so we will have a smaller sample with the following stocks: A2A, Banca Monte Paschi Siena, and Buzzi Unicem;

- then we decide to divide our daily returns in two different time spans: a first in which we consider only a portion of our entire dataset, ranging from January 01, 2000 to December 31, 2002. The second will instead be composed of the whole time period considered at the beginning of the chapter, ranging namely from January 01, 2000 to June 30, 2016.

In order to analyze how the number of stocks we take into account and the length of the time series affect our analysis, we plot the length of the time series and the number of stocks considered separately.

In Figure 3.3 we can see that we have the same outcome in the case in which we consider a shorter time period, namely the two-year period, or the whole period, spanning from January 2000 to June 2016. As a matter of fact, in each of the two time periods considered, when the number of stocks increases passing from the black line which represents the 3 stocks scenario to the red line, which represents the 10 stocks scenario, the efficient frontier increases. If we can see that when we consider a longer period, namely a greater number of observations, the span between the sample efficient frontier and the unbiased estimator proposed by Bodnar and Bodnar decreases.

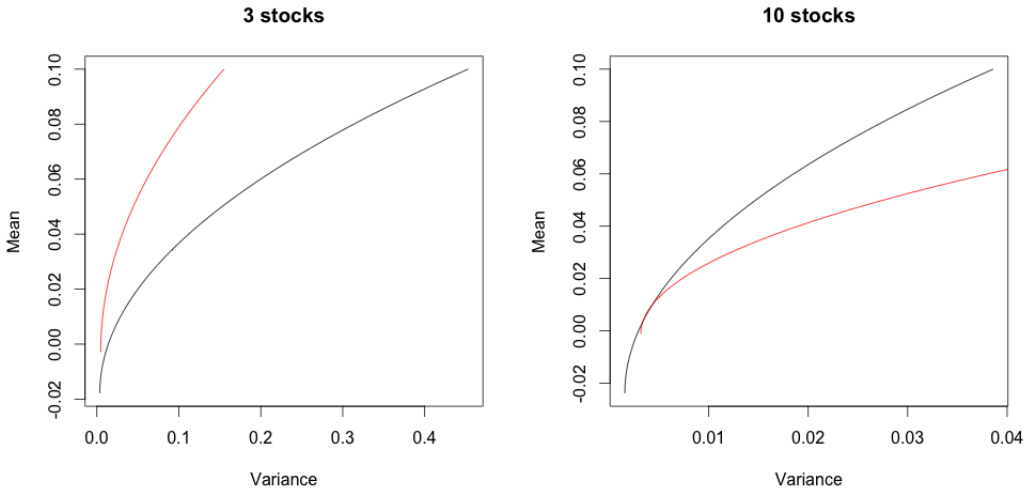
Figure 3.3: Sample efficient frontiers compared according to sample length (Montlhy returns).



From Figure 3.4, instead, we can see that the case in which we have 3 stocks behaves differently from the case in which we have ten stocks. Namely, in the first case we have that when we pass from having 2 years of monthly observations to having 10 years of monthly observations the sample efficient frontier increases. In the case of 10 stocks, instead, we pass from the black line, which represents the 2 years of observations case, to the red line, the 10 years of observations case, where we register a decrease in the sample efficient frontier. The results are coherent with Proposition 1 in Corazza and Pizzi (2014).

This non-univocal behavior, is probably due to the fact that the sample efficient frontier in the case of 10 stocks over the whole time period is obtained by taking into consideration a bigger number of data, thus representing a better estimate of the efficient frontier because, ceteris paribus, characterized by a lower sampling error.

Figure 3.4: Sample efficient frontiers compared according to the number of stocks considered (Monthly returns).



In the case of daily returns, plotted in Figure 3.5 and in Figure 3.6, we have the very same behavior of the sample efficient frontiers, and we can thus obtain the same conclusions.

Figure 3.5: Sample efficient frontiers compared according to sample length (Daily returns).

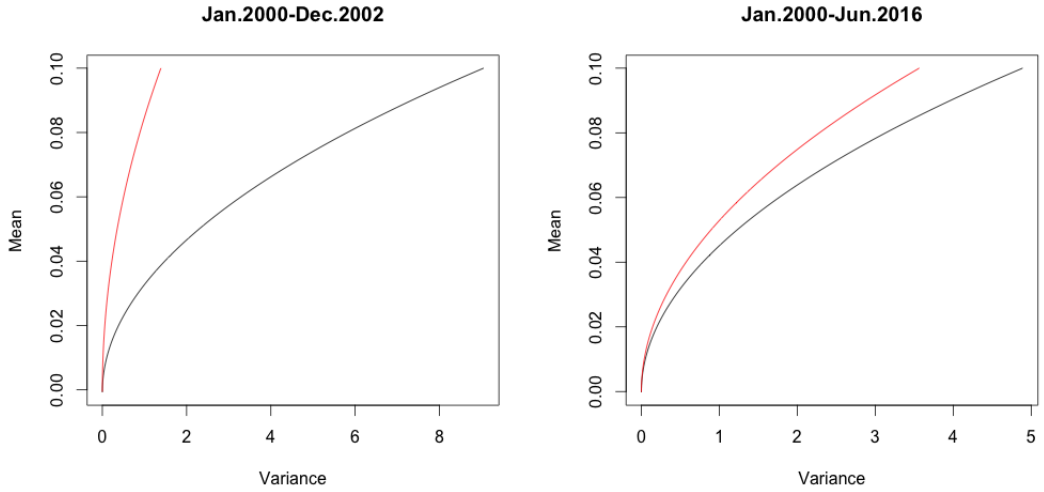
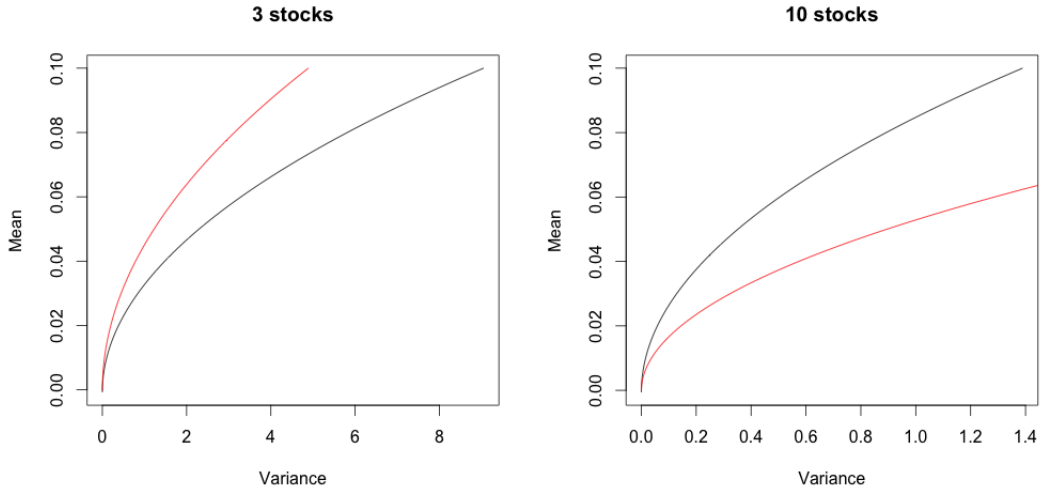


Figure 3.6: Sample efficient frontiers compared according to number of stocks considered (Daily returns).



3.6 Check of the model's hypothesis

By recalling the assumptions on which the unbiased efficient frontier à la Bodnar and Bodnar is based on, we have that in order for the model to work we need:

1. k assets, with $k > 2$,

2. n observations, with $n > k$,

3. normally and independently distributed log-returns.

The first two hypothesis, namely that the number of stocks is greater than two ($k > 2$) and that the length of the time series is greater than the number of stocks ($n > k$), are easy to be obtained in our dataset and are coherent with the real world. The problems come out from the third hypothesis, the one according to which our returns should be multivariate normally distributed and independent.

As we have already checked in the previous sections, when analyzing the hypothesis of the classical formulation of the sample efficient frontier, we have seen that in the case of our specific dataset our returns are not normal, both in the case of daily returns and also in the case of monthly returns.

We have already analyzed in the previous sections also the hypothesis of independence of returns. From the covariance matrix, both in the case of daily and in the case of monthly returns, we can see that we do have correlation between our returns, meaning that we cannot state independence of our data.

Thus, as we have stated before, the main assumptions on which the model, is based do not apply in our real datasets.

Chapter 4

Empirical analysis of the unbiased estimator

4.1 Problems in the application of the unbiased estimator

As we have extensively checked throughout Chapter 3, the main hypothesis on which the unbiased estimator proposed by Bodnar and Bodnar is based on, are difficult to be found in reality. As expected, we have not found evidence of independence either in monthly returns or in daily returns. The same holds true even for the hypothesis of normality. As a matter of fact, we did not found normality either in daily returns or in monthly returns, despite the previsions by Fama (1976) according to which the fat tails which usually characterize financial returns should not have a great impact at least in case of monthly returns.

We thus start our analysis by showing how the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar works in the case in which all its main assumptions are not violated.

We will then try to test its robustness by checking its behavior in the case in which we include very small covariances between our returns, until we come to the extreme case in which we take as our dataset data which clearly exhibit dependence.

4.2 Unbiased efficient frontier obtained on independent monthly returns

The first scenario we consider in this analysis, as we have stated above, is the one in which we account for a completely independent set of asset returns.

For ease of future comparability of all of our results, instead of computing our independent dataset from scratch using resampling techniques, we have decided to directly modify our original dataset which, has we have checked in Chapter 3 was evidently characterized by dependence between returns, in order to obtain the conditions required by the model of Bodnar and Bodnar.

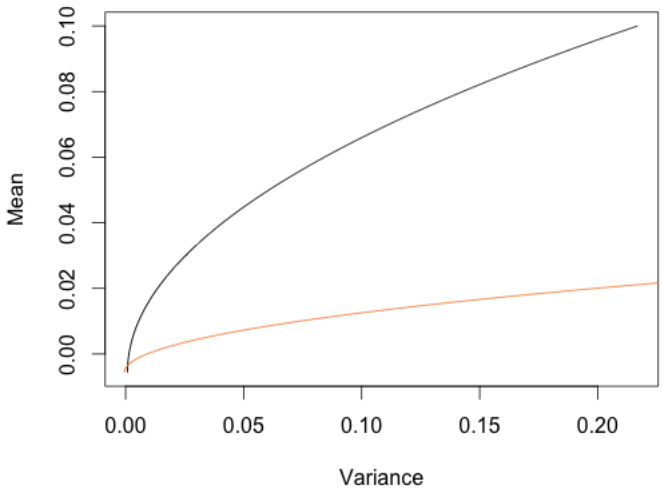
We thus decide to use our original dataset, and in order to make it resemble a portfolio characterized by the covariance matrix in Equation 3.4 we assume as appropriate the vector of the means, at

the same time we consider only the diagonal of variances in the covariance matrix of our dataset leaving them untouched, while forcing equal to zero all the off-diagonal elements. In this way we are considering equal to zero all the covariances between asset returns, meaning that we are implying absence of mutual dependence between them.

4.2.1 Sample efficient frontier versus unbiased efficient frontier à la Bodnar

By applying the results of Bodnar and Bodnar (2010) we are able to plot, by using the code reported in Appendix E the unbiased efficient estimator of Bodnar and Bodnar together with the classical sample efficient frontier, and we obtain the graph represented in Figure 4.1.

Figure 4.1: Sample efficient frontier versus unbiased efficient frontier à la Bodnar and Bodnar, independent monthly returns.



