UNIVERSITÉ DE GENÈVE FACULTÉ DES SCIENCES
Section de mathématiques

A New Portfolio Optimization Based on Entropy.

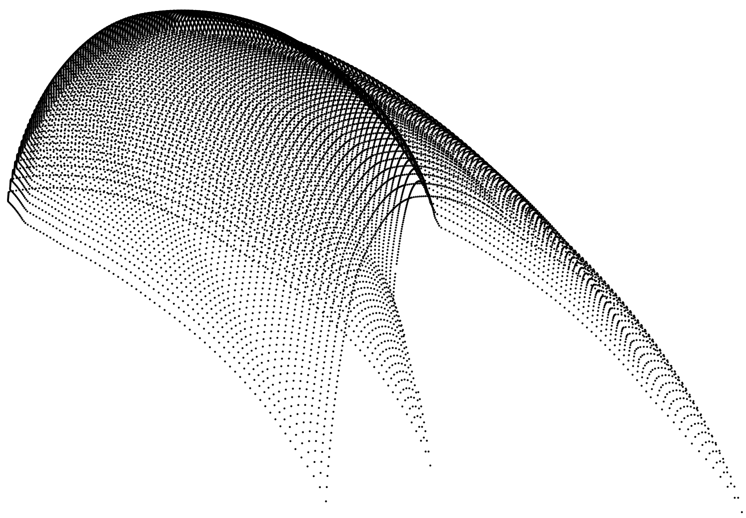
J. Rodriguez

Master thesis under the supervision of
M. Gander (University of Geneva, Section of Mathematics)
M. Gilli (University of Geneva, Department of Econometrics)

Atelier de reproduction de la Section de physique
Geneva, 23th January 2007

Abstract

If nothing or very little is known about a stock market or if there is hardly any trustworthy historical data and one can not estimate a reliable set of parameters in order to construct a consistent probability model for the stock returns or, simply, one is not convinced by the analyzed economical and financial information necessary for decision making, is there a function that takes into account this lack of information and helps us build optimized portfolios?



1 Introduction

In 1952, H. Markowitz published his now famous article *Portfolio Selection* [7], which describes a method to construct portfolios if one has a certain amount of information gathered in two well known statistical parameters, the expected return and the variances and covariances for each stock return. His approach consists of two stages: the first stage is to determine these statistical parameters. As Markowitz says, he does not treat this part in depth. The second stage starts with this estimated information and proposes a rule to construct optimized portfolios.

In our following discussion, we will present and discuss the second stage of Markowitz's approach. This is done in the section 2. At the end of section 2, we will briefly discuss the difficulties one encounters in the first stage Markowitz's approach, and the limitations of how Markowitz defined a tool that he introduced in Stage 2: the risk measure. In the third section, we will discuss another risk measure and in the forth section we suggest an alternative method to construct optimized portfolios when Markowitz's Stage 1 is far from being accurate. To do so, we introduce Shannon's entropy, first in an intuitive manner, second in an axiomatic structure, but both in a financial environment. We then propose our rule, where the notion of entropy plays an important role for determining the portfolio in question. In Section 5, we will show some numerical simulations with portfolios formed with stocks from the Bombay market, stocks chosen from several central-eastern European stock markets, stocks from the Dow Jones Industrial Average Index, stocks from the Shanghai Stock Exchange and stocks from the the Swiss market. We conclude in Section 6.

2 Classical approach: MARKOWITZ

Who controls the past, controls the future: who controls the present controls the past. [1]

We suppose that an economic agent, wanting to invest a certain amount x (in monetary units) into the stock market, knows which market (or markets) he wants to invest in and the let N be the number of stocks to be analyzed and eventually chosen to be part of his portfolio.

[1] G. Orwell, *1984*

This subset is known as the `universe`. We make the assumption that the values of the stocks are not calculated in a deterministic manner, but form a stochastic process, '*since the future is not known with certainty*' [2]. Let $p_{i,t} \in \mathbb{R}_{\geqslant 0}$ be the value of the i^{th} stock at time t. Then one is interested in the behavior and properties of the N-dimensional stochastic process

$$\{\mathcal{P}_t\}_{t\in T},$$

where $T \subset \mathbb{R}_{\geqslant 0}$ is a certain time interval determined in advance and, for all $t \in T$,

$$\begin{array}{rcl} \mathcal{P}_t: \ (\Omega, \mathfrak{A}, \mathbb{P}) & \longrightarrow & (\mathbb{R}^N_{\geqslant 0}, \mathbb{B}_{\mathbb{R}^N_{\geqslant 0}}, P_{\mathcal{P}_t}) \\ \omega & \longmapsto & \mathcal{P}_t(\omega) = (p_{1,t}, \ldots, p_{N,t}), \end{array}$$

where Ω is a sample space, $\mathfrak{A}$ a σ-algebra and $\mathbb{P}$ is a probability measure on $(\Omega, \mathfrak{A})$, and $\mathbb{B}_{\mathbb{R}^N_{\geqslant 0}}$ is the Borel σ-algebra on $\mathbb{R}^N_{\geqslant 0}$ and $P_{\mathcal{P}_t}$ is the distribution measure of the random vector $\mathcal{P}_t$. Although this is the random process that is responsible for all the uncertainty, in what we will discuss, usually, one focuses more on another, very linked, random process: the *returns*, during a certain time $\tau \geqslant 0$, of the stocks. These are defined in

Definition 1. *The* `return` $r_{i,\tau,t}$ *for the* i^{th} *stock at time* t*, over the elapsed time* $\tau \geqslant 0$ *is*

$$r_{i,\tau,t} := \frac{p_{i,t}}{p_{i,t-\tau}}, \ i = 1, \ldots, N \text{ and } t \in T \ .$$

Usually, one is interested in daily, weekly, monthly, quarterly and/or yearly returns. Therefore, it is natural to see τ with one of these time lengths. From now on, except otherwise stated, we will focus on monthly returns. Hence, we remove the index τ work on a discret time basis, $T := \{1, \ldots, n\}$.

To continue our exposition, we need to introduce some definitions and assumptions. The first assumption is very strong: we will suppose, from now on, that the distribution measure (or law) of the returns does not depend on $t \in T$, that is, the random vector r of the returns is

$$\begin{array}{rcl} r: \ (\Omega_r, \mathfrak{A}_r, \mathbb{P}_r) & \longrightarrow & (\mathbb{R}^n_{\geqslant 0}, \mathbb{B}_{\mathbb{R}^N_{\geqslant 0}}, P_r) \\ \omega & \longmapsto & r(\omega) = y, \end{array}$$

[2] Refer to [7].

where Ω_r is the sample space of r, $\mathfrak{A}_r$ the σ-algebra and $\mathbb{P}_r$ is the probability measure on $(\Omega_r, \mathfrak{A}_r)$, and P_r is the distribution measure of the random vector r. To avoid overloading the notation, we will suppose that we are at a certain predefined period of our evolving process, say $t \in T$ is $t = j^* \in \mathbb{Z}_{\geqslant 0}$ for a fixed j^*. All of the following definitions are easily extendable for all $t \in T$.

Definition 2. *The* `budget equation` *for an investor with initial wealth $x > 0$, assuming to hold $X_i \geqslant 0$ shares of stock $i, i = 1, \ldots, N$, is*

$$\sum_{i=1}^{N} X_i p_i = x.$$

Definition 3. *The* `portfolio weight` *vector $\pi = (\pi_1, \ldots, \pi_N)$ is defined by*

$$\pi_i := \frac{X_i p_i}{x}, i = 1, \ldots, N.$$

Definition 4. *The* `portfolio return`*, for a portfolio weight vector π, is defined by*

$$r^{\pi} := \sum_{i=1}^{N} \pi_i r_i.$$

For Definition 2 to be coherent, one has to assume that each stock is perfectly divisible, that is, one can hold $X_i \in \mathbb{R}_{\geqslant 0}$ shares of stock $i, i = 1, \ldots, N$. Also from Definition 2, we do not allow short selling, that is $X_i \nless 0 \ \forall \ i \in \{1, \ldots, N\}$. We also assume that each random variable r_i is integrable with respect to it's probability measure, that is $\mathbb{E}[r_i] =: \mu_i < +\infty$, which is nothing more than the expected return for the stock i.

We can now define the following important quantity:

Definition 5. *The* `portfolio expected return` *is defined by*

$$\mathbb{E}[r^{\pi}] := \sum_{i=1}^{N} \pi_i \mu_i = \langle \pi | \mu \rangle,$$

where $\pi = (\pi_1, \ldots, \pi_N)$ is the portfolio weight vector, $\mu = (\mu_1, \ldots, \mu_N)$ is the vector containing the expected returns of each stock and $\langle \cdot | \cdot \rangle$ represents the usual scalar product.

Because of the randomness of the value of the stocks, the value of the budget equation may be larger or smaller than x after a certain time. Therefore, since the value of the budget equation being smaller than x is unwanted, there exists a notion of 'risk' when one invests an amount x in the stock market. Automatically, the question of how to measure 'risk', arises. If an answer is given, then, with the above definitions at hand and assuming the choice of the market and a subset of stocks of this market is done, our aim is to answer the following question:

Out of the total amount x, what are the weights $\pi_1, \ldots, \pi_N$ that should be invested in the N stocks respectively if we want to minimize 'risk' but, at the same time, seeking a portfolio expected return greater or equal to p (p being given in advance, and p is to be seen as a return greater or equal to the return of a risk free investment)?

There is no absolute answer to this question, and the answer will of course depend on a very large number of variables, spreading from economical, social and historical context from one side of the spectrum, to the individual himself, his knowledge and information about markets and his adversity to risk and confidence on his decisions, on the other side of the spectrum.

To answer this question, we now present the main ideas of the work of H. MARKOWITZ. His article, *Portfolio Selection* [7], begins with '*The process of selecting a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio*'[3].

It is the second stage that we are interested in. In MARKOWITZ's article, the guiding idea of the rule to be applied in the seconde stage is 'diversification'. '*Diversification is both observed and sensible; a rule of behavior which does not imply the superiority of diversification must be rejected both as a hypothesis* [*about investment behavior*] *and as a maxim* [*about investment behavior*]'[4]. Even though it is stage 2 that delivers an answer to our question, there is an important element in stage 1 that we must take into consideration. This concerns the task of calculating the

[3]Refer to [7]
[4]Refer to [7]

parameter set Φ that determines the probabilistic model for the stock returns. MARKOWITZ's approach suggests to consider (μ, σ) as determining parameters where $\mu = (\mu_1, \ldots, \mu_N)$ is the vector containing the expected returns of each stock and $\sigma := (\sigma_{i,j})_{i,j \in \{1,\ldots,N\}}$ is the variance-covariance matrix of the returns for the N stocks, i.e. $\sigma_{i,j} := \mathbb{C}\text{ov}[r_i, r_j]$ is the variance (when $i = j$) and covariance (when $i \neq j$) for the returns of stocks i and j. MARKOWITZ implicitly makes the hypothesis that the returns are square integrable random variables, i.e. $\mathbb{E}[r_i^2] < +\infty \ \forall \ i$. It is with the help of these two parameters that MARKOWITZ presents his 'expected returns - variance of returns' rule, that '*implies diversification for a wide range of $\mu_i, \sigma_{i,j}$*'[5].

MARKOWITZ considers '*the rule that the investor does (or should) consider expected return a desirable thing and variance of returns an undesirable thing*'[6]. He continues by saying that '*the concepts "yield" and "risk" appear frequently in financial writings. Usually if the term "yield" were replaced by "expected yield" or "expected return," and "risk" by "variance of return," little change of apparent meaning would result.*'[7]. Finally, he states that '*variance is a well-known measure of dispersion about the expected*'[8]. Therefore, MARKOWITZ suggests to take the portfolio variance that is defined below as a measure of risk.

[5]Refer to [7]
[6]Refer to [7]
[7]Refer to [7]
[8]Refer to [7]

Definition 6. *The* `portfolio variance` *is defined by*

$$\begin{aligned}
\mathbb{V}ar[r^\pi] &= \mathbb{E}[(r^\pi - \mathbb{E}[r^\pi])^2] = \mathbb{E}[(\sum_{i=1}^N \pi_i r_i - \sum_{i=1}^N \pi_i \mu_i)^2] \\
&= \mathbb{E}[(\sum_{i=1}^N \pi_i (r_i - \mu_i))^2] \\
&= \mathbb{E}[\sum_{i=1}^N \pi_i^2 (r_i - \mu_i)^2 + \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \pi_i \pi_j (r_i - \mu_i)(r_j - \mu_j)] \\
&= \sum_{i=1}^N \pi_i^2 \underbrace{\mathbb{E}[(r_i - \mu_i)^2]}_{=\sigma_{i,i}} + \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \pi_i \pi_j \underbrace{\mathbb{E}[(r_i - \mu_i)(r_j - \mu_j)]}_{\sigma_{i,j}} \\
&= \sum_{i=1}^N \sum_{j=1}^N \pi_i \pi_j \sigma_{i,j} \\
&= \begin{pmatrix} \pi_1 & \dots & \pi_N \end{pmatrix} \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,N} \\ \sigma_{1,2} & \sigma_{2,2} & \cdots & \sigma_{2,N} \\ \vdots & & \ddots & \vdots \\ \sigma_{1,N} & \sigma_{2,N} & \cdots & \sigma_{N,N} \end{pmatrix} \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix} \\
&= \pi^\top \sigma \pi.
\end{aligned}$$

2.1 Selecting the portfolio weights π_i

We are now ready to follow MARKOWITZ's 'expected returns - variance of returns' rule (E-V) to find an optimal portfolio. '*The E-V rule states that the investor would (or should) want to select one of those portfolios which give rise to the (E, V) combinations* [that are] *efficient* [...] *i.e., those with minimum V for given E or more and maximum E for given V or less*'[9]. We will focus on the first part of the rule, i.e. '*those with minimum V for given E or more*'[10]. This can be seen as the following problem: minimize the portfolio variance $\mathbb{V}\mathrm{ar}[r^\pi]$ but, at the same time, the portfolio expected return $\mathbb{E}[r^\pi]$ must be greater or equal than a given lower bound, that is $\mathbb{E}[r^\pi] \geqslant p$.

[9]Refer to [7]
[10]Refer to [7]

Mathematically, this is formulated as:

Problem 1.

$$\min_{\pi \in \mathbb{R}^N} \{\mathbb{V}ar[r^\pi]\}$$

under the constraints

1. $\pi_i \geqslant 0 \ \forall \ i \in \{1, \ldots, N\}$
2. $\sum_{i=1}^{N} \pi_i = 1$
3. $\mathbb{E}[r^\pi] \geqslant p$ *(for a given* p*)* .

In other words

among all the theoretically possible portfolios $\pi \in \mathbb{R}$, we only consider the portfolios that satisfy the constraints, and within these portfolios, we have to determine the one with the smallest risk (i.e. the smallest variance).

Note: The type of portfolio management where we solve Problem 1 for each monthly period, will be called `Markowitz` portfolio management type.

2.2 Solving the problem

Problem 1 is a quadratic optimization problem. This problem can be solved by using standard quadratic programming algorithms. We will not discuss these algorithms here. For our numerical experiments we will use the quadratic optimization problem solver from Matlab. However, two important questions must be answered:

I When is it possible to find a solution?

II Under what conditions do we have a unique solution, i.e. when do we have a unique π^* that solves Problem 1?

Let us answer the first question. $\mathbb{V}\text{ar}[r^\pi]$ is continuous with respect to π, therefore it will reach a minimum over all compact sets of $\mathbb{R}^N$. The three constraints in Problem 1 define closed sets in $\mathbb{R}^N$ and therefore their intersection is closed. Together, constraints 1) and 2) define a bounded set, hence the three constraints form a closed bounded set in $\mathbb{R}^N$, thus compact, and so, it is not difficult to see that for a solution to exist, it must be in the feasible set, that is, it must satisfy the constraints in Problem 1, i.e., p must be well chosen.

Concerning the second question, if the matrix σ is positive definite and invertible, then there is one and only one π that solves Problem 1 (refer to Appendix A).

This concludes Stage 2. This can be summarized by the following diagram:

<u>STAGE 1</u>	$\longrightarrow$	<u>STAGE 2</u>	$\longrightarrow$	<u>OUTPUT OF RULE</u>
Information and analysis Determining μ and σ		Rule with constraints $\sum_{i=1}^{N} \pi_i = 1$ $\pi_i \geqslant 0,\ i = 1, \ldots, N$ $\mathbb{E}[r^\pi] \geqslant p$		Optimized portfolio $\pi^0 = (\pi_1, \ldots, \pi_n)$

If we work with three stocks, we can give a geometric illustration of the situation. We will proceed in the same way as MARKOWITZ did in [7, pp 83]. The main idea is that one can express the variable π_3 in function of the other two (π_1 and π_2) with the help of the third constraint (i.e. $\sum_{i=1}^{3} \pi_i = 1$). We can therefore work in two dimensional geometry. We have chosen three different stocks

I `Český Telecom`, a telecommunication operator quoted on the Prague Stock Exchange,

II `Mol Magyar Olaj- ES Gazipari`, an oil and gas company quoted on the Budapest Stock Exchange,

III `Millennium Bank`, a bank quoted on the Warsaw Stock Exchange.

Figure 2.1: Company symbols of the three stocks.