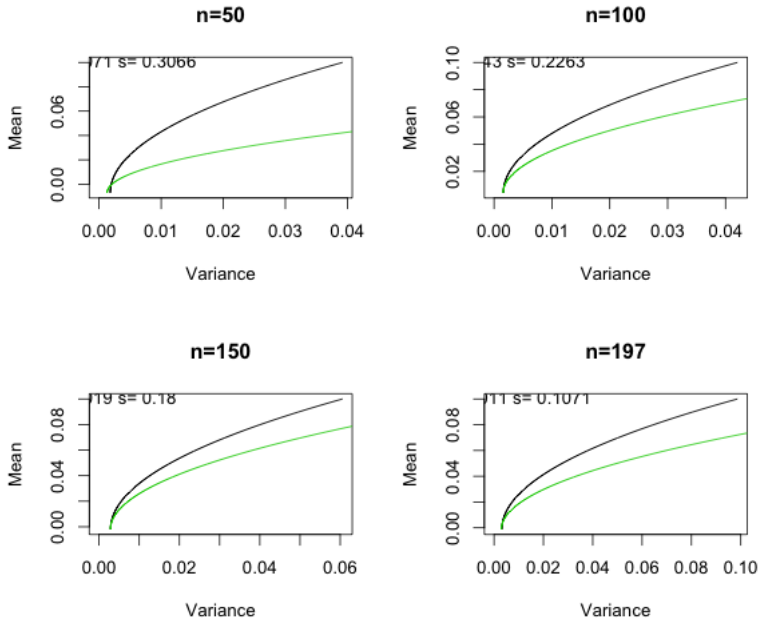
We highlight that the black curve represents the classical sample efficient frontier, while the orange one represents the unbiased estimator proposed by Bodnar and Bodnar. Since it is evident from the picture that they are pretty far apart one from the other, we can say that the classical sample efficient frontier seems to be an overoptimistic estimator, providing higher return for any given level of risk, and the unbiased efficient estimator of the efficient frontier proposed by Bodnar and Bodnar seems to be able to correct for this bias.

4.2.2 Asymptotic behavior of the unbiased efficient frontier à la Bodnar and Bodnar

In order to check how Equation 3.5 in Proposition 1 of Corazza and Pizzi (2014) applies in the case of independent dataset, we split our observation into 4 different subset with $n \in \{50, 100, 150, 197\}$. We obtain the graph in Figure 4.2. The curve in black represents the classical sample efficient

frontier, while the one in green represents the unbiased efficient frontier à la Bodnar and Bodnar for an increasing number of observations, n.

Figure 4.2: Asymptotic behavior of the unbiased efficient frontier, monthly independent returns.



The span between the black and the green line diminishes as n increases. The results in Proposition 1 are thus in accordance with the empirical results we obtained through our dataset. Since in the plots for $n = 150$ and $n = 197$ the values on the $x-$axis increase, it initially appears that they do not follow the asymptotic behavior predicted in Proposition 1. By looking more closely and taking into consideration the changes in the values of the $x-$ and $y-$axis we can immediately understand that the behavior is consistent with Proposition 1.

4.2.3 Operational effectiveness of the unbiased efficient frontier à la Bodnar and Bodnar

At last, we can check the operational effectiveness of this unbiased estimator proposed by Bodnar and Bodnar.

We should check whether the implementation of this new estimator has some advantage if compared to the classical sample efficient frontier. In order to do so, we have to test whether the unbiased estimator proposed by Bodnar and Bodnar falls inside the confidence interval constructed on the the classical sample efficient frontier.

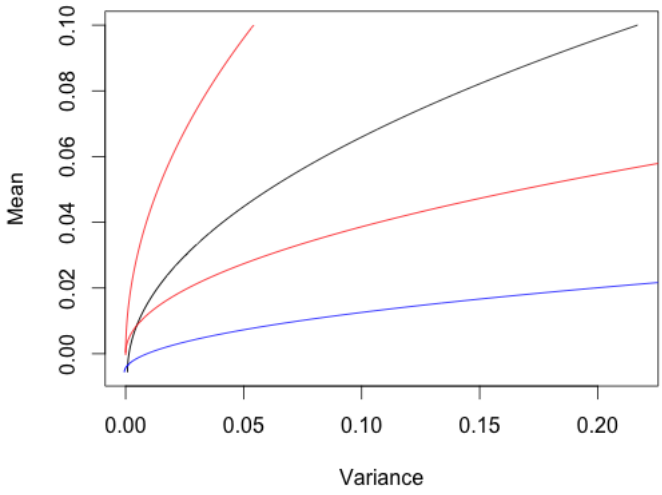
This confidence interval calculated at a significance level $\alpha = 5\%$ is obtained using bootstrapping technique[1].

By pointing out that the blue line represents the unbiased efficient frontier à la Bodnar and Bodnar, the black line represents the classical sample efficient frontier, and the two red curves represent the

[1]See Appendix F for further details on the bootstrap and on the bootstrapping techniques used for the purpose of this work.

confidence interval constructed on the classical sample efficient frontier, it is evident from Figure 4.3 that when we are dealing with completely independent returns, the unbiased efficient frontier obtained by using the model of Bodnar and Bodnar, lies outside the confidence interval.

Figure 4.3: Operational effectiveness of the unbiased efficient frontier à la Bodnar and Bodnar, independent returns.



This situation, thus, suggests that the unbiased efficient frontier à la Bodnar and Bodnar in this case of independent returns, clearly represents an improvement from the classical technique of estimation of the efficient frontier and it should be preferred to the classical sample efficient frontier estimator.

4.3 Unbiased efficient frontier obtained on independent daily returns

In this section we repeat the very same analysis made in the section above, also in the case of daily returns in order to check if the behavior is still consistent with what we have been saying in Chapter 3.

We force again to zero all the off-diagonal elements of the covariance matrix of our daily returns, so that we can consider our dataset to be independent while leaving untouched the vector of the means and the vector of the covariances.

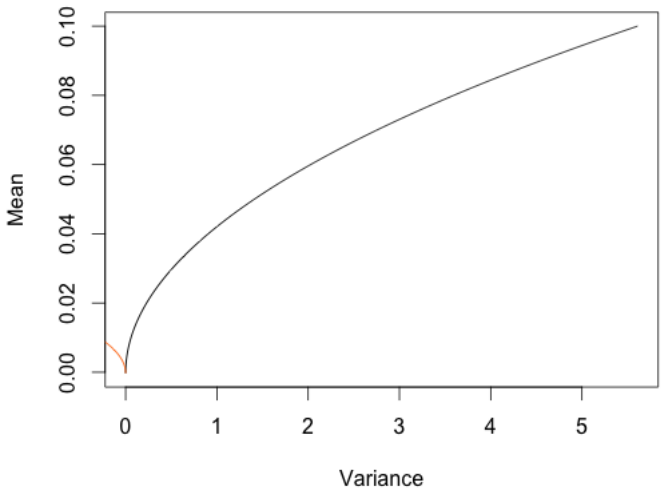
4.3.1 Sample efficient frontier versus unbiased efficient frontier à la Bodnar and Bodnar

By doing this manipulation of the covariance matrix in order to obtain a covariance matrix resembling the one in Equation 3.4, we come up with a diagonal matrix, with all the off-diagonal

elements equal to zero. This represents an independent set of asset returns. We, thus, can continue our analysis and by plotting the classical sample efficient frontier together with the unbiased efficient frontier à la Bodnar and Bodnar what we obtain is represented in Figure 4.4.

As in the case of monthly returns, the curve plotted in black represents the classical sample efficient frontier, while the orange one represents our unbiased efficient frontier à la Bodnar and Bodnar.

Figure 4.4: Sample efficient frontier versus unbiased efficient frontier à la Bodnar and Bodnar, independence of daily returns.



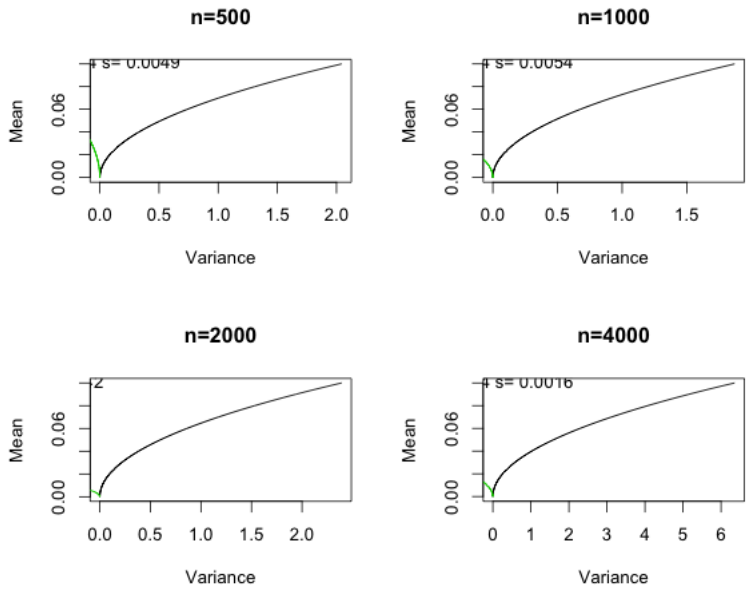
Unfortunately, as we have explained with Proposition 2, since we are dealing with daily logarithmic returns, it is highly more probable to witness the erratic behavior which leads to a frontier characterized by negative variance, which as we have already pointed out is not possible.

Due to this anomaly in the dataset, we cannot say anything about the mutual position of the classical sample efficient frontier and the unbiased one.

4.3.2 Asymptotic behavior of the unbiased efficient frontier à la Bodnar and Bodnar

Even if we have witnessed the erratic behavior according to which our unbiased efficient frontier shows negative variance, we can try to check anyway if the asymptotic behavior of the unbiased efficient frontier proposed by Proposition 1 of Corazza and Pizzi (2014) is present. We choose the following intervals for the number of observations, $n \in \{500, 1000, 2000, 4000\}$, but what we get, which is plotted in the following Figure 4.5, is not very satisfactory.

Figure 4.5: Asymptotic behavior of the unbiased efficient frontier à la Bodnar and Bodnar, daily independent returns.



As a matter of fact, the erratic behavior we have witnessed when considering the whole dataset, namely when $n = 4185$, is present also for all the n we choose for our asymptotic analysis. Again, this analysis is not very informative.

4.3.3 Operational effectiveness of the unbiased efficient frontier à la Bodnar and Bodnar

Since the unbiased efficient frontier for this number of observations shows the erratic behavior explained in Proposition 2 by Corazza and Pizzi, and since the bootstrap technique for such a number of observations is highly time-consuming, we decide not to compute the confidence interval for this dataset. It would also be quite useless, because it would not be informative on the position of the unbiased efficient frontier compared to the position of the classical sample frontier and its confidence interval.

4.4 Slope of the unbiased efficient frontier in case of independent returns

Since, in the words of Bodnar and Bodnar:

> "The slope parameter plays an important role in the construction of the efficient frontier. It shows how profitable is the market, i.e. how large is the increase in the portfolio's profit relative to the unit increase of the portfolio's variance.[2]"

[2]Bodnar, O. and Bodnar, T. (2010). On the unbiased estimator of the efficient frontier. Int. J. Theor. Appl. Finan., 13(07), pp.1065-1073.

it means that this slope is an important quantity in portfolio theory, and it is thus worth investigating it.

As we have already mentioned and reported in Chapter 3, Corazza and Pizzi (2014) have proved with their Proposition 2 that in order not to witness an erratic behavior of the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar it is necessary to have that the slope of the unbiased efficient frontier $\hat{s}$, is above a quantity which we have labeled as s limit. This can be seen in Proposition 2 and it can be summarized in the following equation:

$$\hat{s} < \frac{(n-1)(k-1)}{n(n-k-1)}, \tag{4.1}$$

$$\text{where } \hat{s} = \sum_{i=1}^{k} \frac{\hat{\mu}_i^2}{\hat{\sigma}_i^2}(1-\hat{p}_i) - \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\hat{\mu}_i \hat{\mu}_j}{\hat{\sigma}_j^2}\hat{p}_j,$$

$$\text{with } \hat{p}_i = \frac{\frac{1}{\hat{\sigma}_i^2}}{\sum_{j=1}^{K}\frac{1}{\hat{\sigma}_j^2}} \in (0,1).$$

While the Equation for $\hat{s}$ is reported in Equation 2.107.

In order to plot these two quantities, we have stored in two separate vectors their values for each increasing number of observations considered. Thanks to these vectors, we are also able to compare for each number of observations if the slope of the unbiased efficient frontier was above or below s limit, and count them.

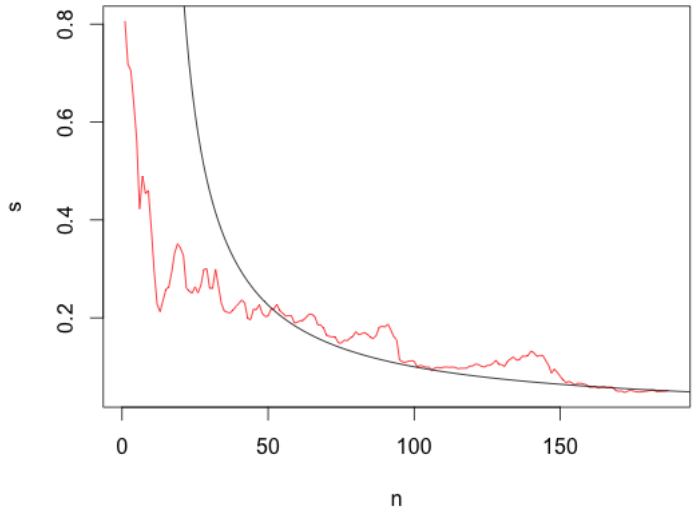
We report the results obtained both in case of monthly and daily returns.

4.4.1 Slope of the unbiased efficient frontier, monthly return

Theory says that the case of monthly returns should not be characterize by a great number of cases in which the slope of the unbiased estimator of the efficient frontier is below s limit.

If we plot the evolution of s limit and that of $\hat{s}$, according to the increasing number of observations n taken into consideration, what we get is the plot reported in the following Figure 4.6.

Figure 4.6: s limit vs $\hat{s}$, monthly independent returns.



If we only compare visually the plots of the black line, representing s limit, to the red line, representing $\hat{s}$, in Figure 4.6 probably we will not be able to get a sense of how many times the slope of the unbiased efficient frontier is above or below the critic value, s limit. In order to check for this specific aspect, what we do is numerically count how many times in this case of independent returns, the slope of the unbiased efficient frontier $\hat{s}$ is below s limit.

Out of a total of 186 observations, we have that in case of independent logarithmic returns, the slope of the unbiased efficient frontier is below s limit 50 times, meaning that the anomalous behavior of the unbiased efficient frontier happens more than 26% of the times.

We consider 186 observations out of the 197 available, since as it is clear from the denominator of Equation 4.1 we cannot compute s limit for a number of observations n below $(k+1)$, where k is the number of assets considered in the analysis.

In the Table 4.1 below we report the results of $\hat{s}$ and of s limit:

Table 4.1: $\hat{s}$ and s limit for monthly returns.

	$n = 197$
$\hat{s}$	0.05160925
s limit	0.05377967

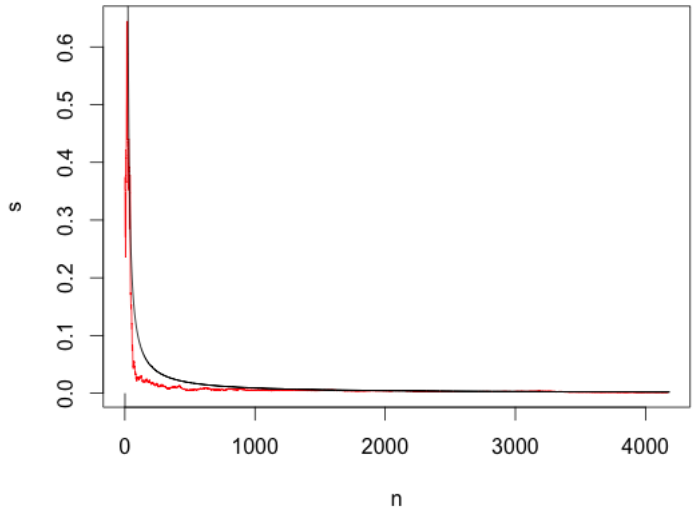
From the table it is clear to see that the condition in Proposition 2 by Corazza and Pizzi (2014) according to which $\hat{s}$ should be smaller than s limit is in this case not violated. This is the reason why we do not witness any erratic behavior of the unbiased efficient estimator of the efficient frontier.

4.4.2 Slope of the unbiased efficient frontier, daily return

We repeat the analysis of the unbiased estimator of the efficient frontier even in the case of daily independent returns.

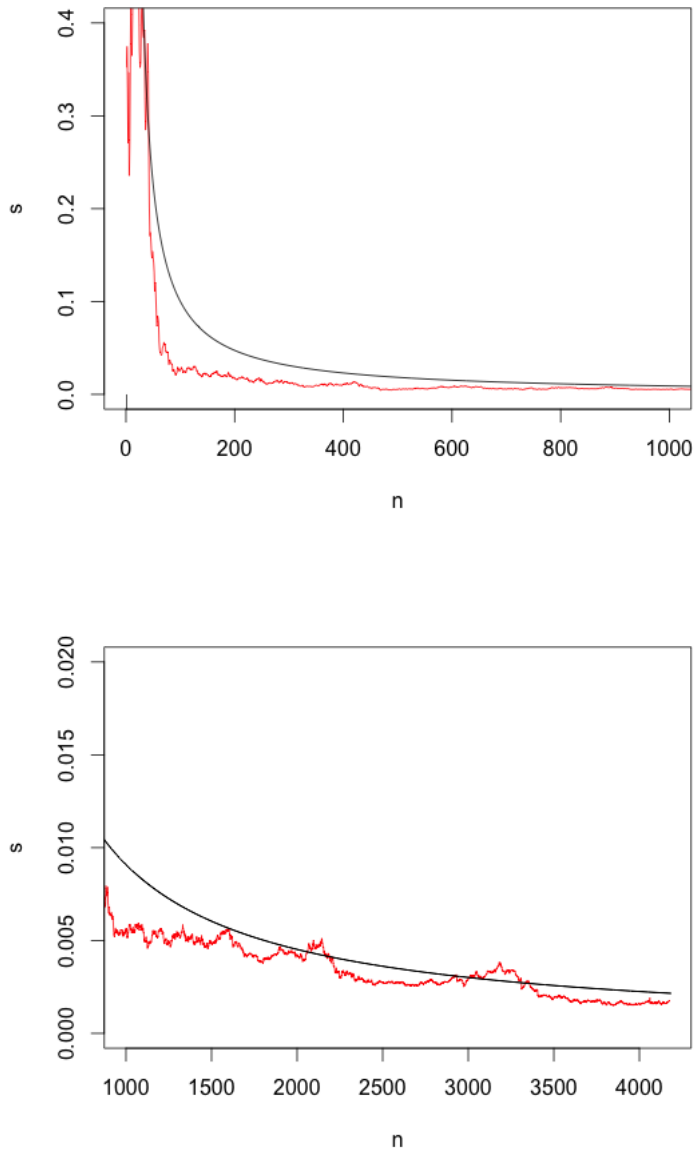
By plotting the slope of the unbiased efficient frontier against the plot of s limit when the number of observations increases, what we get is the plot reported in Figure 4.7.

Figure 4.7: s limit vs $\hat{s}$, daily independent returns.



Due to the big amount of observations by which this dataset is characterized, if left as it is, it is not very informative. We, thus, split it into two parts so that we can more easily see what happens to the evolution of s limit and $\hat{s}$. By splitting the plot in Figure 4.7 we get the plots in Figure 4.8.

Figure 4.8: s limit vs $\hat{s}$, daily independent returns split.



Even in this case is probable that if we only compare how the red and the black lines move in Figure 4.7 we will not be able to get a sense of how many times the $\hat{s}$ will be above or blow s limit. We thus compute this quantities numerically.

What we get is that out of 4174 observations, the times in which the slope of the unbiased efficient frontier is below s limit are 3681, meaning that 88% of the times we will witness the erratic behavior of the unbiased estimator of the efficient frontier.

Again, we consider 4174 observations when computing the evolution of the slope of the unbiased efficient frontier for the same reason explained in the case of daily returns; we must exclude the first 11 observations.

Even in this case of daily independent returns, we report in the following Table the results of $\hat{s}$ and s limit for this number of observations:

Table 4.2: $\hat{s}$ and s limit for daily returns.

	$n = 4185$
$\hat{s}$	0.002815273
s limit	0.002395785

From Table 4.2 we can immediately see that the condition in Equation 3.8 of Proposition 2 by Corazza and Pizzi (2014) is violated. This is the reason why in Figure 4.4 we witness the erratic behavior leading to an unbiased estimator characterized by negative variance.

Chapter 5

Robustness analysis

At this point of our analysis, we have checked how the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar works, if it follows the behavior proved by the propositions of Corazza and Pizzi (2014), and if it is operationally effective.

What we can do now is check whether the assumption of independence is a crucial one in the model. We can do this by testing its robustness, namely by inserting progressively in our model small departures from the assumption and check to what extent the model is able to bear them.

5.1 Dependence again, reintroduction of covariances in our dataset

In order to accomplish the tasks just described, we will go ahead in our analysis by trying to implement one couple of covariances at a time in the independent matrix computed in Chapter 4. In this way we will check at which degree of dependence the model is no more able to bear this violation of its assumption of independence.

5.1.1 One-by-one reintroduction of covariances, increasing order

In order to accomplish the task we have just mentioned, we proceed by reintroducing some degree of dependence in the covariance matrix we analyzed and used in the case of independent returns.

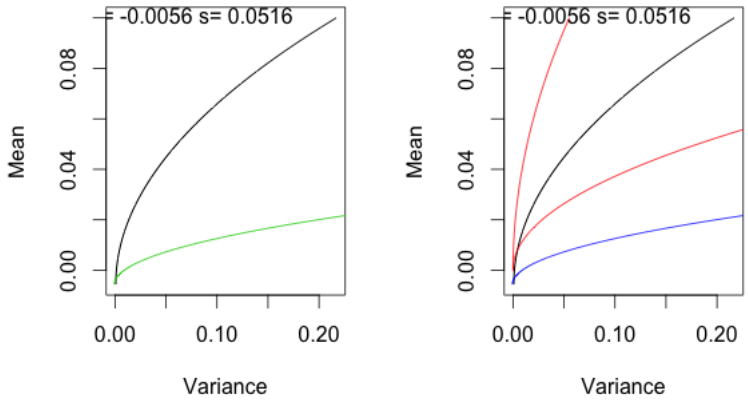
We will keep untouched the covariance matrix of independent returns, and this will imply that we will keep as they were throughout Chapter 4: the vector of means, the vector of variances, and we will keep all the off-diagonal elements of the covariance matrix equal to zero.

Since we have said that we want to reintroduce step by step some degree of dependence, we take the original covariance matrix, where dependence had not been forced to zero, and we put this covariances in a chart having care of keeping their coordinates, so that we will be able to relocate them in the same place into the covariance matrix. We will order them from the smaller to the bigger, and we will respect this order when reimplementing them into the independent covariance matrix. However, we will not report the table with the results in here, because it would be space-consuming and we suspect that we will not need all the couples of covariances, but only the firsts

of them. What it is important here, is to underline that we try to introduce some small degree of dependence, and we start this process from the smallest couple of covariances available, onward.

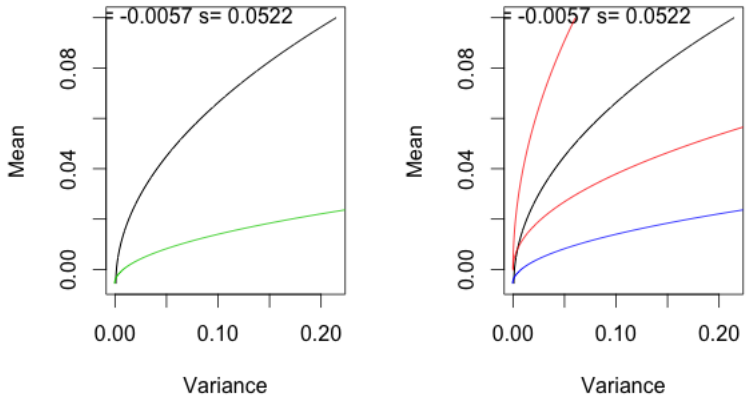
We report below the plots of the case of independent returns, so that we can quickly have a comparison with our new results and the initial situations. Figure 5.1 shows both the plot of the unbiased efficient frontier versus the classical sample one, and these same two curves with the confidence interval constructed on the classical sample efficient frontier, so that we will be able to check when the unbiased estimator of the efficient frontier enters into it.

Figure 5.1: Sample versus unbiased estimator in case of monthly independent returns.