In diagram below, the feasible set is the triangle $\overline{(0,0)(1,0)}, \overline{(1,0)(0,1)}, \overline{(0,1)(0,0)}$ (black segments). In blue, we have the contour lines of the portfolio variance $\mathbb{V}\mathrm{ar}[r^\pi]$, where we have made the adequate substitution: $\pi_3 = 1 - \pi_1 - \pi_2$. Analytically, this gives us the two variable function

$$\begin{aligned} V(\pi_1, \pi_2) &= \pi_1^2(\sigma_{1,1} - 2\sigma_{1,2} + \sigma_{3,3}) + \pi_2^2(\sigma_{2,2} - 2\sigma_{2,3} + \sigma_{3,3}) \\ &\quad +2\pi_1\pi_2(\sigma_{1,2} - \sigma_{1,3} - \sigma_{2,3} + \sigma_{3,3}) + 2\pi_1(\sigma_{1,3} - \sigma_{3,3}) \\ &\quad +2\pi_2(\sigma_{2,3} - \sigma_{3,3}) + \sigma_{3,3}, \end{aligned}$$

where $\sigma_{i,j}$ are the components of the variance-covariance matrix of the returns. The small black dots as well as the small black circles represent the contour lines of the portfolio expected return $\mathbb{E}[r^\pi]$ with the same substitution as above. This gives us, analytically, the two variable function

$$E(\pi_1, \pi_2) = \mu_3 + \pi_1(\mu_1 - \mu_3) + \pi_2(\mu_2 - \mu_3),$$

where μ_i are the expected returns for each stock. To not over complicate the diagram, we have shown only two contour lines: $E = 1.5\%$ and $E = 2\%$. The thick blue point shows the respective weights π_1 and π_2 for the stocks of Česk Telecom and Mol and their weights were calculated by solving Problem 1 with $p = 2\%$ (i.e. asking for a 2% or more monthly return). For this, we have taken a sample set of 84 monthly returns, dating from January 1999 to December 2005.

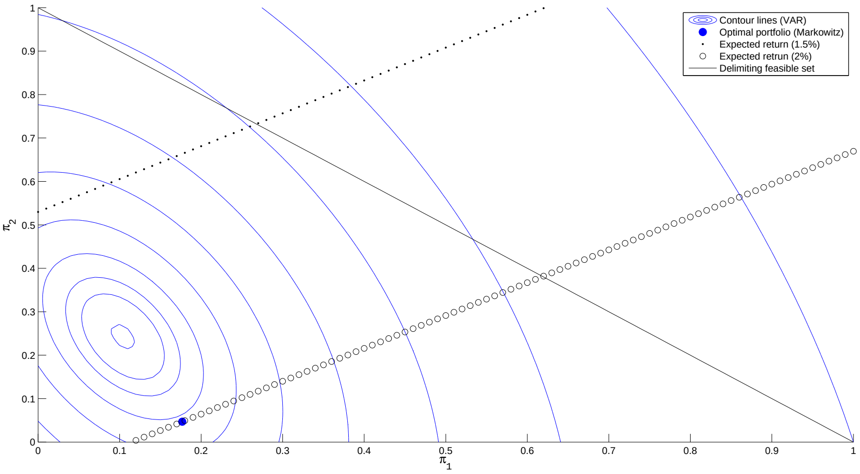


Figure 2.2: Contour lines of the portfolio variance and expected return with the optimal portfolio.

2.3 Difficulties and limitations

What about the first stage? We must deal with it in some way or another, if we want to determine numerical optimized portfolios. Although MARKOWITZ, in a footnote, writes ...

> ‘*This paper does not consider the difficult question of how investors do (or should) form their probability beliefs* [...] (μ_i, σ_{ij}) ’[11].

and in his ending paragraphe he stresses the point that ...

> ‘*In this paper we have considered the second stage in the process of selecting a portfolio. This stage starts with the relevant beliefs about the securities involved and ends with the selection of a portfolio. We have not considered the first stage: the formation of the relevant beliefs on the basis of observation*’[12].

he does, however, say that ...

> ‘*One suggestion as to tentative μ_i, σ_{ij} is to use the observed μ_i, σ_{ij} for some period of the past.*’[13]

[11] Refer to [7]
[12] Refer to [7]
[13] Refer to [7]

However, is the past a reliable source of information? How 'long' is the past, i.e. how much data has to be collected? Last, but not least, is the portfolio variance a 'good' measure of risk? Large positive returns, for a certain stock, would be considered, by the investor, as a positive property. However, this property will make the stock's variance relatively big and this has a negative impact when it comes to minimizing the portfolio variance. In the next section, we introduce another risk measure from the literature.

3 A more recent approach: CVaR

When one invests in stock markets, it is almost inevitable to encounter temporary losses, but large losses are definitely unwanted. Therefore, one is interested in studying the losses to see if it is possible to define 'risk' in another way than only regarding the portfolio variance as a risk measure. One can interpret 'risk' by saying that a 'risky' investment is one that might have large losses. One way to quantify losses is to proceed in the following manner.

We suppose that a random vector r of the returns has a density ρ, that is, by definition,

$$\rho : \mathbb{R}^N_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}$$

is a positive Borel measurable function verifying

$$P_r(A) = \int_A \rho(y)\, dy \qquad \forall\ A \in \mathbb{B}_{\mathbb{R}^N_{\geqslant 0}}.$$

Let $\Pi \subset \mathbb{R}^N$ be the set of the available portfolios, i.e. $\Pi := \{\pi \in \mathbb{R}^N | \pi_i \geqslant 0,\ i = 1, \ldots, N$ and $\sum_{i=1}^N \pi_i = 1\}$ and the i^{th} coordinate corresponds to the weight in the portfolio of the i^{th} stock. We consider a function

$$\begin{array}{rccl} f: & \Pi \times \mathbb{R}^N_{\geqslant 0} & \longrightarrow & \mathbb{R} \\ & (\pi, y) & \longmapsto & f(\pi, y) := 1 - \langle \pi | y \rangle \end{array}$$

that quantifies the loss associated with the portfolio π, where y is the vector containing a sample of returns. Since y is the image of the random vector r, then, for a fixed π, $f(\pi, \cdot)$ is a random variable, that is

$$\begin{array}{rccl} f(\pi, \cdot): & (\mathbb{R}^N_{\geqslant 0}, \mathbb{B}_{\mathbb{R}^N_{\geqslant}}, P_r) & \longrightarrow & (\mathbb{R}, \mathbb{B}_{\mathbb{R}}, P_{f(\pi,\cdot)}) \\ & y & \longmapsto & 1 - \langle \pi | y \rangle, \end{array}$$

where $\mathbb{B}_{\mathbb{R}}$ is the Borel σ-algebra on $\mathbb{R}$ and $P_{f(\pi,\cdot)}$ is the distribution measure of the random variable $f(\pi, \cdot)$.

We are interested in the probability of not exceeding a loss of $\alpha > 0$. This is given by

$$P_{f(\pi,\cdot)}((-\infty, \alpha]) = P_r(\{y \in \mathbb{R}^N_{\geqslant 0} | f(\pi, y) \in (-\infty, \alpha]\}) = \int\limits_{f(\pi,y)\leqslant\alpha} \rho(y)\, dy$$

Therefore, the last term in this equality, can be seen as

$$\begin{aligned} \Psi : \ \Pi \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (\pi, \alpha) &\longmapsto \int\limits_{f(\pi,y)\leqslant\alpha} \rho(y)\, dy. \end{aligned}$$

By definition, for a fixed π, $\Psi(\pi, \cdot) : \mathbb{R} \longrightarrow [0, 1]$, is the cumulative distribution function (CDF) for the losses associated with the portfolio π. It is nondecreasing with respect to α. We assume that the CDF is continuous with respect to α, implying that the set of y with $f(\pi, y) = \alpha$ has probability zero, i.e.,

$$\int\limits_{f(\pi,y)=\alpha} \rho(y)\, dy = 0$$

3.1 Defining Conditional Value-at-Risk (CVaR)

Definition 7. *For a fixed $\pi \in \Pi$, for a fixed $\beta \in]0, 1[$ that is to be seen as a probability level, the VaR and CVaR for the loss $f(\pi, \cdot)$ are denoted by $\alpha_{\beta,\pi}$ and $\phi_{\beta,\pi}$ respectively and they are defined by*

$$\alpha_{\beta,\pi} = \min\{\alpha \in \mathbb{R} | \Psi(\pi, \alpha) \geqslant \beta\}$$

$$\phi_{\beta,\pi} = (1 - \beta)^{-1} \int\limits_{f(\pi,y)\geqslant\alpha_{\beta,\pi}} f(\pi, y)\rho(y)\, dy.$$

Remarks and interpretations

I $\alpha_{\beta,\pi}$ exists. This is because we assumed that $\Psi(\pi, \cdot)$ is continuous and nondecreasing with respect to α so there must be at least one

α_0 such that $\Psi(\pi, \alpha_0) = \beta$. Since $\Psi(\pi, \cdot)$ is not necessarily a strictly increasing function, but it is continuous with the properties that

$$\lim_{\alpha \to \infty} \Psi(\pi, \alpha) = 1 \quad \text{and} \quad \lim_{\alpha \to -\infty} \Psi(\pi, \alpha) = 0,$$

the set of $\alpha \in \mathbb{R}$ such that $\Psi(\pi, \alpha) = \beta$ form a bounded interval, that is closed because $\Psi(\pi, \cdot)$ is continuous. In this case, $\alpha_{\beta,\pi}$ is the left endpoint of this interval.

II VaR is the lowest amount $\alpha_{\beta,\pi}$ such that, with probability β, the loss will not exceed $\alpha_{\beta,\pi}$. In other words, $\alpha_{\beta,\pi}$ can be seen in the following manner: for the portfolio π, there is $(1 - \beta)$ probability of losing $\alpha_{\beta,\pi}$ or more.

III $\phi_{\beta,\pi}$ is the conditional expectation of losses above the amount $\alpha_{\beta,\pi}$. In other words, the probability average (expected value) within the losses that have less than $(1 - \beta)$ probability of occurring, is $\phi_{\beta,\pi}$.

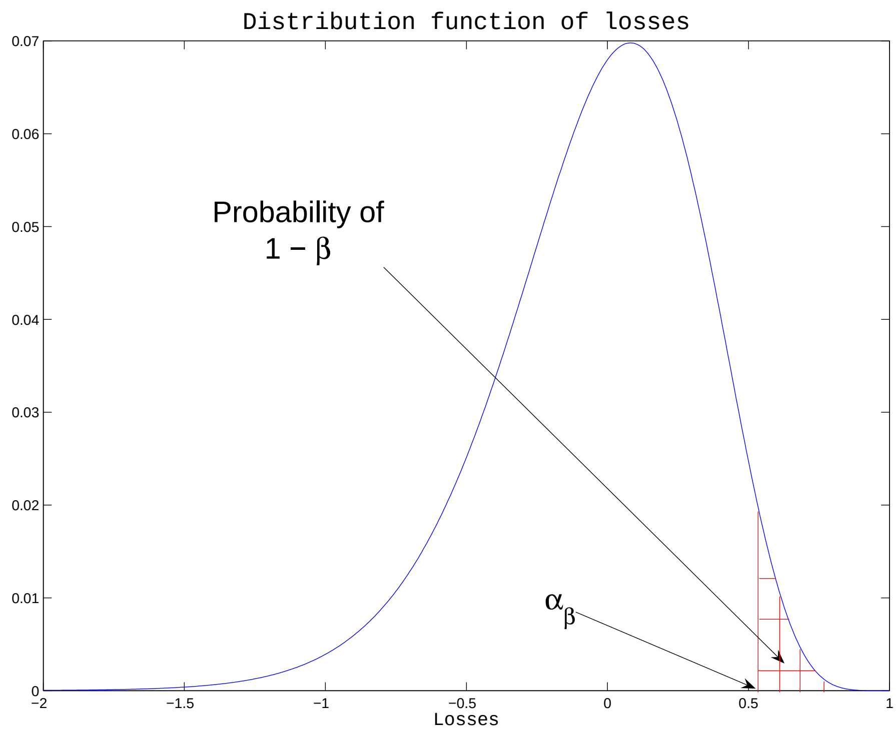


Figure 3.1: Example of a loss distribution with its VaR indicated

With these remarks and interpretations, it is fair to say that CVaR can be seen as a measure of risk. VaR can also be seen as a risk measure. We will not deal with the difficulties in calculating it, which is beyond the scope of this project. However, already from a theoretical point of view, VaR has one main disadvantage compared to CVaR. It may happen that a portfolio has a small VaR but its distribution is such that it may still have very disastrous events (losses of 60% and over, for example) with relatively high probability of occurring. This is not captured in the single real number $\alpha_{\beta,\pi}$. On the other hand, CVaR is well adapted to this scenario since, in this case, the CVaR will be relatively large: being an expected value, the disastrous events with high probability of occurring will have a strong impact on the CVaR and thus the CVaR is sensitive to this circumstance.

We would like to have CVaR as a convex function of π, which would be coherent with the idea of diversifying the assets and not placing the entire investment into one stock. We would also want an efficient method to find, among all the portfolios that we are interested, the portfolio that minimizes the CVaR (i.e., the portfolio with minimum 'risk'). In particular, the method should avoid approximating the density function ρ, which in practice is rarely known.

For this task, we will follow what Rockafellar and Uryasev did in [11]. The idea is to work with the following function:

$$\begin{array}{rccl} F_\beta : & \Pi \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & (\pi, \alpha) & \longmapsto & \alpha + (1-\beta)^{-1} \displaystyle\int_{y \in \mathbb{R}^N_{\geqslant 0}} [f(\pi, y) - \alpha]^+ \rho(y)\, dy \end{array}$$

where

$$[t]^+ \;=\; \begin{cases} t \text{ when } t > 0 \\ 0 \text{ when } t \leqslant 0 \end{cases}$$

This function is well defined (i.e. the integral $\int_{y \in \mathbb{R}^N_{\geqslant 0}} [f(\pi, y) - \alpha]^+ \rho(y)\, dy < +\infty$) if we suppose that $\mathbb{E}[|f(\pi, y)|] < +\infty \ \forall \ \pi \in \Pi$. Since $[f(\pi, y) - \alpha]^+ \rho(y)$ is positive $\forall \ (\pi, \alpha, y) \in \Pi \times \mathbb{R} \times \mathbb{R}^N_{\geqslant 0}$, then to see that the function is well defined, it is sufficient to find an upper bound for the integral $\int_{y \in \mathbb{R}^N_{\geqslant 0}} [f(\pi, y) - \alpha]^+ \rho(y)\, dy$ (i.e. an increasing bounded sequence converges).

$$
\begin{aligned}
0 \;\leqslant\; \int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} [f(\pi,y)-\alpha]^+\rho(y)\,dy \;&\leqslant\; \Big| \int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} [f(\pi,y)-\alpha]^+\rho(y)\,dy \Big| \\
&\leqslant\; \underbrace{\int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} |f(\pi,y)-\alpha|\rho(y)\,dy}_{=\mathbb{E}[|f(\pi,y)|]<+\infty} \;+\; |\alpha| \underbrace{\int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} \rho(y)\,dy}_{=1}\,.
\end{aligned}
$$

The motivation to work with this function comes from the following reasoning: let α be such that $\Psi(\pi,\alpha)=\beta$. Then $P_r(f(\pi,y)>\alpha)=1-\beta$ and

$$
\begin{aligned}
\mathbb{E}[f(\pi,y)|f(\pi,y)>\alpha] &= (1-\beta)^{-1}\int\limits_{f(\pi,y)>\alpha} f(\pi,y)\rho(y)\,dy \\
&= \frac{\mathbb{E}[f(\pi,y)\chi_{\{f(\pi,y)>\alpha\}}(y)]}{P_r(f(\pi,y)>\alpha)} \\
&= \frac{\mathbb{E}[\alpha\chi_{\{f(\pi,y)>\alpha\}}(y)+[f(\pi,y)-\alpha]^+]}{P_r(f(\pi,y)>\alpha)} \\
&= \alpha+(1-\beta)^{-1}\mathbb{E}\big[[f(\pi,y)-\alpha]^+\big]
\end{aligned}
$$

A property of $F_\beta(\pi,\alpha)$ is that it is convex with respect to (π,α). To see this, we will proceed in several steps.

First of all, $\forall\ y\in\mathbb{R}^n_{\geqslant 0}$ fixed, the function $(\pi,\alpha)\longmapsto f(\pi,y)-\alpha$ is affine, and hence convex. Secondly, $\forall\ y\in\mathbb{R}^N_{\geqslant 0}$, the function $(\pi,\alpha)\longmapsto[f(\pi,y)-\alpha]^+$ is convex since it is the composition of the function $(\pi,\alpha)\longmapsto f(\pi,y)-\alpha$ with the nondecreasing convex function $t\longmapsto[t]^+$ and by the following lemma, the composition is convex.