The first couple of covariances we introduce is the one between Mediaset and Saipem, $cov_{7,8} = 0.002333185$. The situation is reported in Figure 5.2.

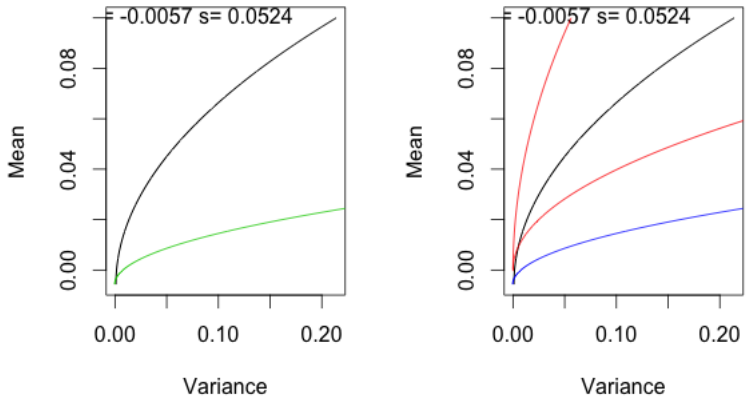
Figure 5.2: Reintroduction of the first couple of covariances.



By simply reintroducing one couple of covariances only, we do not witness a significant change in the relative positions of the classical sample estimator of the efficient frontier and the unbiased one. We thus try to implement the second couple of covariances and see what happens.

We thus have to add a second couple of covariances to the one covariance matrix plotted in the above Figure 5.2. We now introduce the covariance between the returns of A2A and the returns of Generali, $cov_{1,6} = 0.002373962$, and what we obtain is reported in the following Figure 5.3.

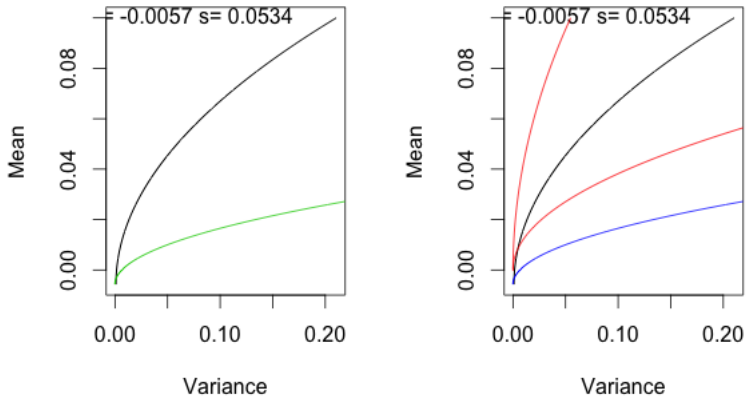
Figure 5.3: Reintroduction of the first two couples of covariances.



Even in this case, the change in the relative position of our curves of interest do not seem to be relevant. We thus try to increment one couple of covariances more.

In order to do this we have to implement the covariance between A2A and Saipem, $cov_{1,8} = 0.002404448$. What we get is plotted in Figure 5.4.

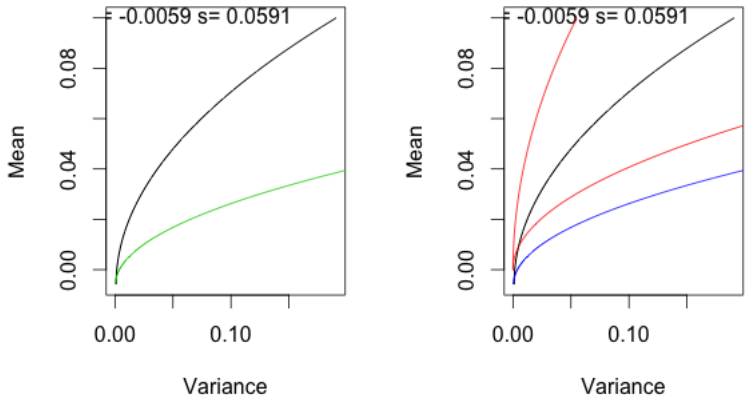
Figure 5.4: Reintroduction of the first three couples of covariances.



The change still appears not to be very relevant, but if we compare this scenario with the one of independent returns in Figure 5.1, we can start seeing that the unbiased efficient frontier is moving toward the classical sample one. However, since the unbiased estimator is still outside of the confidence interval constructed on the sample efficient frontier, we can proceed to the reintroduction of one couple of covariances more.

Then, we can add the covariance between the returns of Buzzi Unicem and Telecom, $cov_{3,10} = 0.002460712$. What we obtain is reported in Figure 5.5.

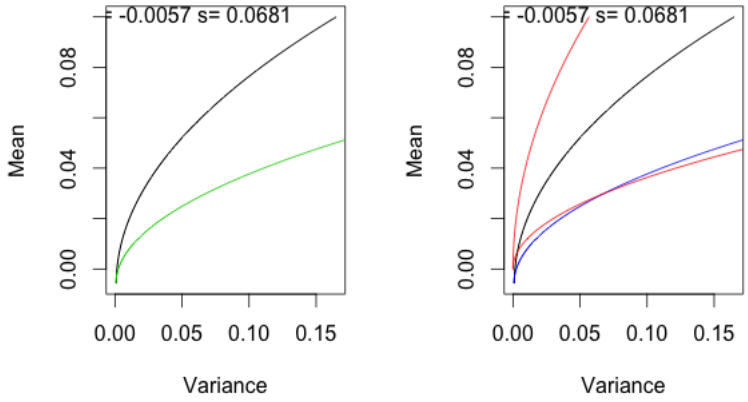
Figure 5.5: Reintroduction of four couples of covariances.



It is now more evident that the classical sample efficient frontier and the unbiased one are slowly converging.

At last, with the introduction of one couple more of covariances, we obtain the plot in Figure 5.6. The covariance we need to add is the one between the returns of A2A and Buzzi Unicem, $cov_{1,3} = 0.002472277$.

Figure 5.6: Reintroduction of five couples of covariances.



From the plot above, we can see that this is the maximum increase in the level of dependence the model can bear, because we can see that the blue line of the second plot -representing the unbiased estimator of the efficient frontier- falls within the confidence interval constructed on the classical sample efficient frontier.

To sum up we can say that in the specific case of our dataset, which was made up by 45 couples of covariances, the model could bear the reintroduction of 5 couples of covariances only. This mean that we have been able to reintroduce only the 11% of the covariance which would have been present in our original dataset, as we downloaded from Bloomberg, thus representing a real-world dataset.

5.2 Extreme case, complete reintroduction of dependence

We conclude this analysis on the robustness of the unbiased estimator proposed by Bodnar and Bodnar by considering the case in which we completely violate the hypothesis of independence. We try to see what happens when considering the data as it they were in reality and try, coupling these results with the ones obtained in the previous section, to grasp some insight on this estimator.

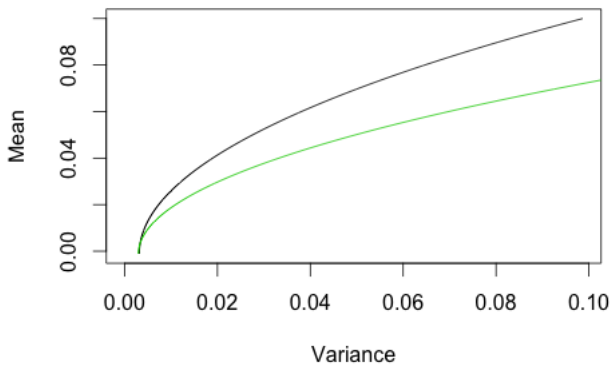
5.2.1 Monthly returns

5.2.1.1 Sample efficient frontier versus unbiased efficient frontier à la Bodnar

By applying the results of Bodnar and Bodnar (2010) we are able to plot, by using the code reported in the Appendix B and letting the flag (Sigma.diag = T), the unbiased efficient estimator of Bodnar and Bodnar together with the classical sample efficient frontier, and what we have obtained is reported in Figure 5.7.

The plot represents with the black curve the classical sample efficient frontier, while with the green curve the new unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar. What we can say by reading this graph and keeping in mind that we are considering a dataset which violates the assumptions of the estimator is that the sample efficient frontier is still an overoptimistic measure, providing higher return for any given level of risk, even though we know that the estimator proposed by Bodnar and Bodnar should have lost its operational effectiveness, due to the violation of the hypothesis of independence, and thus should not provide a correction for the bias of the classical sample estimator.

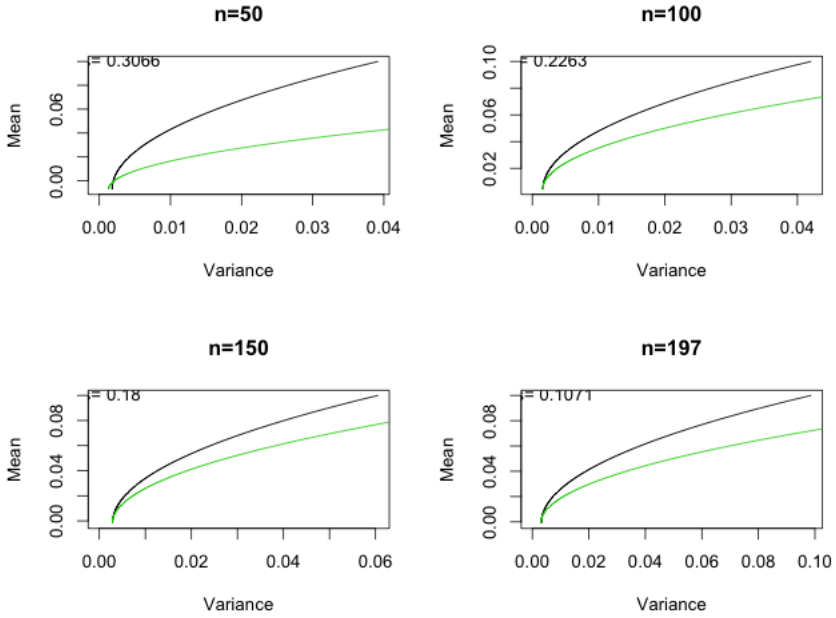
Figure 5.7: Sample efficient frontier versus unbiased efficient frontier à la Bodnar and Bodnar, monthly returns.



5.2.1.2 Asymptotic behavior of the unbiased efficient frontier à la Bodnar and Bodnar

In order to check whether the result obtained in Equation 3.7 applies also when violating the assumption of independence we can proceed just as we did in the case of completely independent returns, and thus we can proceed by considering varying levels of observations, n. In this actual case, we decide to take the following intervals for n: 50, 100, 150, and 197. From the results in Figure 5.8 it is clear how the span between the green and the black line, namely between the unbiased efficient frontier à la Bodnar and Bodnar and the classical sample efficient frontier, progressively diminishes.

Figure 5.8: Asymptotic behavior of the unbiased efficient frontier, monthly returns.



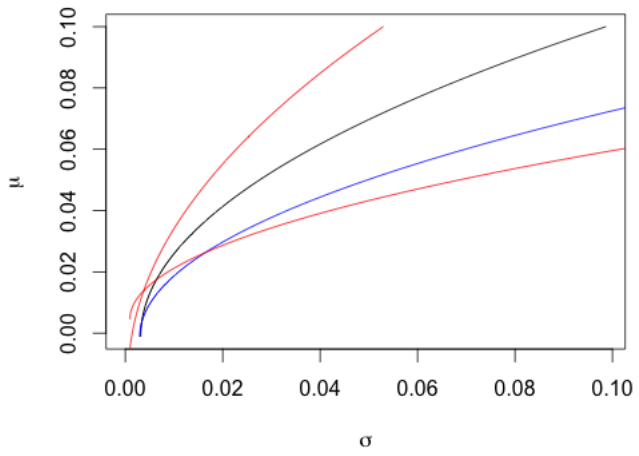
This means that, even when the assumptions of independence is violated, the estimator still shows the asymptotic behavior proved in Proposition1 by Corazza and Pizzi.

5.2.1.3 Operational effectiveness of the unbiased efficient frontier à la Bodnar and Bodnar

At last, we can check what happens to the operational effectiveness we witnessed in the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar in case of independent asset returns.

In Figure 5.9 we have the two red lines, which represent the confidence interval of the classical sample efficient frontier, we have the black line representing the sample efficient frontier, and the blue line which is instead the unbiased efficient frontier proposed by Bodnar and Bodnar.

Figure 5.9: Operational effectiveness of the unbiased efficient frontier à la Bodnar and Bodnar, monthly returns.



To construct the bootstrap in the case of returns which are not independent, we have to use the block bootstrap[1]. It is clear that from the plot that we do not witness any clear advantage in the implementation of this estimator in practice. This means that the hypothesis of independence of our dataset is a crucial one. As a matter of fact, if independence is violated, the estimator loses its fundamental characteristic, that of correcting for the bias of the classical sample efficient frontier.

5.2.1.4 Slope of the unbiased efficient frontier,

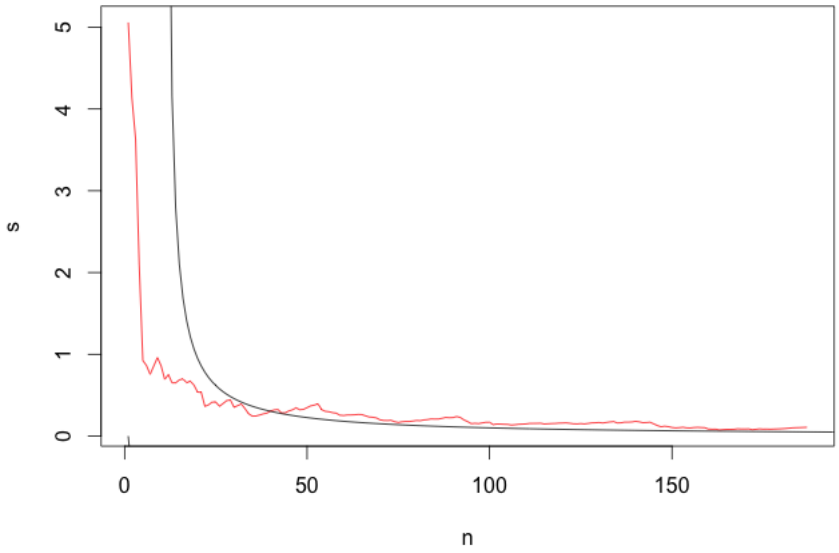
The quantity we have denoted as s limit depends on the number of observations we consider in our analysis, it thus changes depending on n. The slope of the efficient frontier, instead, evolves according to the change we have in the mean and variances, again due to the different number of observations considered.

In order to better understand how the relationship between these two quantities work we will compute for $n \in (12, 197)$ all the values of s limit and of $\hat{s}$. We then plot the evolution of these two quantities on the same graph so that two check when the slope of the unbiased efficient frontier is below the critical value, s limit.

We firstly plot the smooth curve representing the trend followed by s limit and then we proceed by superimposing the plot representing $\hat{s}$, the slope of our unbiased efficient frontier for an increasing number of observations, n. What we obtain is represented in Figure 5.10.

[1]See Appendix F for more details on clock bootstrap.

Figure 5.10: s limit vs $\hat{s}$, monthly returns.



This graph already gives a sense of what we will obtain next, namely that leaving aside the initial values where it is reasonable to witness an $\hat{s} < s$ limit, we have that the slope of the unbiased efficient frontier à la Bodnar and Bodnar is almost always above the critical value for monthly logarithmic returns.

In order to understand numerically on which proportions the this monthly returns will show the erratic behavior proven in previous sections, we have inserted in a chart the values of s limit and of $\hat{s}$ as the number of observations n changed. We have then counted the number of times in which we have that the slope $\hat{s}$ of the unbiased efficient frontier is below or above the quantity s limit and we got that in case of monthly logarithmic returns, the slope of the unbiased efficient frontier $\hat{s}$ is below $s$limit only 17 times out of 186, meaning that we have problems with the slope only 9.14% of the times.

If we compare this results, with those obtained in Chapter 4 when we analyzed the relationship between $\hat{s}$ and s limit in the case of completely independent asset returns, we can see that the incidence of the slope of the unbiased estimator of the efficient frontier being below the critical value we labeled s limit is more heavy in case of independent returns.

5.3 Final remarks

In this chapter what we have tried to grasp some insight on the behavior of the estimator proposed by Bodnar and Bodnar.

We have come up with interesting results, according to which we have witnessed that, for what concerns our specific dataset, out of 45 couples of covariances our model remains consistent with the properties listed and analyzed throughout this work until we introduce 5 couples of covariances.

This means that, we can have a departure from the assumption of independence of the 11%, before the unbiased estimator proposed by Bodnar and Bodnar loses the properties it has in the case in which its assumptions are not violated.

The convergence between the classical sample efficient frontier and the unbiased estimator proposed by Bodnar and Bodnar appears gradually, but once it stands out it is irreversible and it makes the estimator lose its fundamental property, that of correcting for the bias of the classical sample efficient frontier. We can thus deem the assumption of independence to be a crucial one. Also because the model can only bear the reintroduction of 5 couples of covariances, meaning that it can bear a very small departure from the independence assumption.

By applying the very same analysis we have applied in Chapter 4 to our independent datasets, to the original dataset characterized by dependence, even if we know that we are violating the assumption on which the estimator is based on and we thus expect to have results which are not significant, it is anyway interesting to discover that the estimator other than losing its operational effectiveness, namely by converging to its biased counterpart, continues to show its other properties.

Conclusions

Throughout the thesis we have presented and applied the main principles of Modern Portfolio Theory. Starting from the seminal work of Markowitz, passing through the pages written by Merton, we have come to review more recent works, such as those by Tu and Zhou, Basak, Jagannathan and Ma, Siegel and Woodgate, and Kan and Smith, until we reach the piece of Bodnar and Bodnar.

We have concentrated our efforts on their pages and on their results. We have presented and deeply analyzed the unbiased estimator of the efficient frontier they proposed in their work of 2010. We have criticized that the assumptions on which the model is based on, those of normality and of independence of asset returns, appear to be too stringent for such an estimator to be applied to real-world datasets.

In order to grasp some insight on the unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar, we have started our analysis by assuming the very same assumptions made by Bodnar and Bodnar and we have checked whether everything works as they predict. We have also checked for some propositions derived by Corazza and Pizzi, and we contributed to their work by computing some other results regarding the impact mean and variance of the estimator have on its shape.

Thanks to this preliminary analysis we have seen that everything works as predicted in the case in which we do not violate any of the hypothesis on which Bodnar and Bodnar base their unbiased estimator. Since we did not meet any problem with this measure, we have decided to check if the assumption of independence is a crucial one in the model or if there is some ground of departure from it.

In order to do this, we have progressively reimplemented some degree of dependence in our independent dataset. However, our results show that the model can bear only something more than 10% of the actual dependence which our real dataset initially showed. This means that the hypothesis of independence is a crucial one for the model.

We conclude our analysis by showing the results of the very same analysis we implemented in case of independent dataset, for the case of completely dependent dataset. What we obtain is that, except for the lost operational effectiveness, the model still shows all of the other characteristics which we attributed to it throughout this paper.

That is why, in order to broaden the scope of this work, we suggest to try to find a way in which to make the model account for dependent dataset. This aspect, if accomplished, other than interesting on a theoretical ground, it would also be very important in that it would let us apply this unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar to real datasets.

Appendix A. Bayesian inference

Statistical inference is drawing conclusions about an entire population based on data drawn from that population. Two are the approaches commonly used in statistical inference, the Bayesian approach and the frequentist one.

All Bayesian inference is based on the posterior distribution, which contains all the current information about the unknown parameter. Although a plot of the posterior density gives a full graphical description, numeric summaries of the posterior are needed as well.

Bayesian statistical conclusions about a parameter θ, or unobserved data $\tilde{y}$, are made in terms of probability statements. These probabilities are conditional on the observed value of y, and in our notation are written simply as $p(\theta \mid y)$ or $p(\tilde{y} \mid y)$. We also implicitly condition on the known values of any covariates, x. It is at the fundamental level of conditioning on observed data that Bayesian inference departs from the approach of frequentist inference, which is based on a retrospective valuation of the procedure used to estimate θ or $\tilde{y}$ over the distribution of possible y values conditional on the true unknown value of θ.

We now present the basic mathematics and notation of Bayesian inference.

We start the derivation of the basic mathematics underlying Bayesian inference by recalling the Bayes' rule. In order to make probability statement about the parameter θ given y, we begin with a model which provides a joint probability distribution for θ and y. This density function can be written as a product of two densities:

- the prior distribution $p(\theta)$, and
- the sampling, or data, distribution $p(\theta \mid y)$,

which lead to the following result:

$$p(\theta, y) = p(\theta)\, p(y \mid \theta).$$

Simply conditioning on the known value of the data y, using the basic property of conditional probability known as Bayes' rule, yields the posterior density:

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)\, p(y \mid \theta)}{p(y)}, \tag{5.1}$$

where $p(y) = \sum_{\theta} = p(\theta)\, p(y \mid \theta)$.

We can also write Equation 5.1 omitting the factor $p(y)$, which does not depend on θ and, with fixed y, can thus be considered a constant, yielding to the unnormalized posterior density, which becomes:

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta). \tag{5.2}$$

These simple expressions contain the tecnical core of Bayesian inference: the primary task of any specific application is to develop the model $p(\theta, y)$ and perform the necessary computations to summarize $p(\theta \mid y)$ in appropriate ways.

To make inferences about an unknown observable, often called predictive inferences, before data y are considered, its distribution is:

$$p(y) = \int p(y, \theta)\, d\theta = \int p(\theta)\, p(y \mid \theta)\, d\theta, \tag{5.3}$$

which is often called the prior predictive distribution: prior in that it is not conditional on a previous observation of the process, and predictive in that it is the distribution of a quantity that is observable.

After the data y have been observed, we can predict an unknown observable, $\tilde{y}$, from the same process. The distribution of $\tilde{y}$ is called the posterior predictive distribution, posterior because it is conditional on the observed y and predictive in which it is a prediction for an observable $\tilde{y}$:

$$\begin{aligned} p(\tilde{y} \mid y) &= \int p(\tilde{y}, \theta \mid y)\, d\theta \\ &= \int p(\tilde{y} \mid \theta, y)\, p(\theta \mid y)\, d\theta \\ &= \int p(\tilde{y} \mid \theta)\, p(\theta \mid y)\, d\theta. \end{aligned} \tag{5.4}$$

Appendix B. The Jackknife

The jackknife is a technique for estimating the bias and standard error of an estimate.

Suppose we have a sample $\mathrm{x} = (x_1, x_2, ...x_n)$ and an estimator $\hat{\theta} = s(\mathrm{x})$. We wish to estimate the bias and standard error of $\hat{\theta}$. The jackknife focuses on the samples that leave out one observation at a time:

$$\mathrm{x}_{(i)} = (x_1, x_2, ..., x_{i-1}, x_{i+1}, ...x_n)$$

for $i = 1, 2, ..., n$, called the jackknife samples. The i^{th} jackknife sample consists of the dataset with the i^{th} observation removed. Let

$$\hat{\theta}_{(i)} = s\left(\mathrm{x}_{(i)}\right)$$

be the i^{th} jackknife replication of $\hat{\theta}$. The jackknife estimate of bias is defined by:

$$\widehat{bias}_{jack} = (n-1)\left(\hat{\theta}_{(\cdot)} - \hat{\theta}\right),$$

where $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n$.

The jackknife estimate of the standard error is defined by:

$$\widehat{se}_{jack} = \left[\frac{n-1}{n}\sum\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2\right]^{1/2}.$$

Appendix C. Sensitivity analysis

A possible definition of sensitivity analysis is given by Saltelli (2004) as:

> The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input.

As a matter of fact the aim of sensitivity analysis is that of describing how much model output values are affected by changes in model input values. It is the investigation of the importance of imprecision or uncertainty in model inputs in a decision-making or modeling process. Sensitivity analysis can thus be used to interpret model results and to identify to where the efforts to improve a model and input values should be directed to.

A practice related to sensitivity analysis is uncertainty analysis, which focuses more on quantifying uncertainty in model output. These two analyses should both be performed, with uncertainty analysis preceding sensitivity analysis. An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence, while a sensitivity analysis attempts to determine the change in model output values that results from modest changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs.