Lemma 1. *Let g be a convex function from $\mathbb{R}^N$ to $\mathbb{R}$ and let φ be a convex function from $\mathbb{R}$ to $\mathbb{R}$ which is nondecreasing. Then $h(x):=\varphi(g(x))$ is convex.*

Proof. For all $x,y\in\mathbb{R}^N$ and $\lambda\in\,]0,1[$, we have by definition

$$
g(\lambda x+(1-\lambda)y)\leqslant\lambda g(x)+(1-\lambda)g(y)
$$

Now φ is nondecreasing, therefore, applying φ to both sides of the inequality yields

$$\begin{aligned} h(\lambda x + (1-\lambda)y) &\leqslant \varphi(\lambda g(x) + (1-\lambda)g(y)) \\ &\leqslant \lambda\varphi(g(x)) + (1-\lambda)\varphi(g(y)) \\ &= \lambda h(x) + (1-\lambda)h(y) \end{aligned}$$

Therefore, h is convex. $\square$

Thirdly, it is clear that $F_\beta(\pi,\alpha) = \alpha + (1-\beta)^{-1}\int_{y\in\mathbb{R}^N}[f(\pi,y)-\alpha]^+\rho(y)\,dy$ is convex with respect to (π,α) whenever the integrand

$$g(\pi,\alpha,y) := [f(\pi,y)-\alpha]^+\rho(y)$$

is convex with respect to (π,α). Defining $G(\pi,\alpha) := \int\limits_{y\in\mathbb{R}^N} g(\pi,\alpha,y)\,dy$ and from what we have seen above, we have, $\forall \pi,\breve{\pi} \in \mathbb{R}^N$,$\forall\ \alpha_1,\alpha_2 \in \mathbb{R}$ and $\lambda \in]0,1[$ that

$$g(\lambda(\pi,\alpha_1)+(1-\lambda)(\breve{\pi},\alpha_2),y) \leqslant \lambda g((\pi,\alpha_1),y)+(1-\lambda)g((\breve{\pi},\alpha_2),y)\ \forall\, y\in\mathbb{R}^N,$$

and by taking the integral over all $\mathbb{R}^N$ with respect to y leads to

$$\begin{aligned} \int\limits_{y\in\mathbb{R}^N} g(\lambda(\pi,\alpha_1)+(1-\lambda)(\breve{\pi},\alpha_2),y)\,dy &\leqslant \lambda \int\limits_{y\in\mathbb{R}^N} g((\pi,\alpha_1),y)\,dy \\ &+ (1-\lambda)\int\limits_{y\in\mathbb{R}^N} g((\breve{\pi},\alpha_2),y)\,dy \end{aligned}$$

which, by definition, is

$$G(\lambda(\pi,\alpha_1)+(1-\lambda)(\breve{\pi},\alpha_2),y) \leqslant \lambda G(\pi,\alpha_1,y) + (1-\lambda)G(\breve{\pi},\alpha_2,y),$$

and thus concludes the proof that $F_\beta(\pi,\alpha)$ is convex with respect to (π,α).

Another property of $F_\beta(\pi,\alpha)$ is that it is continuous with respect to (π,α). We will use the theorem in Appendix B. Let $(\breve{\pi},\breve{\alpha}) \in \Pi\times\mathbb{R}$ and define

$$\mathrm{I}\ \mathcal{N}^M_{(\breve{\pi},\breve{\alpha})} := \{(\pi,\alpha)\in\Pi\times\mathbb{R}|\ \|(\pi,\alpha)\|_1 = \sum_{i=1}^N |\pi_i| + |\alpha| \leqslant M\}$$

II $g(\pi,\alpha,y) := [f(\pi,y)-\alpha]^+\rho(y)$

For a fixed $y \in \mathbb{R}^N$, $g(\pi,\alpha,y)$ is continuous $\forall\ (\pi,\alpha) \in \Pi \times \mathbb{R}$ and we have seen that $g(\pi,\alpha,y)$ is integrable with respect to $y \in \mathbb{R}^N, \forall\ (\pi,\alpha) \in \Pi \times \mathbb{R}$. We must find a function h integrable with respect to y such that

$$\forall\ (\pi,\alpha) \in \Pi \times \mathbb{R}, |g(\pi,\alpha,y)| \leqslant h(y), \forall\ y \in \mathbb{R}^N.$$

For $(\pi,\alpha) \in \mathcal{N}^M_{(\breve{\pi},\breve{\alpha_0})}$, we have

$$\begin{aligned}
|[f(\pi,y)-\alpha]^+\rho(y)| &\leqslant |1-\langle\pi|y\rangle-\alpha|\rho(y)\\
&\leqslant (1+|\langle\pi|y\rangle|+|\alpha|)\rho(y)\\
&\leqslant (1+|\alpha|)\rho(y)+\|\pi\|_2\|y\|_2\rho(y) \quad \text{since Cauchy-Schwarz}\\
&\leqslant (1+|\alpha|)\rho(y)+K^2\|\pi\|_1\|y\|_1\rho(y) \quad \text{since } \|\cdot\|_2 \leqslant K\|\cdot\|_1\\
&\leqslant (1+M)\rho(y)+K^2M\|y\|_1\rho(y)
\end{aligned}$$

Therefore, if we suppose that $\int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} \|y\|_1\rho(y)\,dy < +\infty$, then the function

$$h_{(\breve{\pi},\breve{\alpha_0},M)}(y) := (1+M)\rho(y)+K^2M\|y\|_1\rho(y)$$

is integrable and so the function $(\pi,\alpha) \longmapsto \int\limits_{y\in\mathbb{R}^N} g(\pi,\alpha,y)\rho(y)\,dy$ is continuous $\forall\ (\pi,\alpha) \in \Pi \times \mathbb{R}$. The hypothesis $\int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} \|y\|_1\rho(y)\,dy < +\infty$ is equivalent to ask that $\int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} y_i\rho(y)\,dy < +\infty$, $i = 1,\dots,N$, and this, by definition, is the expected return for the stock i, (i.e. $\mathbb{E}[y_i] < +\infty, i = 1,\dots,N$)(refer to Appendix C). Therefore, in our modeling, we only have to require the existence of the first moment of the returns and no longer the second moment, i.e., for the random vector Y of the returns, we must have $\mathbb{E}[Y] < +\infty$, which means

$$\mathbb{E}[Y] = \int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} y\rho(y)\,dy = \begin{pmatrix} \int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} y_1\rho(y)\,dy \\ \vdots \\ \int\limits_{y\in\mathbb{R}^N_{\geqslant 0}} y_N\rho(y)\,dy \end{pmatrix} \begin{matrix} < +\infty \\ \vdots \\ < +\infty. \end{matrix}$$

The next theorem gives us a useful method to calculate the CVaR for a given portfolio $\pi \in \Pi$.

Theorem 1 (Rockafellar & Uryasev (2000)). *[11, Thm.1]*

I As a function of α, $F_\beta(\pi, \cdot)$ is continuously differentiable.

II Defining the set $A_{\beta,\pi}$ by

$$A_{\beta,\pi} := \{s \in \mathbb{R} | F_\beta(\pi, s) = \min_{\alpha \in \mathbb{R}} \{F_\beta(\pi, \alpha)\}\},$$

we have that $A_{\beta,\pi}$ is a nonempty closed bounded interval.

III For all $\pi \in \Pi$, the CVaR is

$$\phi_{\beta,\pi} = \min_{\alpha \in \mathbb{R}} \{F_\beta(\pi, \alpha)\}$$

Proof. We first have to show that $F_\beta(\pi, \cdot)$ is continuously differentiable with respect to α. For $\pi \in \Pi$, let us compute the limit

$$
\begin{aligned}
&\lim_{h\to 0} \frac{F_\beta(\pi,\alpha+h)-F_\beta(\pi,\alpha)}{h} \\
&\quad = \lim_{h\to 0}\frac{1}{h}\Big[\alpha+h \;+\; (1-\beta)^{-1}\int\limits_{y\in\mathbb{R}^N}[f(\pi,y)-(\alpha+h)]^+\rho(y)\,dy \\
&\qquad\qquad -\;\alpha-(1-\beta)^{-1}\int\limits_{y\in\mathbb{R}^N}[f(\pi,y)-\alpha]^+\rho(y)\,dy\Big] \\
&\quad = \lim_{h\to 0}\frac{1}{h}\Big[\,h \;+\; (1-\beta)^{-1}\Big(\int\limits_{f(\pi,y)\geqslant\alpha+h}(f(\pi,y)-(\alpha+h))\rho(y)\,dy \\
&\qquad\qquad -\int\limits_{f(\pi,y)\geqslant\alpha}(f(\pi,y)-\alpha)\rho(y)\,dy\Big)\Big] \\
&\quad = \lim_{h\to 0}\frac{1}{h}\Big[\,h \;+\; (1-\beta)^{-1}\Big(\int\limits_{f(\pi,y)\geqslant\alpha+h}(f(\pi,y)-\alpha)\rho(y)\,dy \\
&\qquad\qquad -\int\limits_{f(\pi,y)\geqslant\alpha+h}h\rho(y)\,dy-\int\limits_{f(\pi,y)\geqslant\alpha}(f(\pi,y)-\alpha)\rho(y)\,dy\Big)\Big] \\
&\quad = \lim_{h\to 0}\Big[1 \;+\; \frac{(1-\beta)^{-1}}{h}\int\limits_{f(\pi,y)\in\overline{(\alpha,\alpha+h)}}(f(\pi,y)-\alpha)\rho(y)\,dy \\
&\qquad\qquad -\;(1-\beta)^{-1}\int\limits_{f(\pi,y)\geqslant\alpha+h}\rho(y)\,dy\Big],
\end{aligned}
$$

where $\overline{(\alpha,\alpha+h)}$ is one of the two intervals $[\alpha,\alpha+h]$ or $[\alpha+h,\alpha]$, depending on the sign of h.

We will show that the first term,

$$
\frac{(1-\beta)^{-1}}{h}\int\limits_{f(\pi,y)\in\overline{(\alpha,\alpha+h)}}(f(\pi,y)-\alpha)\rho(y)\,dy,
$$

converges to zero when $h \to 0$. Taking the absolute value, we have:

$$\begin{aligned} 0 &\leqslant \left|\frac{(1-\beta)^{-1}}{h} \int\limits_{f(\pi,y)\in\overline{(\alpha,\alpha+h)}} (f(\pi,y)-\alpha)\rho(y)\,dy\right| \\ &\leqslant \frac{(1-\beta)^{-1}}{|h|} \int\limits_{f(\pi,y)-\alpha\in\overline{(0,h)}} |h|\rho(y)\,dy \\ &= (1-\beta)^{-1} \int\limits_{f(\pi,y)-\alpha\in\overline{(0,h)}} \rho(y)\,dy. \end{aligned}$$

Now, to see that

$$\int\limits_{f(\pi,y)-\alpha\in\overline{(0,h)}} \rho(y)\,dy \xrightarrow[h\to 0]{} 0, \tag{3.1}$$

we define the sets

$$A_h := \{y \in \mathbb{R}^N_{\geqslant 0} | f(\pi,y)-\alpha \in (0,h)\} \quad A_0 := \{y \in \mathbb{R}^N_{\geqslant 0} | f(\pi,y)-\alpha = 0\}$$

and the two functions:

$$\begin{array}{rccl} \zeta_h : & \mathbb{R}^N_{\geqslant 0} & \longrightarrow & \mathbb{R} \\ & y & \longmapsto & \rho(y)\chi_{A_h}(y) \end{array} \qquad \begin{array}{rccl} \zeta_0 : & \mathbb{R}^N_{\geqslant 0} & \longrightarrow & \mathbb{R} \\ & y & \longmapsto & \rho(y)\chi_{A}(y) \end{array}$$

We have that:

I $\zeta_h \longrightarrow \zeta_0$ pointwise on y, if $h \to 0$

II $|\zeta_h(y)| \leqslant \rho(y)\ \forall\ h$ and $\forall\ y \in \mathbb{R}^N$

Therefore, by the Dominated Convergence Theorem (D.C.T.) (refer to Appendix D), we have:

$$\int\limits_{f(\pi,y)-\alpha\in\overline{(0,h)}} \rho(y)\,dy = \int\limits_{y\in\mathbb{R}^N} \zeta_h(y)\,dy \xrightarrow[h\to 0]{D.C.T} \int\limits_{y\in\mathbb{R}^N} \zeta_0(y)\,dy. \tag{3.2}$$

Since f is affine, the set A is a subspace of R^N of dimension $N-1$. Therefore, it is a negligible set for the N dimensional Lebesgue measure. Consequently, the value of the integral of ζ_0 in $\mathbb{R}^N$ will be zero.

Looking at the last term, we have

$$
\begin{aligned}
\lim_{h\to 0}\,(1-\beta)^{-1}\int\limits_{f(\pi,y)\geqslant \alpha+h}\rho(y)\,dy &= (1-\beta)^{-1}\Big(\int\limits_{f(\pi,y)\geqslant\alpha}\rho(y)\,dy \;+\int\limits_{f(\pi,y)-\alpha\in\overline{(0,h)}}\rho(y)\,dy\Big)\\
&= (1-\beta)^{-1}\int\limits_{f(\pi,y)\geqslant\alpha}\rho(y)\,dy.
\end{aligned}
$$

This is because the second term converges to zero because of the Limit 3.1. and hence

$$
\lim_{h\to 0}\frac{F_\beta(\pi,\alpha+h)-F_\beta(\pi,\alpha)}{h} = 1-(1-\beta)^{-1}\int\limits_{f(\pi,y)\geqslant\alpha}\rho(y)\,dy.
$$

Defining a function γ by

$$
\begin{aligned}
\gamma:\ \mathbb{R} &\longrightarrow \mathbb{R}\\
\alpha &\longmapsto \int\limits_{f(\pi,y)-\alpha\geqslant 0}\rho(y)\,dy,
\end{aligned}
$$

and given α and $\breve{\alpha}$ in $\mathbb{R}$ close to each other (for example, $\alpha=\breve{\alpha}+h$ for a certain h), then, by the Limit 3.2, the difference

$$
|\gamma(\alpha)-\gamma(\breve{\alpha})| \xrightarrow[h\to 0]{} 0
$$

converges to zeros, thus showing the continuity with respect to α of our derivative.

By what we have seen then, F_β is continuously differentiable with respect to α with derivative

$$
\frac{\partial F_\beta}{\partial\alpha}(\pi,\alpha) = 1-(1-\beta)^{-1}\int\limits_{f(\pi,y)\geqslant\alpha}\rho(y)\,dy,
$$

and by definition of Ψ, we have

$$
\begin{aligned}
\frac{\partial F_\beta}{\partial\alpha}(\pi,\alpha) &= 1-(1-\beta)^{-1}(1-\Psi(\pi,\alpha))\\
&= (1-\beta)^{-1}[1-\beta-1+\Psi(\pi,\alpha)]\\
&= (1-\beta)^{-1}[\Psi(\pi,\alpha)-\beta].
\end{aligned}
$$

Knowing that F_β is convex, for a given $\pi\in\Pi$, a point where the derivative of $F_\beta(\pi,\cdot)$ vanishes, is the minimum of the function (refer to Appendix E). This minimum is attained by one or more $\alpha \in \mathbb{R}$, that, by definition, are in $A_{\beta,\pi}$, and are precisely the ones that satisfy $\frac{\partial F_\beta}{\partial \alpha}(\pi, \alpha) = 0$ which is equivalent to $\Psi(\pi, \alpha) - \beta = 0$. As discussed above in the remarks, the set of α that satisfy this last condition form an nonempty closed bounded interval. This concludes the second point of the theorem.

With the following calculations, we will show the third point of the theorem. By definition of $\alpha_{\beta,\pi}$, we have $\alpha_{\beta,\pi} \in A_{\beta,\pi}$, so that

$$\min_{\alpha \in \mathbb{R}} F_\beta(\pi, \alpha) = F_\beta(\pi, \alpha_{\beta,\pi}) = \alpha_{\beta,\pi} + (1-\beta)^{-1} \int\limits_{y \in \mathbb{R}^N} [f(\pi, y) - \alpha_{\beta,\pi}]^+ \rho(y)\, dy.$$

Looking at the integral, we have

$$\int\limits_{f(\pi,y) \geqslant \alpha_{\beta,\pi}} (f(\pi, y) - \alpha_{\beta,\pi})\rho(y)\, dy = \int\limits_{f(\pi,y) \geqslant \alpha_{\beta,\pi}} f(\pi, y)\rho(y)\, dy - \alpha_{\beta,\pi} \int\limits_{f(\pi,y) \geqslant \alpha_{\beta,\pi}} \rho(y) dy.$$

In the above, the first integral is, by Definition 7, $(1-\beta)\phi_{\beta,\pi}$ and the second integral is $1 - \Psi(\pi, \alpha_{\beta,\pi})$, but $\Psi(\pi, \alpha_{\beta,\pi}) = \beta$. Thus,

$$\min_{\alpha \in \mathbb{R}} F_\beta(\pi, \alpha) = \alpha_{\beta,\pi} + (1-\beta)^{-1}[(1-\beta)\phi_{\beta,\pi} - \alpha_{\beta,\pi}(1-\beta)] = \phi_{\beta,\pi} \ .$$

This concludes the third point of Theorem 1 and completes the proof. $\square$

The theorem above permits us to see, without ambiguity, $\phi_{\beta,\pi}$ as a function of π, that is:

$$\begin{array}{rccl} \phi_\beta : & \Pi & \longrightarrow & \mathbb{R} \\ & \pi & \longmapsto & \phi_\beta(\pi) := \min\limits_{\alpha \in \mathbb{R}} \{F_\beta(\pi, \alpha)\}, \end{array}$$

and we are interested to find a minimum with respect to $\pi \in \Pi$. For the CVaR seen as a measure of risk, this would represent the search for the portfolio $\pi \in \Pi$ that has minimum risk. Since ϕ_β is continuous (because it is the minimum with respect to another variable of a continuous function) over a compact set, then the minimum $\min_{\pi \in \Pi}\{\phi_\beta(\pi)\}$ is attained and so we have (refer to Appendix F)

$$\min_{\pi \in \Pi}\{\phi_\beta(\pi)\} = \min_{\pi \in \Pi}\{\min_{\alpha \in \mathbb{R}}\{F_\beta(\pi, \alpha)\}\} = \min_{(\pi,\alpha) \in \Pi \times \mathbb{R}}\{F_\beta(\pi, \alpha)\}.$$

We are also interested in having ϕ_β as a convex function with respect to π, that is, we want to have $\forall\ \pi, \breve{\pi} \in \Pi$ and $\lambda \in (0,1)$

$$\phi_\beta(\lambda\pi + (1-\lambda)\breve{\pi}) \leqslant \lambda\phi_\beta(\pi) + (1-\lambda)\phi_\beta(\breve{\pi})$$

To see this, we proceed as follow: by definition of min, we have for all $\epsilon > 0$, $\exists\ \alpha_1, \alpha_2$ such that

$$F_\beta(\pi, \alpha_1) \leqslant \phi_\beta(\breve{\pi}) + \epsilon \text{ and } F_\beta(\breve{\pi}, \alpha_2) \leqslant \phi_\beta(\breve{\pi}) + \epsilon.$$

Multiplying the first inequality by λ, the second by $(1-\lambda)$ and adding them together, yields in:

$$\begin{aligned}
\lambda(\phi_\beta(\pi) + \epsilon) + (1-\lambda)(\phi_\beta(\breve{\pi}) + \epsilon) &\geqslant \lambda F_\beta(\pi, \alpha_1) + (1-\lambda)F_\beta(\breve{\pi}, \alpha_2) \\
&\geqslant F_\beta(\lambda(\pi, \alpha_1) + (1-\lambda)(\breve{\pi}, \alpha_2)) \\
&= F_\beta(\lambda\pi + (1-\lambda)\breve{\pi}, \lambda\alpha_1 + (1-\lambda)\alpha_2) \\
&\geqslant \phi_\beta(\lambda\pi + (1-\lambda)\breve{\pi})
\end{aligned}$$

and this gives us

$$\lambda\phi_\beta(\pi) + (1-\lambda)\phi_\beta(\breve{\pi}) + \epsilon \geqslant \phi_\beta(\lambda\pi + (1-\lambda)\breve{\pi})$$

Letting $\epsilon \to 0$, we obtain the desired result, the convexity of the risk function.