Many authors, when referring to the degree to which an input parameter affects the model output, use the terms sensitive, important, most influential, major contributor, effective or correlated interchangeably. Crick et al. (1987) have made a distinction by referring to important parameters as those whose uncertainty contributes substantially to the uncertainty in assessment results, and sensitive parameters as those which have a significant influence on assessment results.

The consensus among authors is that models are indeed sensitive to input parameters in two distinct ways:

- the variability, or uncertainty, associated with a sensitive input parameter is propagated through the model resulting in a large contribution to the overall output variability, and
- model results can be highly correlated with an input parameter so that small changes in the input value result in significant changes in the output.

The necessary distinction between important and sensitive parameters is in the type of analysis being conducted: uncertainty analysis (parameter importance) or sensitivity analysis (parameter sensitivity). An important parameter is always sensitive because parameter variability will not appear in the output unless the model is sensitive to the input. A sensitive parameter, however,

is not necessarily important because it may be known precisely, thereby having little variability to add to the output. At the completion of an analysis on parameter sensitivity the analyst holds a list, or 'sensitivity ranking', of the input parameters sorted by the amount of influence each has on the model output.

Many are the approaches proposed in the literature to sensitivity analysis, but we will use differential analysis, also referred to as the direct method.

For the scope of our analysis we will apply differential sensitivity analysis.

Differential sensitivity analysis

Differential analysis is also referred to as the direct method to sensitivity analysis. In this framework a sensitivity coefficient ϕ_i is the ratio of the change in the output to the change in the input while all other parameters remain constant.

Sensitivity analysis using partial differentiation techniques is computationally efficient but when we have to deal with complicated models it often requires complex mathematical procedures. In the case in which, instead, the relationship between the independent variables and the dependent one is described by an easy explicit algebraic equation, the sensitivity analysis is easy to be performed. In this case, the sensitivity coefficient ϕ_i for a particular independent variable can be calculated from the partial derivative of the dependent variable with respect to the independent variable:

$$\phi_i = \frac{\partial Y}{\partial X_i},$$

with X_i being the independent variable, and Y the dependent one.

Appendix D. Granger causality

Since in real world we do not deal with single stationary time series, we must take into consideration also the causal relationships between them. There are two different approaches for taking into consideration such relationship:

- following a bottom up strategy, where we first assume that the data generating processes of the different time series are independent of each other, and then we ask whether some specific time series are related to each other;
- the alternative is a top down strategy, which assumes that generating processes are not independent and then asks whether some specific times series generated independently of the other time series are considered.

The first approach, was firstly introduced by Clive W. J. Granger (1969) and it is today still employed when causality test are performed. The definition Granger proposes of causality looks at whether the forecasts of the future values of a variable, y, can be improved if - besides all other available information - the current and lagged values of the other variable, x, are also taken into account.

We assume to have a weakly stationary time series. Let I_t be the total information set available at time t. This information set includes, above all, the to time series x and y, we have mentioned above. Let $\bar{x}_t$ be the set of all current and past values of x, i.e. $\bar{x}_t := \{x_t, x_{t-1}, ..., x_{t-k},\}$ and analogously of y. Let $\sigma^2(\cdot)$ be the vairance of the corresponding forecast error. For such a situation, Granger (1969) proposed the following definition of causality between x and y:

- Granger causality: x is simply Granger causal to y if and only if the application of an optimal prediction function leads to

$$\sigma^2(y_{t+1} \mid I_t) < \sigma^2(y_{t+1} \mid I_t - \bar{x}_t),$$

  i.e. if future values of y can be better predicted, i.e. with a smaller forecast error variance, if current and past values of x are used.

- Instantaneous Granger causality: x is instantaneously Granger causal to y if and only of the application of an optimal linear prediction function leads to

$$\sigma^2(y_{t+1} \mid \{I_t, x_{t+1}\}) < \sigma^2(y_{t+1} \mid I_t),$$

i.e. if the future value of y, y_{t+1}, can be better predicted, i.e. with smaller forecast error variance, if the future value of x, x_{t+1}, is used in addition to the current and past values of x.

- Feedback: there is a feedback between x and y if x is causal to y and y is causal to x. Feedback is only defined for the case of simple causal relations. The reason is that the direction of instantaneously causal relations cannot be identified without additional information or assumptions. Thus the following theorem holds:

  Theorem 3: x is instantaneously causal to y if and only if y is instantaneously causal to x.

According to Granger's definition of causality there are eight different, exclusive possibilities of causal relations between two time series:

1. x and y are independent;
2. there is only instantaneous causality;
3. x is causal to y, without instantaneous causality;
4. y is causal to x, without instantaneous causality;
5. x is causal to y, with instantaneous causality;
6. y is causal to x, with instantaneous causality;
7. there is feedback without instantaneous causality;
8. there is feedback with instantaneous causality.

Application to the dataset

By applying the Granger causality test to our dataset, we get a matrix of causality p-values reported in Table 5.1

Table 5.1: Granger causality test, monthly returns.

	A2A	BMPS	BZU	EXP	FCA	G	MS	SPM	STM	TIT
A2A	NA	0,22	0,15	0,19	0,05	0,09	0,14	0,01	0,01	0,05
BMPS	0,22	NA	0,44	0,63	0,41	0,27	0,11	0,01	0,33	0,02
BZU	0,15	0,44	NA	0,20	0,07	0,33	0,38	0,26	0,03	0,02
EXP	0,19	0,63	0,20	NA	0,09	0,27	0,11	0,33	0,04	0,02
FCA	0,05	0,41	0,07	0,09	NA	0,33	0,20	0,02	0,03	0,03
G	0,09	0,27	0,33	0,27	0,33	NA	0,10	0,16	0,13	0,07
MS	0,14	0,11	0,38	0,11	0,20	0,10	NA	0,01	0,59	0,11
SPM	0,01	0,01	0,26	0,33	0,02	0,16	0,01	NA	0,51	0,11
STM	0,01	0,33	0,03	0,04	0,03	0,13	0,59	0,51	NA	0,22
TIT	0,05	0,02	0,02	0,02	0,03	0,07	0,11	0,11	0,22	NA

Since, if we continue to maintain a significance level of $\alpha = 0.05$, almost all of the returns above show a $p-$value greater than α, we can conclude that if we use causality as a proxy for non-linear dependence, our monthly returns show non-linear dependence.

Appendix E. R code

We here report the four, most important functions used in this thesis for the construction of all our analysis. The functions have been constructed by professor Pizzi, and then have been arranged by the author of this piece to develop all the analysis reported in the previous chapters. These codes have been the building blocks of our analysis.

```
library(tseries)

library(FinTS)

library(forecast)

library(fGarch)

library(corpcor)

#-------------------------------------------------------------------------------------------------------------------
-----#

ef.boot=function (dati, n.start = 20, ordine=c(5,5), order.garch=c(1,1), rseed=NA, covar=F, npoints=100,
alfa=0.05, qq=T, doprint=F, nboot=1000, sieve=F, type="boot", H0="S", prob=0.05,sigma=F,
check=T, rmin=NULL) {

#-------------------------------------------------------------------------------------------------------------------
-----#

#

# dati = matrix of returns

# nboot = number of bootstrap resampling

# n.start = tems for initialization

# extended = multivariate random numbers generation

# rseed = initialization

# ordine = maximum (p,q) order

# covar = if TRUE bootstrap resampling tries to keep also the covariance

# npoints = number of valuation point for the estimate of the EF

# alfa = 1 - confidence level

# qq = if TRUE then LL e UL are computed quantile vs quantile
```

```
# doprint = if TRUE some tags are printed on the graph
# H0 = if S confidence interval is constructed on the sample EF, if U it is constructed on unbiased EF
# type = type of bootstrap applied
# prob = parameter for block bootstrap reinitialization
# sigma = if TRUE off-diagonal elements are forced to zero
# check = if TRUE check of the sign of s for bootstrap resampling
# rmin = if not NULL it is used to estimate the frontiers
#-----------------------------------------------------
# setting the variables for the procedure
n.obs=dim(dati)[1] # number of observations
n.serie=dim(dati)[2] # number of series
res=matrix(NA,ncol=n.serie,nrow=n.obs) # matrix of residuals of mean model
resg=matrix(NA,ncol=n.serie,nrow=n.obs) # matrix of residuals of volatility model
ar.par=matrix(0,ncol=n.serie,nrow=ordine[1]) # matrix of AR model's parameters
ma.par=matrix(0,ncol=n.serie,nrow=ordine[2]) # matrix of MA model's parameters
order.ar=double() # number of AR parameters
order.ma=double() # number of MA parameters
mu=double() # vector of intercepts
omega=double() # vector of GARCH model's constants
alpha=double() # vector of ARCH parameters
beta=double() # vector of GARCH parameters
delta = 2
deltainv=1/delta
replicheR=matrix(NA,ncol=nboot,nrow=npoints) # bootstrap unbiased frontier matrix
replicheV=matrix(NA,ncol=nboot,nrow=npoints) # bootstrap unbiased frontier matrix
s.val=double() # vector of s
s.boot=double() # vector of flags for finding bootstrap resampling with negative variances
#
# computation of efficient frontier and plot of graphs
#
out.dati=EF.unbiased(dati, npoints=npoints, doprint=F, doplot=F, rmin=rmin)
# choice of type of bootstrap
if(type=="boot"){
```

```
#
# cycle for estimation of each AR model (mean) and GARCH model (volatility)
#
for(j in 1:n.serie){
#
# AR model
#
fit=auto.arima(dati[,j]) # estimate of Ar model
order.ar[j]=ifelse(length(fit$model$phi)>0,length(fit$model$phi),0) # order AR
order.ma[j]=ifelse(length(fit$model$theta)>0,length(fit$model$theta),0) # order MA
mu[j]=ifelse(length(fit$coef)>(order.ar[j]+order.ma[j]),fit$coef[order.ar[j]+order.ma[j]+1],0) # intercepts
if(order.ar[j]>0) # check if there is an AR model
ar.par[1:order.ar[j],j]=rev(fit$model$phi)
if(order.ma[j]>0) # check if there is an AR model
ma.par[1:order.ma[j],j]=rev(fit$model$theta)
res[,j]=fit$residuals # residuals of AR model
#
# GARCH(1,1) for volatility
#
test=ArchTest(res[,j]) # check presence of ARCH effects
if(test$p.value<alfa){
options(warn=-1)
fitg=garchFit(~garch(1,1),res[,j],include.mean=F,trace=F)
options(warn=0)
omega[j]=fitg@fit$matcoef[1,1]
if(!is.na(sum(fitg@fit$se.coef))){ # check for problems in the estimation of the model
if(fitg@fit$matcoef[3,4]<0.05){
beta[j]=fitg@fit$matcoef[3,1]
alpha[j]=fitg@fit$matcoef[2,1]
}else{
beta[j]=0
alpha[j]=ifelse(fitg@fit$matcoef[2,4]<0.05,fitg@fit$matcoef[2,1],0)
```

```
}
}else{
beta[j]=0
options(warn=-1)
fitg=garchFit(~garch(1,0), res[,j], include.mean=F) #try ARCH estimation
options(warn=0)
if(!is.na(sum(fitg@fit$se.coef))){ # check for problems in the estimation of the model
alpha[j]=ifelse(fitg@fit$matcoef[2,4]<0.05,fitg@fit$matcoef[2,1],0)
omega[j]=ifelse(fitg@fit$matcoef[2,4]<0.05,fitg@fit$matcoef[1,1],0)
}else{
alpha[j]=0
omega[j]=0
}
}
}else{ # test arch
alpha[j]=0
beta[j]=0
omega[j]=0
}
if(alpha[j]!=0){
resg[,j]=fitg@residuals
}else{
resg[,j]=res[,j]
}
}
model=list(mu=mu, ar=ar.par, ma=ma.par, omega=omega, alpha=alpha, beta=beta)
#-----------------------------------------------------------
# end of model's estimation
#-----------------------------------------------------------
}
if (!is.na(rseed)) # initialization generation of psudo random number
set.seed(rseed)
#-----------------------------------------------------------
```

```
# generation of bootstrap resampling
#-------------------------------------------------------------
if(type=="boot"){
cat("Bootstrap of residuals\n")
boot.series=boot.g(model=model, res=resg, n=n.obs, n.start=n.start, nboot=nboot, sieve=F, type="boot",
order.garch=c(1,1))
yboot=boot.series$dati.boot # bootstrap resampling
}
if(type=="block"){
cat("Bootstrap a blocchi\n")
yboot=block.boot(dati, R=nboot, gprob=prob,check)$x
}
if(type=="surrogate"){
cat("dati surrogati\n")
yboot=surrog(dati, R=nboot)
}
if(sum(type==c("boot","block","surrogate"))==0)
stop("Missing or wrong type")
l=1
for(j in 1:nboot){
if(H0=="U"){
if(qq){
out=EF.unbiased(yboot[,j], npoints=npoints, doprint=F,doplot=F,Sigma.diag=sigma,rmin=rmin)
}else{
out=EF.unbiased(yboot[,j], npoints=npoints, rmin=out.dati$R, doprint=F, doplot=F,Sigma.diag=sigma)

}
s.boot[l]=(out$S[2]>out$S[1] & out$S[3]>out$S[1])
replicheR[,j]=out$R.UB
replicheV[,j]=out$V.UB
s.val[j]=out$S[1]
l=ifelse(s.boot[l],l,(l+1))
}else{
```

```
if(qq){
out=EF.sample(yboot[,,j], npoints=npoints, doprint=F, maxv=0.05)
}else{
out=EF.sample(yboot[,,j], npoints=npoints, rmin=out.dati$R, doprint=F, maxv=0.05)
}
replicheR[,j]=out$R
replicheV[,j]=out$V
s.val[j]=out$S[1]
}
}
EF.unbiased(dati,addplot=F,color=4,Sigma.diag=sigma)
if(qq){
lines(x=apply(replicheV,1,quantile,probs=(alfa/2)), y=apply(replicheR,1,quantile,probs=(alfa/2)),col=2)
lines(x=apply(replicheV,1,quantile,prob=(1-alfa/2)), y=apply(replicheR,1,quantile,probs=(1-alfa/2)),col=2)
}else{
lines(x=apply(replicheV,1,quantile,probs=(alfa/2)), y=replicheR[,1],col=2)
lines(x=apply(replicheV,1,quantile,prob=(1-alfa/2)), y=replicheR[,1],col=2)
}
if(doprint){
text(x=0.045,y=0.01,"Unbiased Efficient Frontier",cex=.75)
text(x=0.04,y=0.012,"Sample Efficient Frontier",cex=.75)
text(x=0.04,y=0.006,"Sample Efficient Frontier LL = 0.05",cex=.75)
text(x=0.045,y=0.006,expression(alpha),cex=.75)
text(x=0.016,y=0.012,"Sample Efficient Frontier UL = 0.05",cex=.75)
text(x=0.021,y=0.012,expression(alpha),cex=.75)
}
return=list(yboot=yboot,par.ar=ar.par,alpha=alpha,omega=omega,beta=beta,V.star=replicheV,
R.star=replicheR,s=s.val,
V=out.dati$V,R=out.dati$R)
}
#-------------------------------------------------------------------------------------------------------------------
-----#
boot.g=function (model , res, n = 100, n.start = 100, nboot=1000, sieve=F, type="boot", or-
der.garch=c(1,1)) {
```

```
#-------------------------------------------------------------------------------------------------------------------------
-----#
#
# model = list with all the parameter of the model ARIMA(p,d,q) + GARCH(r,s)
# res = matrix of the residuals of the ARIMA+GARCH model for each asset
# n = number of observations
# n.start =number of observations for inizialization
# nboot = numebr of bootstrap (of residuals) resampling
# sieve = logical if sieve boostrap
# type = distribution for sieve bootstrap
#
#-----------------------------------------------------------------------------
n = n + n.start
delta = 2
deltainv = 1/delta
#---------------------------------------------------------
# retrieve models' parameters
#---------------------------------------------------------
n.serie=length(model$mu)
mu = matrix(model$mu,ncol=n.serie,nrow=1)
ar = model$ar
ma = model$ma
omega = matrix(model$omega,ncol=n.serie,nrow=1)
alpha = matrix(model$alpha,ncol=n.serie,nrow=order.garch[1])
beta = matrix(model$beta,ncol=n.serie,nrow=order.garch[2])
#---------------------------------------------------------
# models' order
#---------------------------------------------------------
order.ar = nrow(ar)
order.ma = nrow(ma)
order.alpha = nrow(alpha)
order.beta = nrow(beta)
m=max(order.ar,order.ma,order.alpha,order.beta)
```

```
#-------------------------------------------------------
# setting matrices for bootstrap
#-------------------------------------------------------
# create an array for bootstrap series
z=array(NA,dim=c(n,n.serie,nboot))
h=array(NA,dim=c(n,n.serie,nboot))
yboot=array(NA,dim=c(n,n.serie,nboot))
for(i in 1:n.serie){
z[,i,]=sample(res[,i],n*nboot,replace=T)
if(alpha[1,i]!=0){
if(beta[1,i]!=0){
h[1:n.start,i,]=omega[i]/(alpha[1,i]+beta[1,i])
}
}else{
h[,i,]=1
}
}
eps = h^deltainv * z
yboot[1:m,,]=z[1:m,,]
if (type=="sieve") {
# to write
}
if (type=="wild") {
# to write
}
#-------------------------------------------------------
# build bootstrap series
#-------------------------------------------------------
for(l in 1:n.serie){
for(j in 1:nboot){
for (i in (m + 1):n) {
if(alpha[1,l]!=0){
if(beta[1,l]!=0){
```

```
h[i,l,j] = omega[l] + sum(alpha[,l] * eps[(i - (1:order.alpha)),l,j]^2) + sum(beta[,l] * h[(i - (1:order.beta)),l,j])

eps[i,l,j] = h[i,l,j]^deltainv * z[i,l,j]

}

}

yboot[i,l,j] = mu[l] + sum(ar[,l] * yboot[i - (1:order.ar),l,j]) + sum(ma [,l]* eps[i - (1:order.ma),l,j]) + eps[i,l,j]

}

}

}

dati = yboot[-(1:n.start),,]

return=list(dati.boot=dati, eps.boot=eps, res.boot=z, h.boot=h)

}

#-------------------------------------------------------------------------------------------------------------------------
-----#

EF.unbiased=function(dati, addplot=F, npoints=1000, rmin=NULL, Sigma.diag=T, color=3, doprint=T, doplot=T){

#-------------------------------------------------------------------------------------------------------------------------
-----#

# dati = matrix of log-returns

# addplot = if FALSE new graph, if TRUE add to previous graph

# npoints = numebr of points on which to calculate the EF

# rmin = foces at a predetermined minimum value of r

# Sigma.diag = forces off-diagonal elements to zero

# color = color for unbiased EF

# doprint = if TRUE prints some info

#-----------------------------------------------------------------------

k=dim(dati)[2]

n=dim(dati)[1]

mu=apply(dati,2,mean,na.rm=T)

Sigma=cov(dati)

if(Sigma.diag)

Sigma=diag(diag(Sigma))

if(det(Sigma)==0){
```

```
S1=pseudoinverse(Sigma)
}else{
S1=solve(Sigma)
}
one=rep(1,dim(dati)[2])
if(doprint){
cat("_____________________\n\n",sum(S1),"\t",sum(mu^2*diag(S1)),"\t",
sum(mu*diag(S1)),"\n")
cat(sum(S1)*sum(mu^2*diag(S1)),"\t",sum(mu*diag(S1))^2,"\n\n_______
_____________\n\n")
}
#---------------------------------------------------------------------------
# Sample efficient frontier
#---------------------------------------------------------------------------
Vgmv=1/(t(one)%*%S1%*%one)
Rgmv=(t(one)%*%S1%*%mu)*Vgmv
R=S1-(S1%*%one%*%t(one)%*%S1)/c(t(one)%*%S1%*%one)
s=c(t(mu)%*%R%*%mu)
#---------------------------------------------------------------------------
# Unbiased estimator of the efficient frontier
#---------------------------------------------------------------------------
a1=(n-2)*(n-1)/(n*(n-k)*(n-k-1))
a2=(n-k-1)/(n-1)*s-(k-1)/n
a3=(n-k-2)/(n-k)*(n-1)/(n-k-1)
s.hat3=(n-1)*(-2*n+2*n*k-k*(k+1)+2)/(n*(n-k)*(n-k-1))
#---------------------------------------------------------------------------
# check conditions for variance>0
#---------------------------------------------------------------------------
s.hat2=(n-1)*k /(n*(n-k-2))
s.hat1=(n-1)*(k-1) /(n*(n-k-1))
if(s.hat2>s & s.hat1>s & doprint){
warning(paste("Il parametro s non soddisfa condizione positivita varianza\n"),
cat("s=",round(s,5),"\t s1=",round(s.hat1,5),"\t s2=",round(s.hat2,5),"\n\n"))
```

```
}
if(s.hat2<s & s.hat1>s & doprint ){
warning("Il parametro s soddisfa parzialmente condizione positivita varianza")
cat("s=",round(s,5),"\t s1=",round(s.hat1,5),"\t s2=",round(s.hat2,5),"\n\n")
}
if(is.null(rmin)){
r=seq(from=Rgmv,to=0.1,length=npoints)
}else{
r=seq(from=rmin,to=0.1,length=npoints)
}
v=1/s*(r-Rgmv)^2+Vgmv
vUB=((r-Rgmv)^2 - a1*Vgmv)/a2+a3*Vgmv # da formula 2.4 p.1070 articolo Bodnar and Bodnar
if(doprint){
cat(" Medie dei rendimenti logaritmici giornalieri \n")
print(round(mu,5))
cat("\n Deviazioni standard dei rendimenti logaritmici giornalieri\n")
print(round(sqrt(diag(Sigma)),5))
cat("\n Global Minimum Variance Portfolio\n")
cat("Expected return\t",round(Rgmv,5),"\t")
cat("Variance\t\t\t",round(Vgmv,5),"\n")
cat("s\tt\t\t\t\t\t",s,"\n")
cat("s1\tt\t\t\t\t\t",s.hat1,"***\n")
cat("s2\tt\t\t\t\t\t",s.hat2,"\n")
}
if(doplot){
if(!addplot){
plot(0,0,cex=0,xlim=c(0,range(v)[2]),ylim=range(r),ylab=expression(mu),xlab=expression(sigma))
text(-range(v)[2],range(r)[2],paste("Vmgv =",round(Vgmv,4),"Rgmv=",round(Rgmv,4),
"s=",round(s,4)),pos=4)
lines(x=v,y=r,cex=.25)
}
lines(x=vUB,y=r,cex=.25,col=color)
}
```

```
return=list(V=Vgmv,R=Rgmv,S=c(s,s.hat1,s.hat2,s.hat3),mu=mu,Sigma=Sigma,V.UB=vUB,R.UB=r)
}
#-------------------------------------------------------------------------------------------------------------------
--------#
EF.sample=function(dati, npoints=100, Sigma.diag=T, RGMV=NULL, VGMV=NULL, rmin=NULL,
doprint=F, maxv=0.1){
#-------------------------------------------------------------------------------------------------------------------
--------#
# dati = matrix of log-returns
# npoints =number of points on which to calculate the EF
# Sigma.diag = forces to zero the off-diagonal elements of the matrix
# RGMV = enables to set the minimum value of returns on which to calculate EF
# VGMV = enables to set the minimum value of variance on which to calculate EF
# rmin = computes EF imposing mu or sigma values
# doprint = if TRUE prints range r
# maxv = upper limit of the plot
#-----------------------------------------------------------------------
k=dim(dati)[2]
n=dim(dati)[1]
mu=apply(dati,2,mean,na.rm=T)
if(Sigma.diag){
Sigma=diag(apply(dati, 2, var, na.rm=T))
}else{
Sigma=cov(dati)
}
if(det(Sigma)>0){
S1=solve(Sigma)
}else{
S1=pseudoinverse(Sigma)
}
one=rep(1,dim(dati)[2])
#-----------------------------------------------------------------------
# Sample efficient frontier
#-----------------------------------------------------------------------
```

```
Vgmv=1/(t(one)%*%S1%*%one) # formula 1.3 p.1066
Rgmv=(t(one)%*%S1%*%mu)*Vgmv # formula 1.3 p.1066
R=S1-(S1%*%one%*%t(one)%*%S1)/c(t(one)%*%S1%*%one) # formula 1.2 p.1066
s=c(t(mu)%*%R%*%mu) # formula 1.2 p.1066
RGMV=ifelse(is.null(RGMV),Rgmv,RGMV)
if(is.null(rmin)){
r=seq(from=Rgmv,to=0.1,length=npoints)
# v=seq(from=0,to=0.4,length=npoints)
}else{
r=seq(from=rmin,to=0.1,length=npoints)
# v=seq(from=0,to=0.4,length=npoints)
}
v=1/s*(r-Rgmv)^2+Vgmv # formula 2.2 p. 1066 (uguagliata a 0)
# r=Rgmv+sqrt(s*(v-Vmgv))
return=list(V=v, R=r) # gives back parameters for the construction of EF sample
}
block.boot=function(dati, RR, gprob, check=F){
#-------------------------------------------------------------------------------------------
# dati = matrix of values
# R = number of bootstrap resampling
# gprob =parameter for block reinizialization
#-------------------------------------------------------------------------------------------
n=nrow(dati)
k=ncol(dati)
yboot=array(NA,dim=c(n,k,RR))
change=rep(1,RR)
r=1
while(r <= RR){
loc = round(runif(1,1,n)) #finds first block element
for (i in 1:n){
g1 = runif(1,0,1)
# In probability gprob, we take next observation, otherwise we start a new block
if(g1>gprob){
```

```
loc = loc+1 # continues with the block
}else{
loc = round(runif(1,1,n)) # re-initializes the block
change[r]=change[r]+1
}
if(loc>n)
loc = loc-n # wrap the series as a circule
yboot[i,,r] = as.numeric(dati[loc,])
#-------------------------------------------------
# check correlation structure
#-------------------------------------------------
}
if(check){
Sigma=cov(as.matrix(yboot[,,r]))
if(det(Sigma)>0){
S1=solve(Sigma)
mu=apply(yboot[,,r],2,mean,na.rm=T)
one=rep(1,dim(dati)[2])
R=S1-(S1%*%one%*%t(one)%*%S1)/c(t(one)%*%S1%*%one)
if(c(t(mu)%*%R%*%mu) > (k-1)/n/((n-k-1)/(n-1)))
r=r+1
}
}else{
r=r+1
}
}
return(list(x=yboot,ch=change))
}
surrog=function(dati, R, fft=T, amplitude=T){
#-----------------------------------------------------------------------------------------
# dati = matrix of data
# R = number of boostrap resampling
# gprob = parameter for block reinitialization
```

```
#-------------------------------------------------------------------------------------
n=nrow(dati)
m=ncol(dati)
yboot=array(dim=c(n,m,R))
for (r in 1:R){
for (i in 1:m){
yboot[,i,r] = surrogate(dati[,i],ns=1,fft=T,amplitude=T)
}
}
return(x=yboot)
}
```