Having introduced another measure of risk, and studied its properties, we can now define a new strategy for portfolio selection.

Problem 2.

$$\min_{(\pi,\alpha)\in\Pi\times\mathbb{R}} \{F_\beta(\pi, \alpha)\}$$

under the constraints

1. $\pi_i \geqslant 0\ \forall\ i \in \{1, \dots, N\}$
2. $\sum_{i=1}^{N} \pi_i = 1$
3. $\mathbb{E}[r^\pi] \geqslant p$ *(for a given p)*

Note: the first two constraints are redundant since all $\pi \in \Pi$ satisfy them. However, we decide to write them out again to have the same presentation as Problem 1, and thus making the analogy easier in one's mind.

3.2 Determining the portfolio with smallest CVaR

Taking variance as a risk measure, we were confronted to solve a quadratic programming problem. This new risk measure, CVaR, seems difficult to minimize because it involves the distribution ρ, which is a difficult quantity to approximate numerically. However, we now show that one can reduce the optimization problem to a linear programming problem.

The expectation term in $F_\beta(\pi, \alpha)$ (that is, the integral), can be approximated by using a historical sample set $\{y_j\}_{j=1}^n$ of stock returns, and this gives the approximated function

$$F(\pi, \alpha) = \alpha + (1-\beta)^{-1} \int\limits_{y \in \mathbb{R}^N_{\geqslant 0}} [f(\pi, y) - \alpha]^+ \rho(y)\, dy \simeq \underbrace{\alpha + \mu \sum_{j=1}^{n} [1 - \langle \pi | y_j \rangle - \alpha]^+}_{=:\tilde{F}_\beta(\pi,\alpha)},$$

where $\mu = (n(1-\beta))^{-1}$. If we introduce the auxiliary real variables $z_j,\ j = 1, \ldots, n$, then solving Problem 2 (with the function $\tilde{F}_\beta$) , is equivalent (refer to Appendix G) to solving the following linear programming problem:

Problem 3.

$$\min_{(\pi,\alpha,z) \in \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^n} \{\alpha + \mu \sum_{j=1}^{n} z_j\}$$

under the constraints

1. $\pi_i \geqslant 0\ \forall\ i \in \{1, \ldots, N\}$
2. $\sum_{i=1}^{N} \pi_i = 1$
3. $\mathbb{E}[r^\pi] \geqslant p$ *(for a given p)*
4. $z_j \geqslant 0\ \forall\ j \in \{1, \ldots, n\}$
5. $z_j \geqslant 1 - \langle \pi | y_j \rangle - \alpha\ \forall\ j \in \{1, \ldots, n\}$

where $\mu = (n(1-\beta))^{-1}$.

Our objective function in Problem 3, is only a scalar product between the variable $(\pi_1, \ldots, \pi_N, \alpha, z_1, \ldots, z_n)$ and the vector $(\underbrace{0, \ldots, 0}_{=N}, 1, \underbrace{\mu, \ldots, \mu}_{=n})$

and all the constraints form convex sets and so the feasible set (i.e. the intersection of all the constraint sets) is convex. Therefore the optimization Problem 2 has been reduced to a linear programming problem.

Note: The type of portfolio management where we solve Problem 3 for each monthly period, will be referred as `Min CVaR` portfolio management type.

Again we can ask about the existence and uniqueness of the solution π_* to Problem 2. From the discussion above we have

$$\min_{(\pi,\alpha)\in\Pi\times\mathbb{R}}\{F_\beta(\pi,\alpha)\} = \min_{\pi\in\Pi}\{\min_{\alpha\in\mathbb{R}}\{F_\beta(\pi,\alpha)\}\} = \min_{\pi\in\Pi}\{\phi_\beta(\pi)\}.$$

Because $\phi_\beta(\pi)$ is continuous, our theoretical Problem 2 will have a solution on all compact set of $\mathbb{R}^N$. Again, if there is an appropriate choice for p, then the constraints for Problem 2 are a nonempty compact set in $\mathbb{R}^N$, and thus a minimum exists. For uniqueness, it can be shown (refer to Appendix H), that for $N = 2$, the Hessian matrix of $F_\beta(\pi,\alpha)$ is a positive definite matrix, and since the function is convex, it therefore has one, and only one (π^*,α^*) that realizes the minimum. We still need to further investigate for the case where N is arbitrary.

As for Problem 1, we show a two dimensional illustration of the situation. In red, we have the contour lines for the objective function $\tilde{F}_\beta$, were we have fixed $\alpha = 0.1296$. This α corresponds to the α that, with the optimal portfolio (the thick red point), minimizes $\tilde{F}_\beta$. The same data and the same value of p were taken as for Figure 2.2.

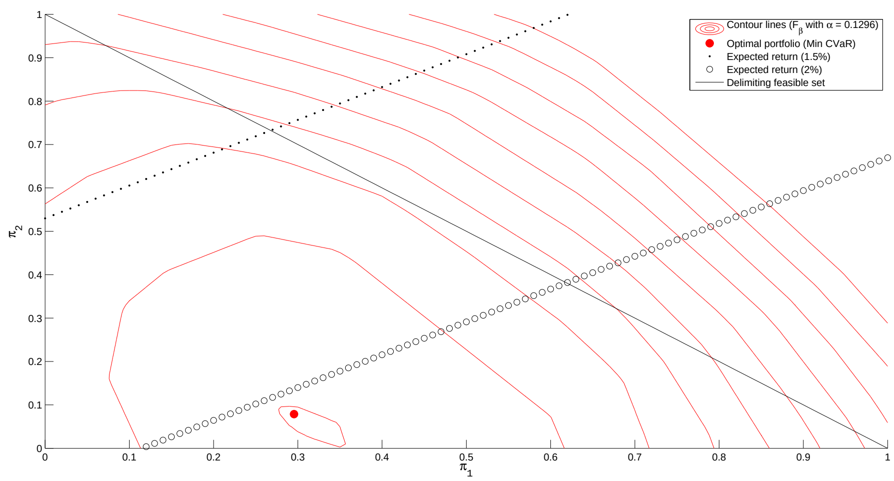


Figure 3.2: Contour lines of the the objective function and expected return with the optimal portfolio.

4 The new approach using entropy

'*I believe that better methods* [...to find μ_i and $\sigma_{i,j}$ ...]*, which take into account more information, can be found.*'[14], says MARKOWITZ. Even if we have this '*information*', are we certain that it is reliable? And, what about if we do not have this '*information*'? Supposing that we are not able to determine the probability model, to calculate the parameter set Φ or to verify the necessary assumptions and to quantify other variables. What can we do?

4.1 The SHANNON entropy

> 1) Phys Grandeur qui, en thermodynamique, permet d'valuer la dgradation de l'nergie d'un systme.
>
> 'L'entropie d'un systme caractrise son degr de dsordre.'
>
> 2) Cybern Dans la thorie de la communication, nombre qui mesure l'incertitude de la nature d'un message donn partir de celui qui le prcde (l'entropie est nulle quand il n'existe pas d'incertitude). [15]

If one states that 'randomness' is omnipresent, we can agree up to some point where, we will argue that there are some phenomena that are 'more random' than others, or, in other words, there are some systems that work very deterministically or, are said to be very `certain`, while others seem to be very hazardous, or, are considered to be very `uncertain`. Very quickly, we are faced with the following questions

I How 'much' randomness is contained in a system?

II How uncertain is the phenomenon?

The notion of `entropy` tries to answer this question.

4.2 Empirical approach

We will place our discussion in a financial environment and introduce the notion in two steps.

I If we are very certain that the k^{th} stock in our universe has good chances for positive returns and its risk is low and that we are

[14] Refer to [7]
[15] Petit Larousse

confident about our choice, then we should invest a relatively large proportion of x (our initial amount) in it. This gives

$$\pi_k \simeq 1 \Longleftrightarrow \frac{1}{\pi_k} \simeq 1 \quad \text{‘small’}.$$

If, on the other hand, we are very uncertain about the returns and risk of the j^{th} stock, then we should invest a relatively small proportion of x in it. This gives

$$\pi_j \ll 1 \Longleftrightarrow \frac{1}{\pi_j} \gg 1 \quad \text{‘large’}.$$

Therefore, as a measure of uncertainty about our decision, we can take $\frac{1}{\pi_i}$.

II It is a well established fact that our perception (or our understanding) of a signal (this could be a variable or some information that we managed to quantify) is proportional to the logarithm of this signal, i.e.

$$\log(\text{‘signal’}) \sim \text{perception (or understanding)}$$

With these two ideas at hand, we can summarize our situation by

$$\text{Uncertainty in the decision concerning the } i^{th} \text{ stock} \sim \log(\frac{1}{\pi_i}) \ .$$

Since we are concerned with N stocks, then we are interested in the aggregated quantity

$$\frac{1}{N}\sum_{i=1}^{N} -\log(\pi_i).$$

However, it seems more realistic to take the weighted average, i.e.

$$-\sum_{i=1}^{N} \pi_i \log(\pi_i).$$

This can be defined as the entropy of a portfolio. We give a formal definition after we have presented the axiomatic approach, in the following subsection.

4.3 Axiomatic approach

> My greatest concern was what to call it. I thought of calling it 'information'. But the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John Von Neumann, he had a better idea. He told me: ≪ You should call it entropy, for two reasons. In first place, your uncertainty has been used in statistical mechanics under that name, so it already has a name. In second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.≫ [16]

We will now present the axiomatic version for defining the entropy of a portfolio. We will not give the axioms in one block but rather present them along our discussion. Let us consider the two extreme situations. If on one side, we reject the assumption that the future is unpredictable and that everything can be calculated deterministically, or that, after gathering all the necessary information, we conclude that the i^{th} stock will have the largest return and we are certain about our decision, then the portfolio π should be of the form $\pi = (0, \ldots, \underbrace{1}_{i}, \ldots, 0)$. If, on the other hand, we are uncertain about everything or we have no information at all, then the most rational way to construct a portfolio would be $\pi = (\frac{1}{N}, \ldots, \frac{1}{N})$. Between these two extremes situations, is it possible to construct a function H that takes a portfolio as its argument and associates to this portfolio, a quantity that reflects the certainty (or uncertainty) of our decision, which is linked to the portfolio, or more precisely, which is linked to how the portfolio is composed? We might answer this question if we introduce several assumptions on the function H. Up to now, we can summarize the situation by

It is known which investment will have the largest return.	$\longrightarrow$	We know what to do.	$\longrightarrow$	Portfolio $\pi = (0, \ldots, 1, \ldots, 0)$ $H(\pi) = 0$
It is unknown which investment will have the largest return.	$\longrightarrow$	We do not know what to do.	$\longrightarrow$	Portfolio $\pi = (\frac{1}{N}, \ldots, \frac{1}{N})$ $H(\pi) = \max$

[16] C. SHANNON, quoted in [13], page 20

Note: by convention, we have chosen $H(\pi) = 0$ when 'We know what to do' and $H(\pi) = \max$ when 'We do not know what to do'.

As a preliminary axiom, we impose

$$H(\pi_1, \dots, \pi_N) = H(\pi_{\sigma(1)}, \dots, \pi_{\sigma(N)})$$

for all $\sigma \in Sym(N)$, where $Sym(N)$ is the group of all the permutations of $\{1, \dots, N\}$. This means that our certainty of our decision in investing, for example (if $N = 3$), 70% in the first stock, 20% and 10% in the second and third stock respectively, is the same as if we would have invested 20% in the first stock, 10% and 70% in the seconde and third stock respectively.

Having introduce this idea, we are able to define the domain set of the function H

$$H : \ \Pi/Sym(N) \ \longrightarrow \ \mathbb{R}_{\geqslant 0},$$

where $\Pi/Sym(N)$ is the set of all the orbits. This means that for all $\pi \in \Pi$, we define the action of the group by

$$\sigma \centerdot \pi = \sigma \centerdot (\pi_1, \dots, \pi_N) := (\pi_{\sigma(1)}, \dots, \pi_{\sigma(N)})$$

where $\sigma \in Sym(N)$. The orbit $\bar{\pi}$ is the set $\{\sigma \centerdot \pi \mid \sigma \in Sym(N)\}$. Then

$$\Pi/Sym(N) := \{\bar{x} \mid x \in \Pi\}.$$

Intuitively, this means that an element $(\pi_1, 0, 0, \pi_4, 0, 0, \pi_7)$ in $\Pi/Sym(N)$ is the same as $(\pi_7, 0, \pi_4, 0, 0, \pi_1, 0)$, which again is the same as $(\pi_1, \pi_4, \pi_7, 0, 0, 0, 0)$.

To not overload our notation, we make the following simplification. Let $\pi \in \Pi$. Let $I_\pi := \{i_1, \dots, i_k\} \subset \{1, \dots, N\}$ be the set of indices such that $\pi_i = 0$ for all $i \in \{1, \dots, N\} \setminus I_\pi$. From what we have seen above, we have $H(\pi) = H(\pi_{i_1}, \dots, \pi_{i_k}, 0, \dots, 0)$. Our simplification is

$$H(\underbrace{\pi_{i_1}, \dots, \pi_{i_k}}_{k}, \underbrace{0, \dots, 0}_{N-k}) = H(\underbrace{\pi_{i_1}, \dots, \pi_{i_k}}_{k})$$

We now introduce axioms for the function H. Let us look at the case where there is very little information, but enough to decide if there are some stocks that will not be chosen. Let N be the number of stocks in the universe. We define $f(j) := H(\underbrace{\frac{1}{j}, \dots, \frac{1}{j}}_{j})$ for $j \in \mathbb{Z}_{\geqslant 1}$. We impose the following conditions on f called *Axiom I*:

I $f(N)$ must be the maximum value of H.

II $f(\cdot)$, as a function of the number of chosen stocks in the universe, is a strictly increasing function. The idea behind this is that if one takes $n < N$ securities, this means that the investor has 'some' information since he decides not to invest in the $N-n$ other stocks.

III $f(1) = 0$, meaning that there is either total determinism or that the investor is completely sure about investing the hole of the value x into one stock.

Suppose that there exists $M, L \in \mathbb{Z}_{\geqslant 0}$ such that $ML = N$. Let us consider M baskets of stocks in the universe, all of them having L stocks. Suppose that we are still in the case where there is very little information available, or that the market seems totally unpredictable. We impose on

$$f(ML) = H(\frac{1}{ML}, \dots, \frac{1}{ML})$$

the following property: the certitude of deciding to invest in the j^{th} basket does not influence the uncertainty in deciding how to invest among the L stocks. In other words, if we decide to invest only in one basket, then the remaining uncertainty in deciding how to compose the portfolio after having chosen the basket, is $f(L)$, that is

$$f(ML) - f(M) = f(L) \Leftrightarrow f(ML) = f(M) + f(L) \qquad \textit{Axiom II}$$

Let us now look at the case where we have some information that we can rely on to make our decision, that is, among the $j \leqslant N$ stocks, we decide to invest with the portfolio $\pi = (\pi_1, \dots, \pi_j)$. Therefore, the quantity $H(\pi)$ reflects our certainty (or uncertainty) of our decision.

Let A et B be two nonintersecting subsets of stocks of the portfolio with $|A| = r$ and $|B| = j - r$ as their respective cardinals. The weights of A and B are

$$\pi_A = \pi_1 + \dots + \pi_r \quad \text{and} \quad \pi_B = \pi_{r+1} + \dots + \pi_j$$

respectively. Our aim is to link $H(\pi)$ and $H(\pi_A, \pi_B)$, where $H(\pi_A, \pi_B)$ is to been seen as the uncertainty in deciding between the subset A or the subset B.

If we know for sure that we have to invest in the subset A, then the uncertainty in deciding among the portfolio based on the subset A is

$$H(\frac{\pi_1}{\pi_A}, \dots, \frac{\pi_r}{\pi_A})$$

On the other hand, if we know for sure that we have to invest in the subset B, then the uncertainty in deciding among the portfolio based on the subset B is

$$H(\frac{\pi_{r+1}}{\pi_B}, \dots, \frac{\pi_j}{\pi_B})$$

We claim (postulate) that the remaining uncertainty in deciding the portfolio composition, knowing in which subset (either A or B) one must invest in, must be the average of the two 'conditional' uncertainties in deciding. This leads to

$$H(\pi) = H(\pi_A, \pi_B) + \pi_A H(\frac{\pi_1}{\pi_A}, \dots, \frac{\pi_r}{\pi_A}) + \pi_B H(\frac{\pi_{r+1}}{\pi_B}, \dots, \frac{\pi_j}{\pi_B}) \quad \textit{Axiom III}$$

For technical reasons, we need to impose a continuity condition,

$$q \longrightarrow H(q, 1-q) \qquad \textit{Axiom IV}$$

is continuous over $(0,1)$ (and $q \in (0,1)$).

The task to find such a function might seem difficult, but with the help of the following theorem, we know exactly how H is defined.

Theorem 2. *[13] H is a function that satisfies Axioms I,...,IV if, and only if*

$$H(\pi_1, \dots, \pi_N) = -C \sum_{i=1}^{N} \pi_i \log(\pi_i),$$

where $C \in \mathbb{R}_{>0}$.

Remark: $\pi_i \log(\pi_i) = 0$ if $\pi_i = 0$ for a certain $i \in \{1, \dots, N\}$. This is because of the fact that: $\lim_{x \to 0} x \log(x) = 0$.

Proof. [$\Rightarrow$] The first step of the proof is to show that $f(M) = C\log(M)$ where $M \in \mathbb{Z}_{\geqslant 2}$ and $C \in \mathbb{R}_{>0}$. By *Axiom II* we have, for $s \in \mathbb{Z}_{\geqslant 1}$,

$$f(s^2) = f(ss) = f(s) + f(s) = 2f(s)$$

and by induction $f(s^k) = kf(s)$ for all $k \in \mathbb{Z}_{\geqslant 1}$. We fix $M \in \mathbb{Z}_{\geqslant 2}$ and let $r \in \mathbb{Z}_{>0}$. We know that there is a unique k such that $2^r \in [M^k, M^{k+1})$. By *Axiom I* we have

$$\text{since } M^k \leqslant 2^r < M^{k+1} \text{ then } f(M^k) \leqslant f(2^r) < f(M^{k+1})$$

which yields, by the induction argument above,

$$kf(M) \leqslant rf(2) < (k+1)f(M).$$

By the properties of the log function, we have

$$\log(M^k) \leqslant \log(2^r) < \log(M^{k+1}) \quad \text{and} \quad k\log(M) \leqslant r\log(2) < (k+1)\log(M)$$

Hence

$$k \leqslant r\frac{f(2)}{f(M)} < k+1 \quad \text{and} \quad k \leqslant r\frac{\log(2)}{\log(M)} < k+1,$$

and therefore

$$|\frac{f(2)}{f(M)} - \frac{\log(2)}{\log M}| < \frac{1}{r}.$$

Since r was chosen arbitrarily, we have $f(M) = (f(2)/\log(2))\log(M)$, which concludes the first part of the proof.

The second part of the proof consists of showing that

$$H(q, 1-q) = -C[q\log(q) + (1-q)\log(1-q)]. \tag{4.1}$$

From now on, we will let $C := \frac{f(2)}{\log(2)}$ and keep in mind that the value of C depends on the 'initial value' of $f(2)$. We show (4.1) for $q = \frac{r}{s} \in \mathbb{Q}$ and by the continuity of the function H (*Axiom IV*), the results holds for all $q \in (1,0)$. Lets consider a portfolio with s stocks which is uniformly weighted (that is, for the security i, its weight is $\pi_i = \frac{1}{s}$ for all $i \in \{1, \ldots, s\}$). We divide the set containing all the s stocks into two subsets A and B, with $|A| = r$ and $|B| = s - r$ as their respective cardinals. We therefore have, for the weights of A and B, $\pi_A = r\frac{1}{s}$ and $\pi_B = (s-r)\frac{1}{s}$ respectively. We also have

$$H(\frac{\pi_1}{\pi_A}, \ldots, \frac{\pi_r}{\pi_A}) = H(\frac{1/s}{r/s}, \ldots, \frac{1/s}{r/s}) = H(\frac{1}{r}, \ldots, \frac{1}{r}) = f(r),$$

and

$$
\begin{aligned}
H(\frac{\pi_{r+1}}{\pi_B}, \dots, \frac{\pi_s}{\pi_B}) &= H(\frac{1/s}{(s-r)/s}, \dots, \frac{1/s}{(s-r)/s}) \\
&= H(\frac{1}{(s-r)}, \dots, \frac{1}{(s-r)}) = f(s-r).
\end{aligned}
$$

From *Axiom III*, we obtain

$$
f(s) = H(\frac{r}{s}, \frac{s-r}{s}) + \frac{r}{s} f(r) + \frac{s-r}{s} f(s-r).
$$

By the first step of the proof, we know the value of $f(\cdot)$ and letting $q := \frac{r}{s}$, we get

$$
C \log(s) \;=\; H(q, 1-q) + qC\log(r) + (1-q)C\log(s-r),
$$

hence

$$
\begin{aligned}
H(q, 1-q) &= -C[q\log(r) - \log(s) + (1-q)\log(s - sq)] \quad \text{since } r = sq \\
&= -C[q\log(r) - \log(s) + (1-q)\log(s) + (1-q)\log(1-q)] \\
&= -C[q\log(r) - q\log(s) + (1-q)\log(1-q)] \\
&= -C[q\log(\frac{r}{s}) + (1-q)\log(1-q)] \\
&= -C[q\log(q) + (1-q)\log(1-q)].
\end{aligned}
$$

We have proved the second part of the proof.

In the third part of the proof, we proceed by induction. Let $n \in \mathbb{Z}_{\geqslant 1}$ be the number of stocks under consideration. The second part of the proof (4.1) shows the result of the theorem when $n = 2$. We suppose that the result holds for $n \leqslant N-1$, where $N \in \mathbb{Z}_{\geqslant 2}$ is the number of stocks in the universe. Let $\pi = (\pi_1, \dots, \pi_N)$ be a chosen portfolio. We again divide the set containing all the stocks into two subsets A and B, with $|A| = 1$ and $|B| = N-1$ as their respective cardinals. Without loss of generality, we place the first stock $i = 1$ in the subset A. We therefore have, for the weights of A and B, $\pi_A = \pi_1$ and $\pi_B = 1 - \pi_1$ respectively.

By the third axiom, we have

$$\begin{aligned}
H(\pi_1,\ldots,\pi_N) &= H(\pi_1, 1-\pi_1) + \pi_1 H(\tfrac{\pi_1}{\pi_1})\\
&+ (1-\pi_1)\underbrace{H(\frac{\pi_2}{1-\pi_1},\ldots,\frac{\pi_N}{1-\pi_1})}_{\text{to calculate } H\text{,we use the induction hypothesis}}\\
&= -C[\pi_1\log(\pi_1) + (1-\pi_1)\log(1-\pi_1)] \; + \pi_1 \underbrace{H(1)}_{=0}\\
&\quad (1-\pi_1)(-C)\sum_{i=2}^{N}\tfrac{\pi_i}{(1-\pi_1)}\log(\tfrac{\pi_i}{(1-\pi_1)})\\
&= -C\Big[\pi_1\log(\pi_1) + \log(1-\pi_1) - \pi_1\log(1-\pi_1) +\\
&\quad \sum_{i=2}^{N}\pi_i(\log(\pi_i) - \log(1-\pi_1))\Big]\\
&= -C\Big[\pi_1\log(\pi_1) + \sum_{i=2}^{N}\pi_i\log(\pi_i) + \log(1-\pi_1)\\
&\quad -\log(1-\pi_1)\underbrace{\sum_{i=1}^{N}\pi_1}_{=1}\Big]
\end{aligned}$$

Therefore $H(\pi_1,\ldots,\pi_N) = -C\sum_{i=1}^{N}\pi_i\log(\pi_i)$.

[$\Leftarrow$] We will verify the axioms.

Verifying Axiom I Since log is an increasing function, we obtain, for $n < N, n, N \in \mathbb{Z}_{\geqslant 0}$ and for $C > 0$,

$$\begin{aligned}
\log(n) &< \log(N)\\
-C\log(\tfrac{1}{n}) &< -C\log(\tfrac{1}{N})\\
-C\sum_{i=1}^{n}\tfrac{1}{n}\log(\tfrac{1}{n}) &< -C\sum_{i=1}^{N}\tfrac{1}{N}\log(\tfrac{1}{N})\\
f(\tfrac{1}{n},\ldots,\tfrac{1}{n}) &< f(\tfrac{1}{N},\ldots,\tfrac{1}{N}).
\end{aligned}$$

Therefore f as a function of the number of stocks in the universe, is a strictly increasing function.

Verifying Axiom II By definition of H, we have

$$\begin{aligned} H(\tfrac{1}{ML},\dots,\tfrac{1}{ML}) &= -C\sum_{i=1}^{ML}\tfrac{1}{ML}\log(\tfrac{1}{ML}) \\ &= -C\log(\tfrac{1}{ML}) \\ &= -C(\log(\tfrac{1}{M})+\log(\tfrac{1}{L})) \\ &= -C\sum_{i=1}^{M}\tfrac{1}{M}\log(\tfrac{1}{M}) - C\sum_{i=1}^{L}\tfrac{1}{L}\log(\tfrac{1}{L}) \\ &= H(\tfrac{1}{M},\dots,\tfrac{1}{M}) + H(\tfrac{1}{L},\dots,\tfrac{1}{L}), \end{aligned}$$

and by the definition of f we have the second axiom.

Verifying Axiom III By hypothesis, we have

$$H(\pi_1,\dots,\pi_N) = -C\sum_{i=1}^{N}\pi_i\log(\pi_i) \text{ with } C\in\mathbb{R}_{>0}$$

Let

$$\pi_A := \pi_1+\cdots+\pi_r \text{ and } \pi_B := \pi_{r+1}+\cdots+\pi_n$$

then

$$\begin{aligned} &H(\pi_A,\pi_B)+\pi_A H(\tfrac{\pi_1}{\pi_A},\dots,\tfrac{\pi_r}{\pi_A})+\pi_B H(\tfrac{\pi_{r+1}}{\pi_B},\dots,\tfrac{\pi_N}{\pi_B}) \\ &= -C(\pi_A\log(\pi_A)+\pi_B\log(\pi_B)) + \pi_A(-C\sum_{i=1}^{r}\tfrac{\pi_i}{\pi_A}\log(\tfrac{\pi_i}{\pi_A})) \\ &\quad +\pi_B(-C\sum_{i=r+1}^{N}\tfrac{\pi_i}{\pi_B}\log(\tfrac{\pi_i}{\pi_B})) \\ &= -C[\pi_A\log(\pi_A)+\pi_B\log(\pi_B)+\sum_{i=1}^{r}\pi_i\log(\tfrac{\pi_i}{\pi_A}) \\ &\quad +\sum_{i=r+1}^{N}\pi_i\log(\tfrac{\pi_i}{\pi_B})] \\ &= -C[\pi_A\log(\pi_A)-\log(\pi_A)\underbrace{\sum_{i=1}^{r}\pi_i}_{=\pi_A}+\pi_B\log(\pi_B)-\log(\pi_B)\underbrace{\sum_{i=r+1}^{N}\pi_i}_{=\pi_B} \\ &\quad +\sum_{i=1}^{r}\pi_i\log(\pi_i)+\sum_{i=r+1}^{N}\pi_i\log(\pi_i)] \\ &= -C\sum_{i=1}^{N}\pi_i\log(\pi_i) = H(\pi_1,\dots,\pi_N). \end{aligned}$$

Verifying Axiom IV If $N = 2$, then, because $x \longrightarrow x\log(x)$ is continuous for $x \in \mathbb{R}_{\geqslant 0}$,

$$H(\pi, 1-\pi) = -C(\pi\log(\pi) + (1-\pi)\log(1-\pi))$$

is continuous, and hence the last axiom is verified, and we have completed the proof.

$\square$

From this theorem, we see that all the functions that have the properties that we are interested in, differ only by a constant C. It is therefore reasonable to choose $C = 1$ for the function one wants to work with. This gives us the following definition.

Definition 8. *The* `entropy` *of a portfolio* π *is*

$$H(\pi) = H(\pi_1, \ldots, \pi_N) = -\sum_{i=1}^{N} \pi_i \log(\pi_i).$$

Having introduced the notion of entropy, let us suppose that we know very little about the returns of the stocks under consideration, but assume that their first moment exists. It is then possible to construct the vector μ, either from historical samples, or from qualitative based financial information coming from banks, brokers or financial institutions that we juge reliable. Then, we suppose that we know nothing else and that we will not even introduce a risk measure in our rule to construct our portfolio. This lack of information concerns our uncertainty in deciding, therefore, at this stage, we should maximize the entropy of our portfolio and thus, by doing so, it reflects the fact that we are missing information that would otherwise allow us to construct, what we would think to be more 'certain' portfolios. It is important to note that maximizing the portfolio's entropy does not violate MARKOWITZ's guideline idea of 'diversification', since portfolios with a high entropy are diversified: the one which has maximum entropy is the one that is mostly distributed (diversified) throughout the N stocks of the universe. This new approach having been introduced, we will expose the underlying mathematical problem. Instead of maximizing H, we minimize $-H$, which is of course equivalent.

Problem 4.

$$\min_{\pi \in \Pi} \{-H(\pi)\}$$

under the constraints

1. $\pi_i \geqslant 0 \; \forall \; i \in \{1, \ldots, N\}$
2. $\sum_{i=1}^{N} \pi_i = 1$
3. $\mathbb{E}[r^\pi] \geqslant p$ *(for a given p)*

Note: The type of portfolio management where we solve Problem 4 for each monthly period, is, from now on, called `Max Entropy`.

4.4 Solving the problem

We can use a numerical solver for this problem. In our simulations, we have used the `fmincon` function in Matlab, that minimizes a nonlinear function with linear, and/or none linear constraints.

Asking ourselves about existence and uniqueness of a solution to Problem 4, we can again easily answer the first question by noting that $-H$ is continuous on the compact set Π and a solution exists if p is well chosen.

For uniqueness, refer to Appendix I.

We again show a two dimensional illustration of the situation. In green, we have the contour lines for the entropy function. The thick green point is the optimal portfolio according to this type of portfolio management. The same data and the same value of p were taken as for Figure 2.2 and Figure 3.2.

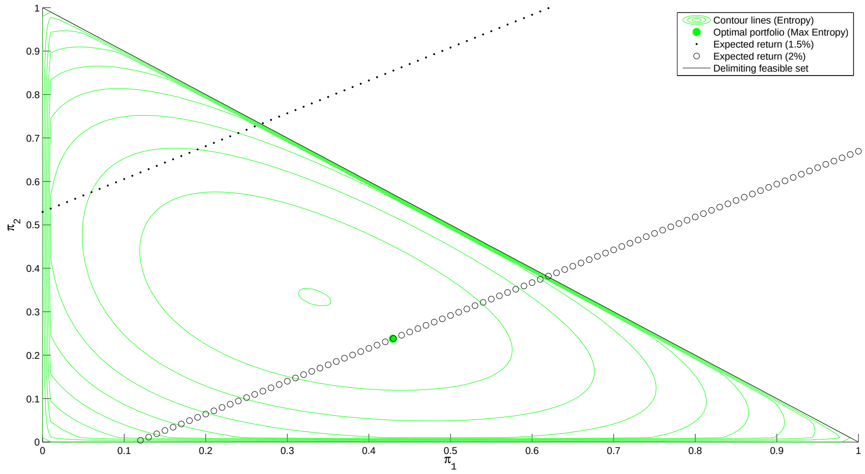


Figure 4.1: Contour lines of the entropy function and expected return with the optimal portfolio.

In the next figure, we superimpose Figure 2.2, Figure 3.2 and Figure 4.1.

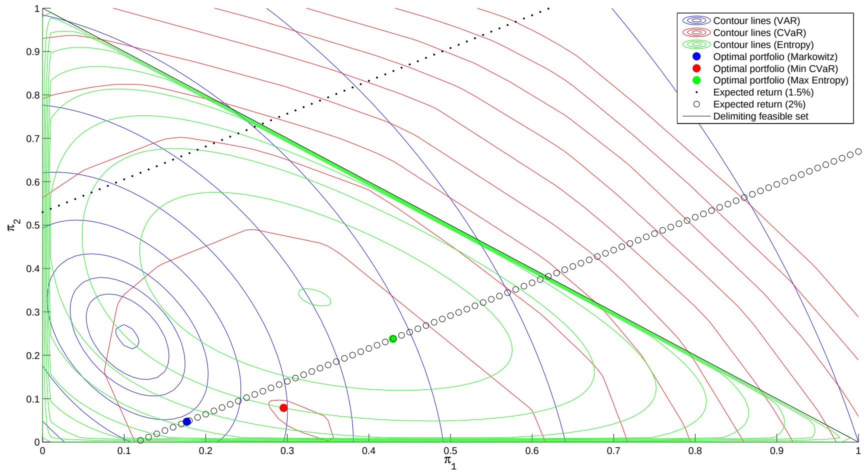


Figure 4.2: Contour lines of the portfolio variance, the objective function, and the entropy function and expected return with the three optimal portfolios.