Appendix F. The bootstrap method

The bootstrap was introduced in 1979 as a computer-based method for estimating the standard error of the estimate $\hat{\theta}$ of a quantity of interest, θ.

Now bootstrapping is intended to be the practice of estimating properties of an estimator by measuring those properties when sampling from an approximating distribution. One standard choice for an approximating distribution is the empirical distribution function of the observed data. In the case where a set of observations can be assumed to be from an independent and identically distributed population, this can be implemented by constructing a number of resamples with replacement, of the observed dataset (and of equal size to the observed dataset).

It may also be used for constructing hypothesis tests. It is often used as an alternative to statistical inference based on the assumption of a parametric model when that assumption is in doubt, or where parametric inference is impossible or requires complicated formulas for the calculation of standard errors.

Bootstrap methods depend on the notion of a bootstrap sample. Suppose we have observed a random sample $\mathrm{x} = (x_1, x_2, ..., x_n)$ from an unknown probability distribution F. Suppose also that we wish to estimate a parameter of interest $\theta = t(F)$ on the basis of x. We thus calculate an estimate $\hat{\theta} = s(\mathrm{x})$ from x. Let $\hat{F}$ be the empirical distribution, putting probability $1/n$ on each of the observed values . A bootstrap sample is defined to be a random sample of size n drawn form $\hat{F}$, say $\mathrm{x}^* = (x_1^*, x_2^*, ..., x_n^*)$, , where the star notation indicates that x^* is not the actual dataset x, but rather a randomized, or resampled, version of x.

When dealing with bootstrap, the first distinction to be made is that between parametric and non-parametric bootstrap. When there is a particular mathematical model, with adjustable constants or parameters that fully determine F, such a model is called parametric and statistical methods based on this model are parametric methods. When no such mathematical model is used, the statistical analysis is nonparametric, and uses only the fact that the random variables x are independent and identically distributed. However, it is still possible to apply bootstrap resampling techniques to data which are dependent, it is the case of block bootstrap.

Block bootstrap is a parametric bootstrap, applied to dataset characterized by correlation. In this case, a simple resampling will fail, as it will not able to replicate the correlation in the data. The block bootstrap tries to replicate the correlation by resampling instead blocks of data. The block bootstrap has been used mainly with data correlated in time but can also be used with data correlated in space, or among groups (so-called cluster data).

In our specific case, we will use parametric bootstrap method for the construction of confidence intervals.

Acknowledgments

I would like to sincerely thank my supervisors, professor Marco Corazza and professor Claudio Pizzi whose expertise and assistance have guided me throughout the course of my research and have contributed in making this study possible. They have been there whenever a doubt or a problem have arisen, always available to support me.

Besides my supervisors, I wish to thank everyone who helped me complete this dissertation and my studies, and who have contributed to make me grow into the person I am today.

I would like to express my gratefulness to my dear venetian housemates, who have been my second family for the last two years. Their door is always open with a hot cup of tea waiting for me on the table.

I would like to thank other special persons I have had the pleasure to meet during my venetian studies; we have been together through good and bad times, through long nights partying and very long days studying. A special mention goes to Luca, with his patience he has been able to be by my side throughout the completion of this thesis and I recognize it has been a hard work. Thanks for having helped me finding solutions when problems came up, encouraged me and not let me give up; your presence has been irreplaceable.

A big thank goes also to those people who I have met during my bachelor's studies, those have seriously been busy and funny days and I am grateful to have had the possibility to share with you my Buenos Aires experience.

Thanks to my long-time friends, which every time I come back home are there waiting for me with their arms, heart, and ears wide open to listen, comfort and encourage me to do my best. It takes courage to grow up and become who you really are, but with them by my side it has been easier. Thanks for having taught me that it is not how much time you spend with a person, but the kind of time you spend with her that matters.

Last but not the least, I would like to thank my family: my parents, for always trying to understand my choices and supporting me no matter what; my brother, for always remembering me where I truly belong; my grandma and my grandpa, always ready to fondly feed my stomach and, most important of all, my soul. Without your continued efforts and support, I would have not been here.

References

Aït-Sahalia, Y. and Hansen, L. (2010). Handbook of financial econometrics tools and techniques. Amsterdam: North-Holland/Elsevier.

Bachelier, L. (1900). Théorie de la spéculation. Annales scientifiques de l'École Normale Supérieure, 17, pp.21-86.

Basak, G., Ma, T. and Jagannathan, R. (2005). Assessing the Risk in Sample Minimum Risk Portfolios. Working Paper, Northwestern Univerisity.

Best, M. and Grauer, R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. Rev. Financ. Stud., 4(2), pp.315-342.

Bodnar, O. and Bodnar, T. (2009). Statistical inference procedure for the mean–variance efficient frontier with estimated parameters. AStA Advances in Statistical Analysis, 93(3), pp.295-306.

Bodnar, O. and Bodnar, T. (2010). On the unbiased estimator of the efficient frontier. Int. J. Theor. Appl. Finan., 13(07), pp.1065-1073.

Bodnar, T. and Schmid, W. (2009). Econometrical analysis of the sample efficient frontier. The European Journal of Finance, 15(3), pp.317-335.

Borges, B. and Knetsch, J. (1998). Tests of market outcomes with asymmetric valuations of gains and losses: Smaller gains, fewer trades, and less value. Journal of Economic Behavior & Organization, 33(2), pp.185-193.

Borges, B. and Knetsch, J. (1998). Tests of market outcomes with asymmetric valuations of gains and losses: Smaller gains, fewer trades, and less value. Journal of Economic Behavior & Organization, 33(2), pp.185-193.

Britten-Jones, M. (1999). The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. The Journal of Finance, 54(2), pp.655-671.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. Annals of Operations Research, 45(1), pp.21-58.

Corazza, M. and Pizzi, C. (2014). The unbiased efficient frontier à la Bodnar & Bodnar: Some insights. MAF 2014, Mathematical and statistical methods for actuarial sciences and finance. Presentation, Ca' Foscari University.

Davison, A. and Hinkley, D. (1997). Bootstrap methods and their application. Cambridge University Press.

DeMiguel, V., Garlappi, L. and Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/ N Portfolio Strategy?. Rev. Financ. Stud., 22(5), pp.1915-1953.

Dickinson, J. (1974). The Reliability of Estimation Procedures in Portfolio Analysis. The Journal of Financial and Quantitative Analysis, 9(3), p.447.

Efron, B. and Tibshirani, R. (1994). An introduction to the bootstrap. New York: Chapman & Hall.

Fama, E. (1976). Foundations of finance. New York: Basic Books.

Frost, P. and Savarino, J. (1986). An Empirical Bayes Approach to Efficient Portfolio Selection. The Journal of Financial and Quantitative Analysis, 21(3), p.293.

Gelman, A. (2004). Bayesian data analysis. Boca Raton, Fla.: Chapman & Hall/CRC.

Gouriéroux, C. and Jouneau, F. (1999). Econometrics of efficient fitted portfolios. Journal of Empirical Finance, 6(1), pp.87-118.

Gourieroux, C. and Monfort, A. (2005). The econometrics of efficient portfolios. Journal of Empirical Finance, 12(1), pp.1-41.

Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37(3), p.424.

Hamby, D. (1994). A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment, 32(2), pp.135-154.

Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. The Journal of Finance, 58(4), pp.1651-1683.

Jarrow, R., Maksimovic, V. and Ziemba, W. (1995). Finance. Amsterdam: Elsevier.

Jobson, J. (1991). Confidence regions for the mean-variance efficient set: An alternative approach to estimation risk. Rev Quant Finan Acc, 1(3), pp.235-257.

Jobson, J. and Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. Journal of the American Statistical Association, 75(371), p.544.

Kan, R. and Smith, D. (2008). The Distribution of the Sample Minimum-Variance Frontier. Management Science, 54(7), pp.1364-1380.

Levy, H. and Markowitz, H. (1979). Approximating expected utility by a function of mean and variance. The American Economic Review, 69, pp.308-317.

Liu, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data?. J. Appl. Econ., 24(4), pp.560-582.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. The Journal of Business, 36(4), p.394.

Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), p.77.

Markowitz, H. (1991). Portfolio selection. Cambridge, Mass.: B. Blackwell.

Merton, R. (1972). An Analytic Derivation of the Efficient Portfolio Frontier. The Journal of Financial and Quantitative Analysis, 7(4), p.1851.

Michaud, R. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal?. Financial Analysts Journal, 45(1), pp.31-42.

Rubinstein, M. (2002). Markowitz's "Portfolio Selection": A Fifty-Year Retrospective. The Journal of Finance, 57(3), pp.1041-1045.

Saltelli, A. (2004). Sensitivity analysis in practice. Chichester: Wiley.

Saltelli, A. (2008). Global sensitivity analysis. Chichester, England: John Wiley.

Schwartz, E. and Ingersoll, J. (1988). Theory of Financial Decision Making. The Journal of Finance, 43(1), p.259.

Siegel, A. and Woodgate, A. (2007). Performance of Portfolios Optimized with Estimation Error. Management Science, 53(6), pp.1005-1015.

Statman, M. and Scheid, J. (2008). Correlation, Return Gaps, and the Benefits of Diversification. The Journal of Portfolio Management, 34(3), pp.132-139.

Szegö, G. (2005). Measures of risk. European Journal of Operational Research, 163(1), pp.5-19.

Tu, J. and Zhou, G. (2004). Data-generating process uncertainty: What difference does it make in portfolio decisions?. Journal of Financial Economics, 72(2), pp.385-421.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. Journal of Financial Economics, 99(1), pp.204-215.