In the table below, we show the composition (in %) of the different optimal portfolios determined by the three different types of portfolio management.

	Český Telecom	Mol	Millennium Bank
Markowitz	17.7%	4.7%	77.6%
Min CVaR	29.5%	7.9%	62.6%
Entropy	42.9%	23.8%	33.3%

5 Numerical comparison

The back testing was done in the following manner

Data The data was taken in different stock markets: one collection, denoted by `INDIA` was taken from the Bombay Stock Exchange (25 stocks from the BSE 30 Index), another, denoted by `OST`, came from several central-eastern European stock markets (22 stocks), another universe, denoted `DOW` was the Dow Jones Industrial Average Index (from from the New York Stock Exchange), another came from the main stock exchange of the People's Republic of China, Shanghai (21 stocks from the SSE 50 Index) and denoted by `PRC`, and the last univers, from the Swiss market (42 stocks from the SMI Expanded), denoted by `CH`. All the prices were expressed in €. All the data was taken from one source : DataStream.

Time Interval We chose 48 monthly periods, that spread out through 4 consecutive years (2002, 2003, 2004, 2005).

Parameters & Samples The calculation of the parameters in Problem 1 and the sample need to estimate the integral in Problem 2 were treated in the following way

I The expected returns and the variance-covariance matrix were calculated for each different period with the last 36 monthly returns preceding the period where we determine the portfolio.

II The sample consisted of the last 36 monthly returns preceding the period where the portfolio was determined.

The index The `index` was calculated by taking the market capitalisation MC (in €) of each stock in the universe, and determining the weights of the portfolio index π^I as a capitalized-weighted basis, that is

$$\pi_i^I = \frac{MC_i}{\sum\limits_{i=1}^{N} MC_i}.$$

The source for the $MC_i, i = 1, \ldots, N$, was DataStream.

The Tests For each monthly period, Problems 1, 3 and 4 were solved. For the universe INDIA and OST, the parameter p in the constraints (the demanded expected return) was 2% (it approximately correspondes to a 26.8% annual return), for the universe DOW and PRC, p was 0.15% (which represents about 1.82% annual return) and for the universe CH, a 0.5% monthly return was asked (corresponding more or less to 6.17% annual return).

Graphics The graphics, that can been seen below, have been produce for the INDIA, OST, DOW, PRC and CH universe. The first graphic shows the evolution of an initial wealth of € 1.- during the 48 monthly periods for the three different types of portfolio management. The `index` is also shown, as well as the `uniform portfolio` (i.e. all the weights are always $\frac{1}{N}$). The second graphic shows the annual performance for each type of portfolio management as well as for the uniform portfolio and the index, i.e. the percentage increase (or decrease) between 12 consecutive monthly periods. A horizontal yellow line is shown, indicating the annual risk free rate that we took as 1.5%. The third graph shows the cumulative amount that was traded (selling and buying stocks) in ordre to re-balance the portfolio at the end of each period. This statistic, that we will call `flux`, is shown for the three types of portfolio management and for the uniform portfolio.

We now present our graphics for our numerical simulation.

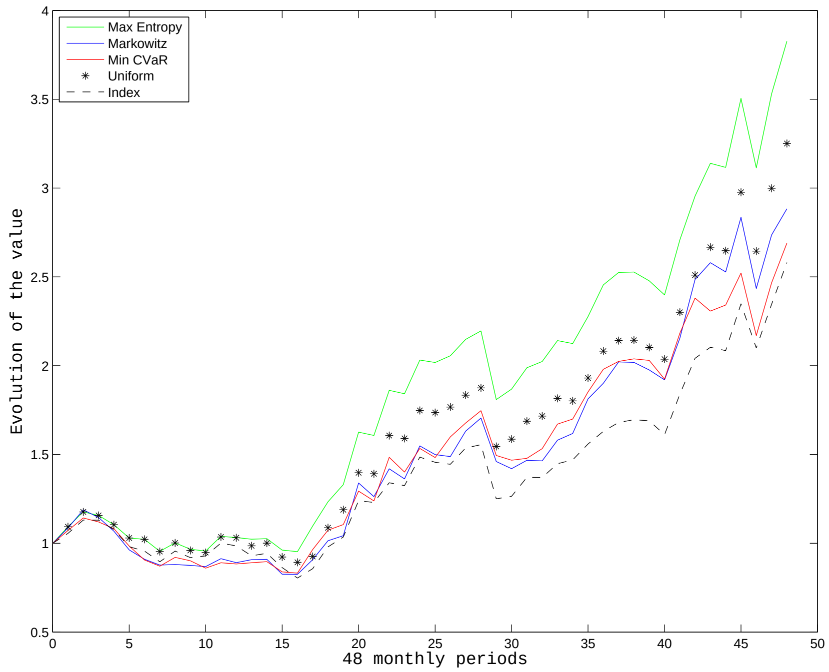


Figure 5.1: Evolution of an initial value of € 1.- with different type of portfolio management asking for a 2% or more monthly return over 48 monthly periods with stocks in INDIA.

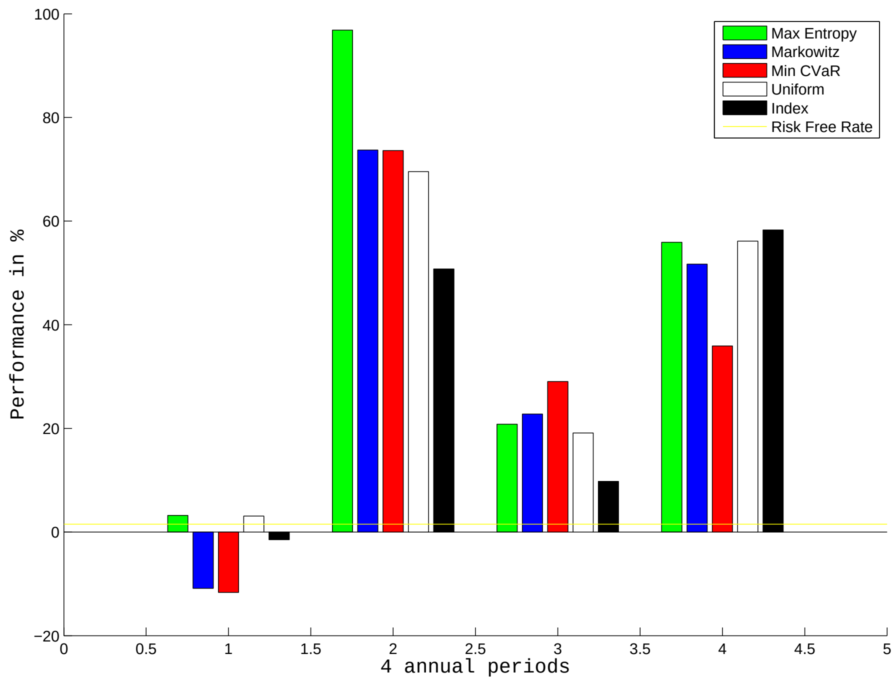


Figure 5.2: The four annual performance in % with stocks in INDIA.

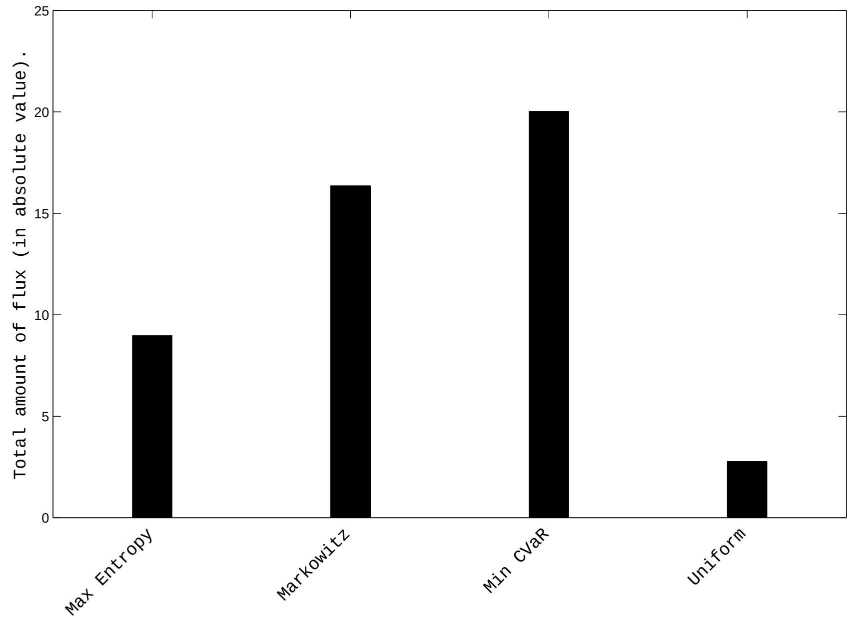


Figure 5.3: The total amount of flux for different type of portfolio management with stocks in INDIA.

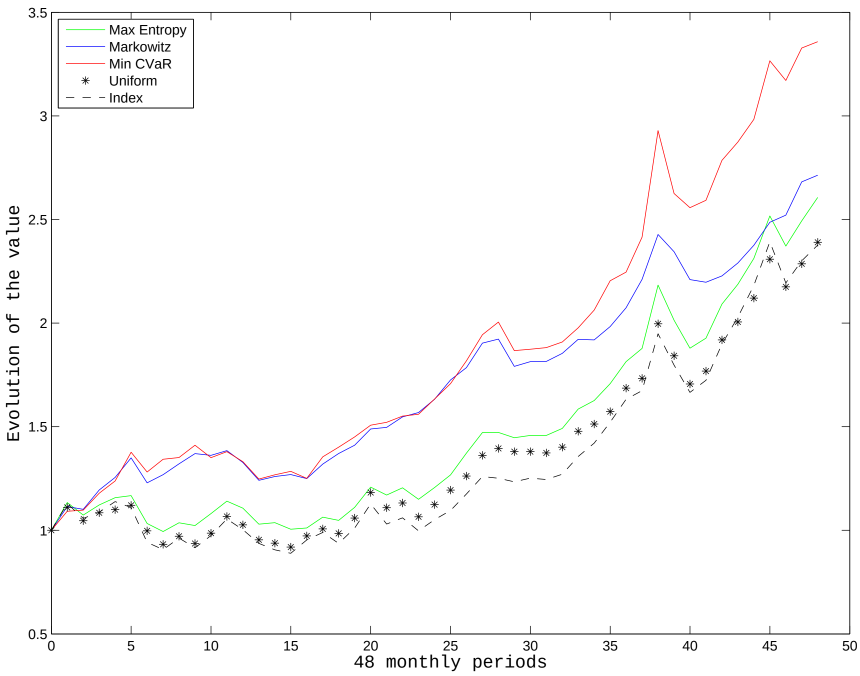


Figure 5.4: Evolution of an initial value of € 1.- with different type of portfolio management asking for a 2% or more monthly return over 48 monthly periods with stocks in OST.

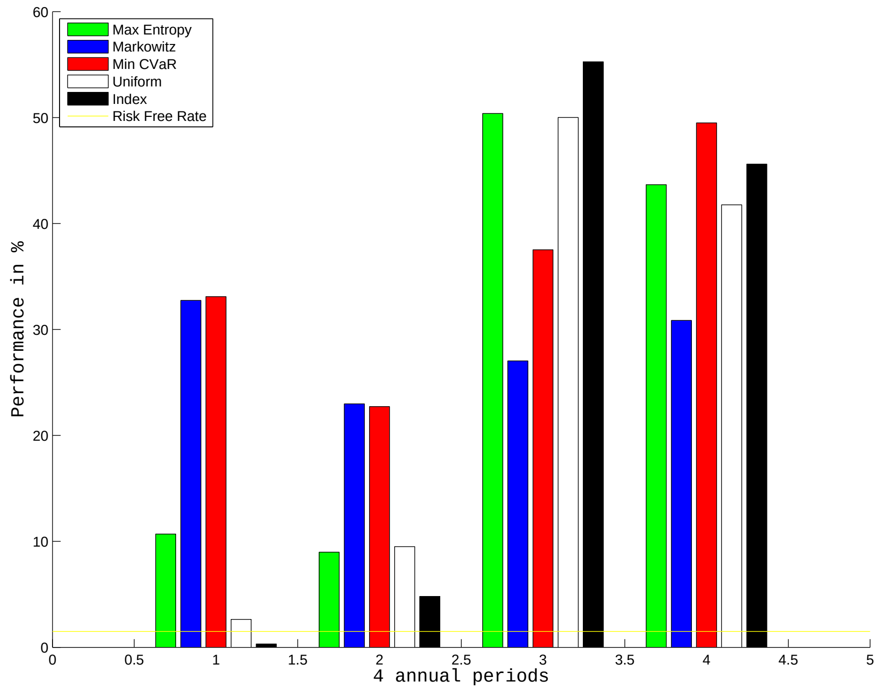


Figure 5.5: The four annual performance in % with stocks in OST.

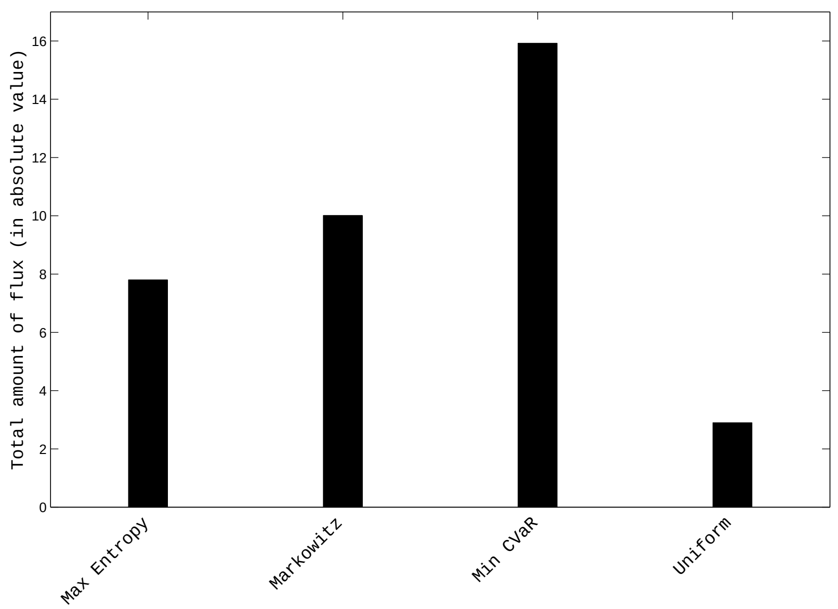


Figure 5.6: The total amount of flux for different type of portfolio management with stocks in OST.

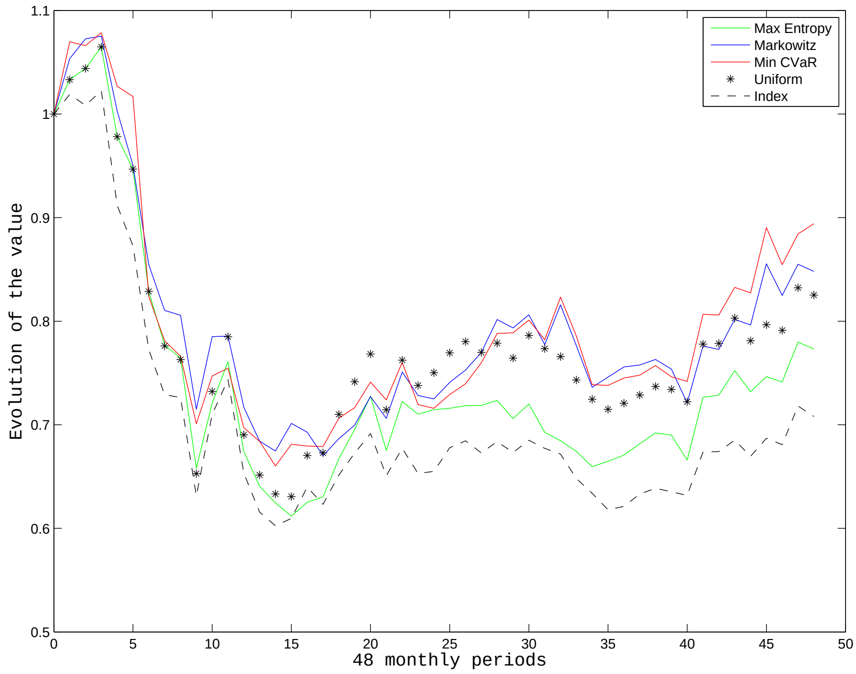


Figure 5.7: Evolution of an initial value of € 1.- with different type of portfolio management asking for a 0.15% or more monthly return over 48 monthly periods with stocks in DOW.

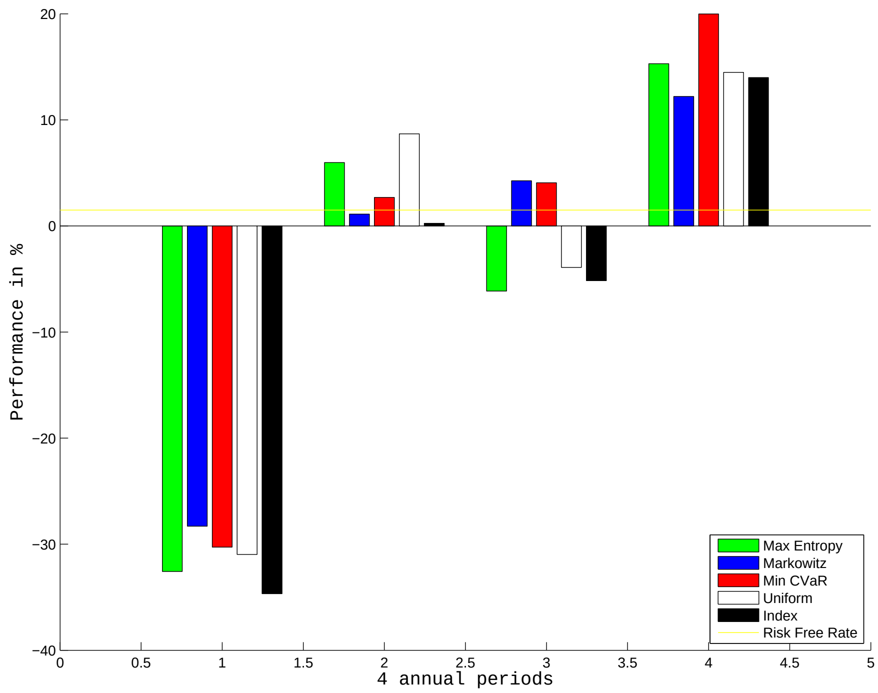


Figure 5.8: The four annual performance in % with stocks in DOW.

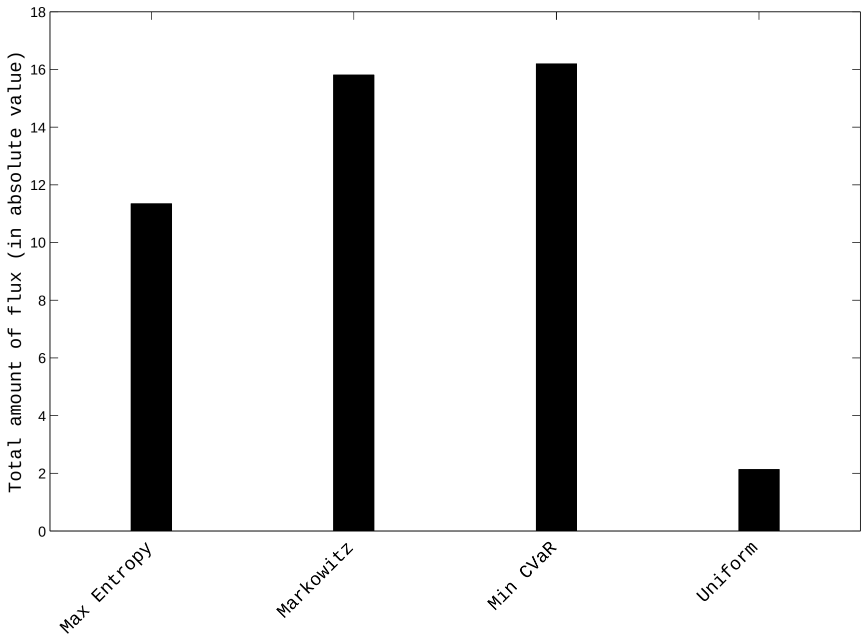


Figure 5.9: The total amount of flux for different type of portfolio management with stocks in DOW.

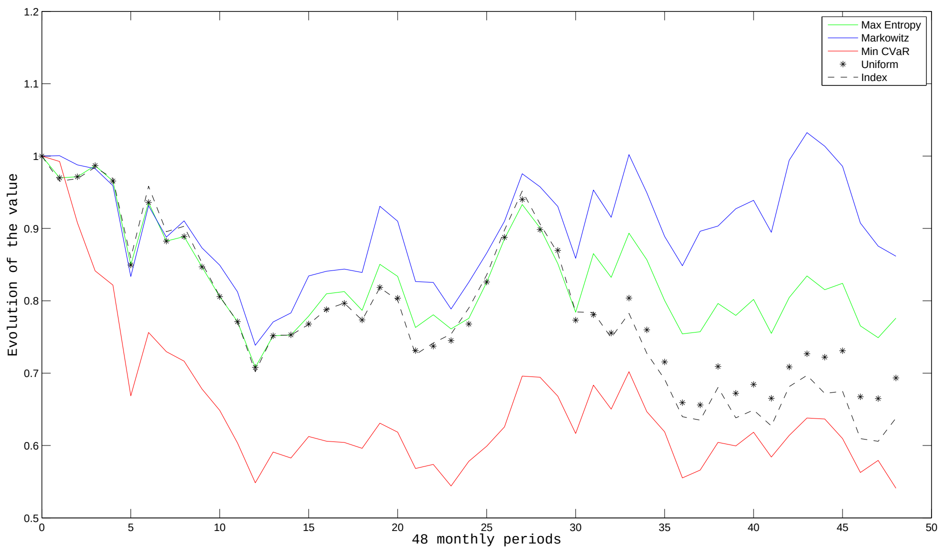


Figure 5.10: Evolution of an initial value of € 1.- with different type of portfolio management asking for a 0.15% or more monthly return over 48 monthly periods with stocks in PRC.

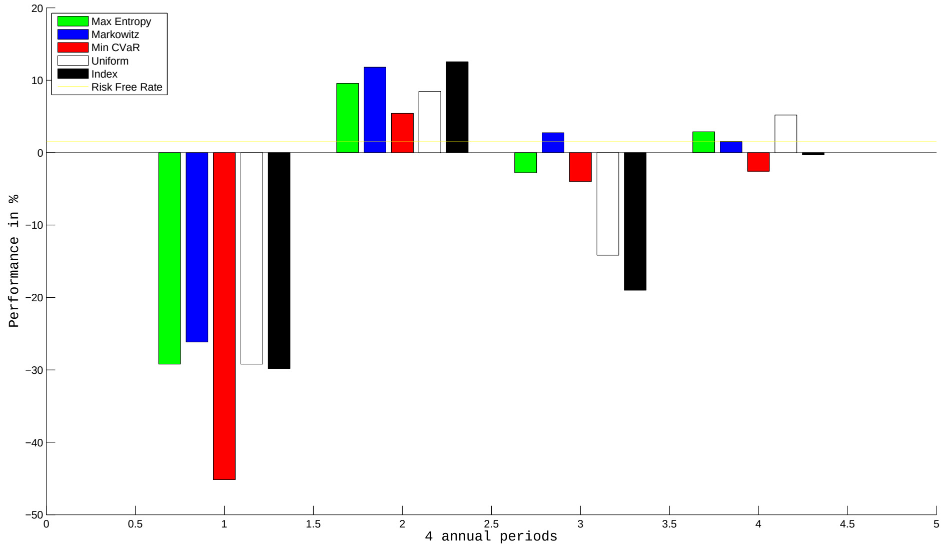


Figure 5.11: The four annual performance in % with stocks in PRC.

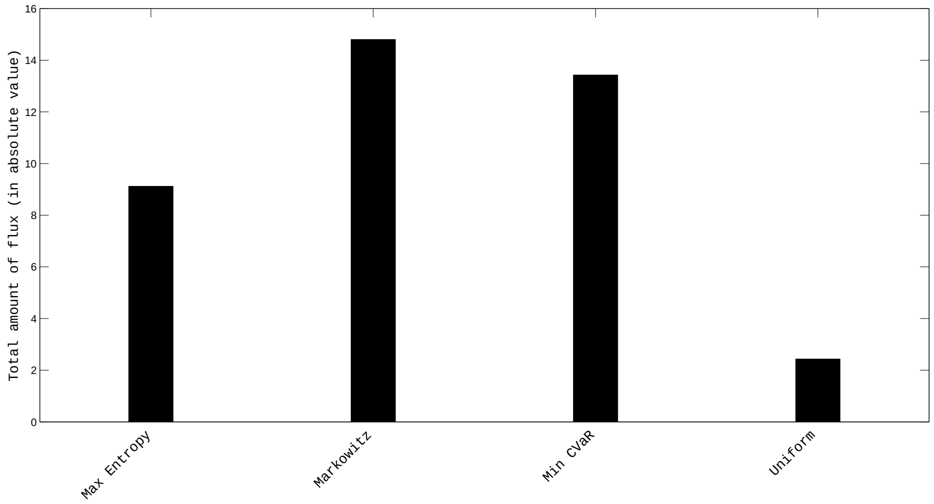


Figure 5.12: The total amount of flux for different type of portfolio management with stocks in PRC.

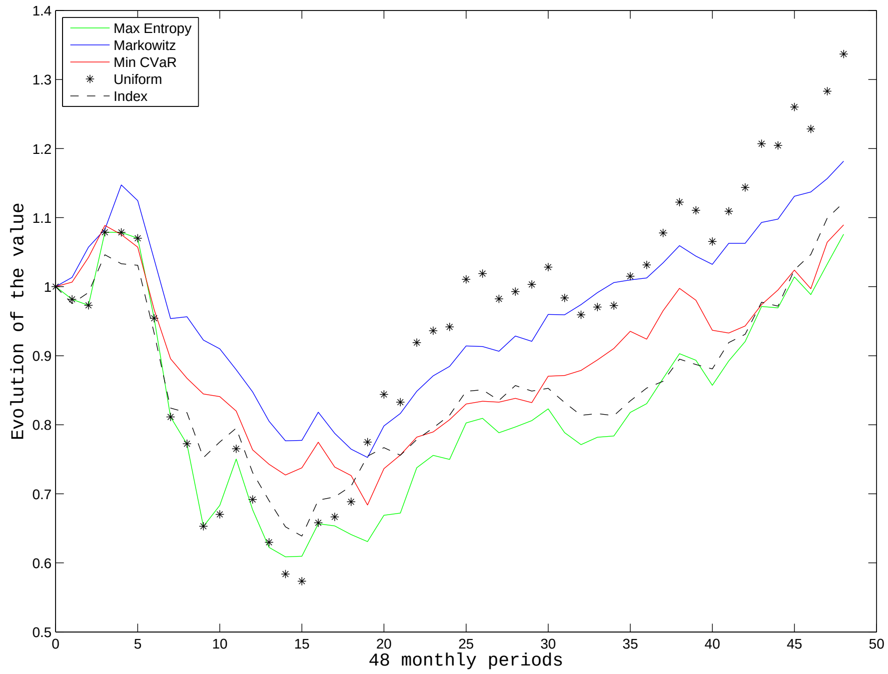


Figure 5.13: Evolution of an initial value of € 1.- with different type of portfolio management asking for a 0.5% or more monthly return over 48 monthly periods with stocks in CH.

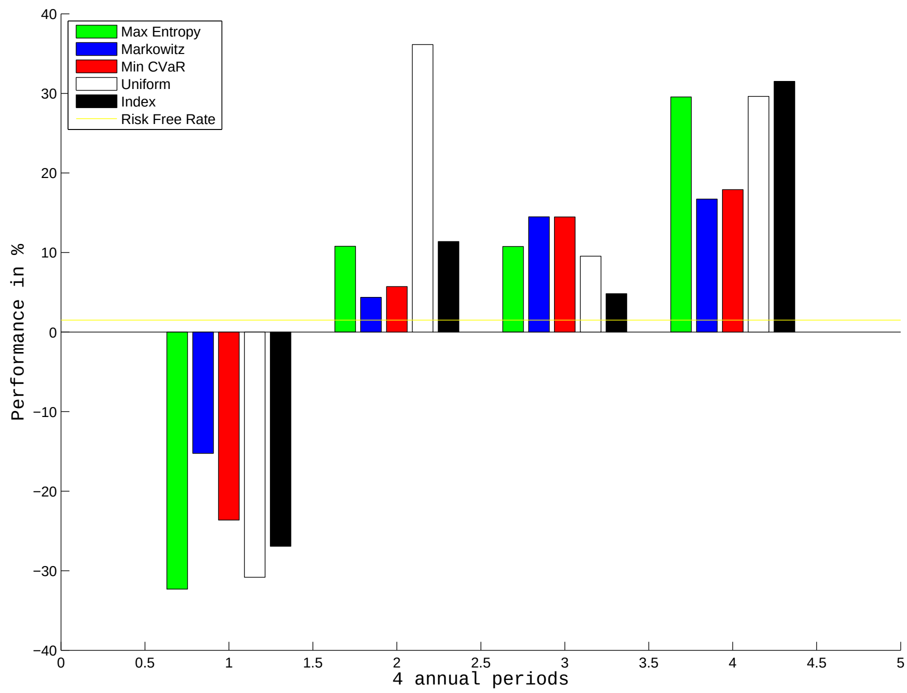


Figure 5.14: The four annual performance in % with stocks in CH.

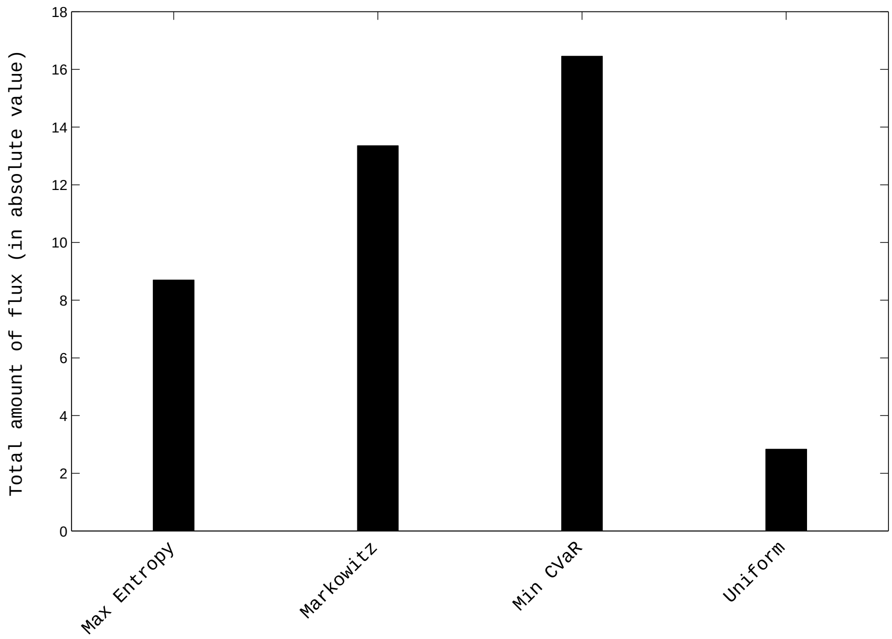


Figure 5.15: The total amount of flux for different type of portfolio management with stocks in CH.

Notes and Remarks From the first graph of the evolution of the initial value, we are tempted to conclude that our approach works well. It performed very well on the INDIA universe and this market is supposedly the one where we have little information. Compare to the other types of portfolio management, Max Entropy also performed well in the PRC univers. Min CVaR, however, was not very convincing in this market.

Concerning the OST univers, maximizing the entropy of the portfolio was very similar to the uniform portfolio. Max Entropy performed relatively well in the last two years but it did not take advantage of the bull market of the first two year, where Markowitz and Min CVaR did. We can see that the Min CVaR type of portfolio management was very performant and consistent, i.e., all the annual performances were 'good'. In the DOW univers, Max Entropy only managed to beat the index. In the third year, while Markowitz and Min CVaR managed to have a positive annual performance, Max Entropy fell back and did not managed to recover this loss in the last year.

However, Max Entropy did not performed well at all in the Swiss market where it did not even beat the index. In the first year, Max Entropy had the the worst annual performance. The three other annual performances were were not that bad, but not enough to catch up with the loss of the first year. On this market, Min CVaR was not very successful either. Both of theses types of portfolio management did not performed better than the indice. It is interesting to see that the uniform portfolio was the one that performed the best and by far in this market.

6 Conclusion

In general, maximizing the entropy of a portfolio can sometimes lead to a better performance than with other methods. However, it is very well known that

....past performances do not guarantee future returns ....

and so, it is difficult, if not impossible, to firmly conclude. One should note that in the five universes that we presented, the uniform portfolio, followed by the Max Entropy portfolio management type, were the ones that had the smallest flux. The flux of the uniform portfolio ranged between € 2.- to € 4.-, while the flux of Min CVaR (which was all ways the largest except for the universe PRC, where it was slightly smaller than the flux of Markowitz), was between € 13.- and € 20.- . In practice, this means that there would have been smaller transaction costs for the uniform portfolio or Max Entropy than for Min CVaR or Markowitz.

We have other types of portfolio management in mind that one could investigate. One problem that we can consider would be:

Problem 5.

$$\min_{\pi \in \mathbb{R}^N} \{-H(\pi)\}$$

with the constraints

1. $\pi_i \geqslant 0 \ \forall \ i \in \{1, ..., N\}$

2. $\sum_{i=1}^{N} \pi_i = 1$

3. $\mathbb{V}ar[r^{\pi}] \leqslant r$ *to guarantee a maximum theoretical risk* r *or less*

Note: The type of portfolio management where we solve Problem 5 for each monthly period, is called `Max Entropy (RISK)` and by `Markowitz (RISK)`, we mean the type of portfolio management where we solve the Problem 6, which we present just below.

Problem 6.

$$\max_{\pi \in \mathbb{R}^N} \{\mathbb{E}[r^{\pi}]\}$$

with the constraints

1. $\pi_i \geqslant 0 \ \forall \ i \in \{1, ..., N\}$
2. $\sum_{i=1}^{N} \pi_i = 1$
3. $\mathbb{V}ar[r^{\pi}] \leqslant r$ *to guarantee a maximum theoretical risk* r *or less*

We have done some tests, and we haven seen that in this case, our approach performed well on the Swiss market. For this test, the parameter r was taken to be $r = 7\%$.

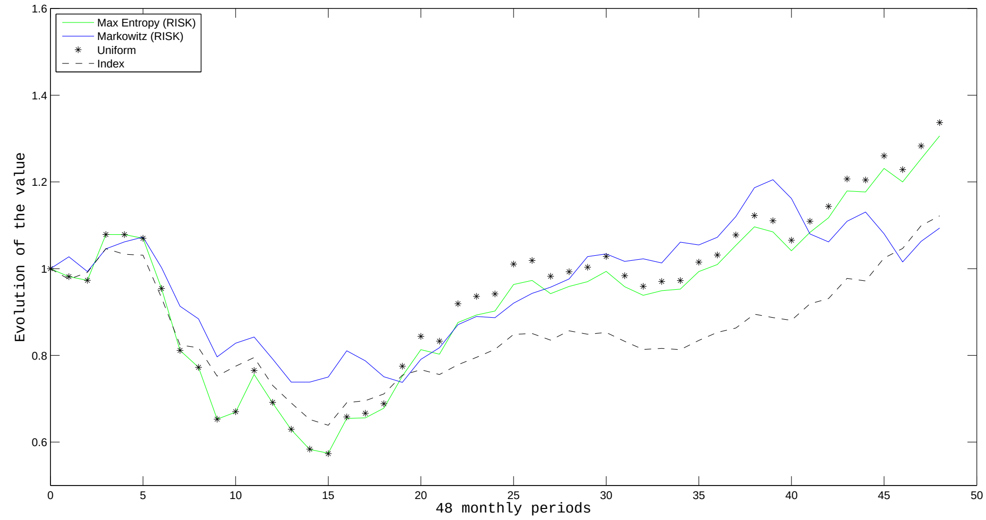


Figure 6.1: Evolution of an initial value of € 1.- with different type of portfolio management demanding for a 7% or less variance over 48 monthly periods with stocks in CH.

One could also investigate the effects if we worked with 'Relative-Entropy'. That is, for fixed portfolio weights q_i (ex. the index),

$$H_{R-E}(\pi) = -\sum_{i=1}^{N} \pi_i \log(\frac{\pi_i}{q_i}) \ .$$

This could be interesting to compare with ETFs (Exchange Traded Funds). One could continue and investigate other strategies such as: for $\lambda_1, \lambda_2, \lambda_3 \in [0,1]$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$

$$\pi_\lambda := \lambda_1 \pi_{VAR} + \lambda_2 \pi_{CVaR} + \lambda_3 \pi_{ME},$$

where π_{VAR} minimizes the variance, π_{CVaR} minimizes CVaR and π_{ME} maximizes entropy, all three under the usual constraints.

Note: The figure on the first page shows the entropy (on the z-axis) of all the possible portfolios composed of the three stocks: Česk Telecom, Mol, Millennium Bank. On the x-axis, we have the risk (measured with the CVaR), on the y-axis, we have the expected return and the z-axis is the entropy. The three 'foot' correspond to the portfolios that are totally concentrated on one stock. At these points, the entropy is zero, and on the x-axis and y-axis we have the risk (CVaR) and expected return respectively for the stock in question. The rest of the black dots show all the other portfolios where the hight of each black dot is the entropy of the corresponding portfolio.

List of Figures

Appendix

A Uniqueness of the solution for Problem 1

We will show that Problem 1 has a unique solution if σ is invertible.

Definition 9. *We say that the matrix $A \in \mathbb{M}\mathrm{at}_{N,N}(\mathbb{R})$ is* `positive semi-definite` *if*

$$x^\top A\, x \geqslant 0 \;\forall\; x \in \mathbb{R}^N.$$

Definition 10. *We say that the matrix $A \in \mathbb{M}\mathrm{at}_{N,N}(\mathbb{R})$ is* `positive definite` *if*

$$x^\top A\, x > 0 \;\forall\; x \in \mathbb{R}^N \setminus \{0\}.$$

Theorem 3. *A variance-covariance matrix σ is positive semi-definite.*

Proof. Since, by definition, the covariance between two random variables X_i and X_j is $\mathbb{C}\mathrm{ov}[X_i, X_j] := \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$, which will be denoted by $\sigma_{i,j}$, we have for all $x \in \mathbb{R}^N$

$$\begin{aligned}
x^\top \sigma\, x &= \sum_{i=1}^N \sum_{j=1}^N x_i x_j \sigma_{i,j} \\
&= \sum_{i=1}^N \sum_{j=1}^N x_i x_j (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]) \\
&= \sum_{i=1}^N \sum_{j=1}^N x_i x_j \mathbb{E}[X_i X_j] - \sum_{i=1}^N \sum_{j=1}^N x_i x_j \mathbb{E}[X_i]\mathbb{E}[X_j] \\
&= \mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N x_i x_j X_i X_j] - \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}[x_i X_i]\mathbb{E}[x_j X_j] \\
&= \mathbb{E}[(\sum_{i=1}^N x_i X_i)^2] - (\mathbb{E}[\sum_{i=1}^N x_i X_i])^2 \\
&= \mathbb{V}\mathrm{ar}[\sum_{i=1}^N x_i X_i]
\end{aligned}$$

and a variance is always greater or equal to zero. □

We now show that if σ admits an inverse, then σ is positive definite. To this end, we use the spectral theorem in $\mathbb{R}$:

Theorem 4. *Let $A \in \mathbb{M}\mathrm{at}_{N,N}(\mathbb{R})$ be a symmetric matrix. Then*

- $\mathbb{S}\mathrm{pec}_A \subset \mathbb{R}$

- *A can be diagonalized*
- *$\exists\ \{x_1, \dots, x_N\} \subset \mathbb{R}^N$ an orthogonal basis of eigenvectors of A ($Ax_i = \lambda_i x_i,\ \lambda_i \in \mathbb{R}\ \forall\ i$ and $\langle x_i | x_j \rangle = 0\ \forall\ i, j$ and $i \neq j$)*

Theorem 5. *Let $A \in \mathrm{Mat}_{N,N}(\mathbb{R})$ be a symmetric, positive semi-definite matrix. Then*

$$\exists\ A^{-1} \Longleftrightarrow A \text{ is positive definite}$$

Proof. [$\Rightarrow$] Let $x \in \mathbb{R}^N \setminus \{0\}$. By the $\mathbb{S}$*pectral* theorem, we can express x in the basis formed by the eigenvectors ($x = \sum_{i=1}^{N} \alpha_i x_i$). Calculating the product, we have

$$\langle x | Ax \rangle = \langle \sum_{i=1}^{N} \alpha_i x_i | \sum_{i=1}^{N} \alpha_i \underbrace{Ax_i}_{=\lambda_i x_i} \rangle = \sum_{i,j} \langle \alpha_i x_i | \lambda_j \alpha_j x_j \rangle.$$

Since $\langle x_i | x_j \rangle = 0$ for all i, j and $i \neq j$, we have

$$\langle x | Ax \rangle = \sum_{i=1}^{N} \alpha_i^2 \lambda_i \|x_i\|^2. \tag{A.1}$$

Since A is positive semi-definite, this last expression must be non-negative, and from

- $\|x_i\|^2 > 0\ \forall\ i$ because $\{x_i\}_{i=1}^{N}$ formes a basis
- $\alpha_i^2 \geqslant 0\ \forall\ i$ and $\exists\ j$ (at least one) such that $\alpha_j > 0$ because $x \in \mathbb{R}^N \setminus \{0\}$
- (A.1) holds for all $x \in \mathbb{R}^N$ (and in particular for any α_i)

we conclude that $\lambda_i \geqslant 0\ \forall\ i$. If $\lambda_i > 0\ \forall\ i$, then we have that A is a positive definite matrix, concluding the first part of the proof. This is true because, by hypothesis, A is invertible. In fact, we have

$$\begin{aligned} \nexists\ A^{-1} &\Longleftrightarrow \exists\ x \neq 0 \text{ such that } Ax = 0 \\ &\Longleftrightarrow 0 \text{ is an eigen value} \end{aligned}$$

Therefore, $A^{-1}\exists \Longleftrightarrow \mathbb{S}pec_A \subset \mathbb{R} \setminus \{0\}$ ($\mathbb{S}pec_A \subset \mathbb{R}$ already since A is symmetric (spectral theorem)). Hence, $\lambda_i > 0\ \forall\ i$.

[$\Leftarrow$] Let $x \in \mathbb{R}^N \setminus \{0\}$. Because A is positive definite, we have

$$\langle x|Ax\rangle = \sum_{i=1}^{N} \alpha_i^2 \lambda_i \|x_i\|^2 > 0.$$

Therefore, as above, this is true for all $x \in \mathbb{R}^N$ (and in particular for any α_i) and so we must have $\lambda_i > 0 \ \forall\ i$. Again, from above, if all the eigenvalues are in $\mathbb{R} \setminus \{0\}$, then the invertible matrix A exists which concludes the proof of the theorem. $\square$

We will now see that if $A \in \mathbb{M}at_{N,N}(\mathbb{R})$ is positive definite, then there exists a unique x that solves the Problem 1. To see this we consider Problem 1 as a particular case of the more general problem:

Problem 7. *Let $A \in \mathrm{Mat}_{N,N}(\mathbb{R})$ either be a positive semi-definite or a positive definite matrix, $a \in \mathbb{R}^N$ be an N-dimensional vector and let ϕ be the function*

$$\phi(x) = \frac{1}{2} x^\top A\ x + a^\top x$$

The problem consist of finding an x that minimizes ϕ under the condition where x belongs to a predefined feasible set $F = Ic\ \cap\ Ec$, where $Ic := \{x \in \mathbb{R}^N | Bx \leqslant \alpha\}$ consists of the inequality constraints and $Ec := \{x \in \mathbb{R}^N | Cx = \beta\}$ consists of the equality constraints, with B and C real matrices of dimension $k \times N$ and $l \times N$ respectively, and α and β are predefined vectors.

Note: By defining the sets

$$Ic := \{\pi \in \mathbb{R}^N | \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & -1 \\ -\mu_1 & \cdots & \cdots & -\mu_N \end{pmatrix} \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix} \leqslant \begin{pmatrix} 0 \\ \vdots \\ 0 \\ p \end{pmatrix} \},$$

$$Ec := \{\pi \in \mathbb{R}^N | \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix} = 1\},$$

in our case, the feasible set is:

$$F := Ic\ \cap\ Ec.$$

Let us show, for Problem 7, that any local minimum is a global minimum if A is positive semi-definite and that the solution to Problem 7 is reduced to a unique x if A is positive definite.

If $F = \emptyset$, the constraints of the problem are such that the problem itself is not interesting since there is no feasible point. We suppose, therefore, that $F \neq \emptyset$. If is reduced to a single point $F = \{\check{x}\}$ (this could be the case if, in the equality constraints Ec, C is an $N \times N$ invertible matrix), then Problem 7 has one, and only one solution. Let us suppose that F is not reduced to a singleton. Let $\check{x} \in F$ be a local minimum, that is, there exists $\epsilon > 0$ such that

$$\forall\ x \in \mathcal{N}_\epsilon(\check{x}) := \{x \in \mathbb{R}^N |\ |\check{x}_i - x_i| \leqslant \epsilon, i = 1, \ldots, N\}$$

we have

$$\phi(\check{x}) \leqslant \phi(x).$$

Let $\bar{x} \in F$ and $\bar{x} \neq \check{x}$. Defining x_λ by

$$x_\lambda := (1-\lambda)\check{x} + \lambda\bar{x}, \lambda \in [0,1]$$

we have that $x_\lambda \in F$ because

$$Bx_\lambda = (1-\lambda)B\check{x} + \lambda B\bar{x} \leqslant (1-\lambda)\alpha + \lambda\alpha = \alpha.$$

Hence $x_\lambda \in Ic$. In addition,

$$Cx_\lambda = (1-\lambda)C\check{x} + \lambda C\bar{x} = (1-\lambda)\beta + \lambda\beta = \beta,$$

and hence $x_\lambda \in Ec$. Therefore, $x_\lambda \in F$ for all $\lambda \in [0,1]$.

We now focus on a particular point,

$$x_\xi := (1-\xi)\check{x} + \xi\bar{x},\ 0 < \xi \leqslant \frac{\epsilon}{\max_i\{|\check{x}_i - \bar{x}_i|\}}.$$

Since $\xi > 0$, we have that $x_\xi \neq \check{x}$. We also have

$$x_\xi \in \mathcal{N}_\epsilon(\check{x}),\ \forall\ \xi \text{ such that } 0 < \xi \leqslant \frac{\epsilon}{\max_i\{|\check{x}_i - \bar{x}_i|\}},$$

because, if one takes the largest value for ξ, one has:

$$x_\xi = \check{x} + \frac{\epsilon}{\max_i\{|\check{x}_i - \bar{x}_i|\}}(\bar{x} - \check{x}),$$

and therefore every coordinate of x_ξ differs from the corresponding coordinate of $\check{x}$ by no more than ϵ. One can take an $\check{\epsilon} \leqslant \epsilon$ sufficiently small so that

$$\frac{\check{\epsilon}}{\max_i\{|\check{x}_i - \bar{x}_i|\}}(\bar{x} - \check{x}) < 1,$$

and therefore, from what we have seen above, we have $x_\xi \in F \cap \mathcal{N}_\epsilon(\check{x})$ and thus

$$\phi(\check{x}) \leqslant \phi(x_\xi).$$

If A is positive semi-definite, we have

$$\frac{1}{2}(\lambda - \lambda^2)(\check{x}^\top - \bar{x}^\top)A(\check{x} - \bar{x}) =: \kappa \geqslant 0, \quad 0 < \lambda < 1.$$

Let us suppose that $\phi(\bar{x}) < \phi(\check{x})$. This implies that $\exists\, \eta > 0$ such that

$$\phi(\bar{x}) = \phi(\check{x}) - \eta.$$

To continue, we must establish the following equation: x_λ defined as above, one has

$$\phi(x_\lambda) = (1 - \lambda)\phi(\check{x}) + \lambda\phi(\bar{x}) - \frac{1}{2}(\lambda - \lambda^2)(\check{x}^\top - \bar{x}^\top)A(\check{x} - \bar{x}), \quad \text{(A.2)}$$

which follows from the simple but relatively long computation.

$$
\begin{aligned}
\phi(x_\lambda) &= \frac{1}{2}x_\lambda^\top A x_\lambda + a^\top x_\lambda \\
&= \frac{1}{2}[(1-\lambda)\breve{x}^\top + \lambda\bar{x}^\top]A[(1-\lambda)\breve{x} + \lambda\bar{x}] + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-\lambda)\breve{x}^\top A + \lambda\bar{x}^\top A][(1-\lambda)\breve{x} + \lambda\bar{x}] + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-\lambda)^2\breve{x}^\top A\breve{x} + \lambda(1-\lambda)\breve{x}^\top A\bar{x} + \lambda(1-\lambda)\bar{x}^\top A\breve{x} + \lambda^2\bar{x}^\top A\bar{x}] \\
&\quad + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-2\lambda+\lambda^2)\breve{x}^\top A\breve{x} + (\lambda-\lambda^2)\breve{x}^\top A\bar{x} + (\lambda-\lambda^2)\bar{x}^\top A\breve{x} + \lambda^2\bar{x}^\top A\bar{x}] \\
&\quad + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[\breve{x}^\top A\breve{x} - 2\lambda\breve{x}^\top A\breve{x} + \lambda^2\breve{x}^\top A\breve{x} + \lambda\breve{x}^\top A\bar{x} - \lambda^2\breve{x}^\top A\bar{x} \\
&\quad + \lambda\bar{x}^\top A\breve{x} - \lambda^2\bar{x}^\top A\breve{x} + \lambda^2\bar{x}^\top A\bar{x}] + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-\lambda)\breve{x}^\top A\breve{x} + \lambda\bar{x}^\top A\bar{x} - \lambda\bar{x}^\top A\bar{x} - \lambda\breve{x}^\top A\breve{x} \\
&\quad + \lambda^2(\breve{x}^\top A\breve{x} - \breve{x}^\top A\bar{x} - \bar{x}^\top A\breve{x} + \bar{x}^\top A\bar{x}) + \lambda\breve{x}^\top A\bar{x} + \lambda\bar{x}^\top A\breve{x}] \\
&\quad + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-\lambda)\breve{x}^\top A\breve{x} + \lambda\bar{x}^\top A\bar{x} - \lambda^2(-\breve{x}^\top A\breve{x} + \breve{x}^\top A\bar{x} + \bar{x}^\top A\breve{x} - \bar{x}^\top A\bar{x}) \\
&\quad + \lambda(-\breve{x}^\top A\breve{x} + \breve{x}^\top A\bar{x} + \bar{x}^\top A\breve{x} - \bar{x}^\top A\bar{x})] + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-\lambda)\breve{x}^\top A\breve{x} + \lambda\bar{x}^\top A\bar{x} + (\lambda-\lambda^2)(\breve{x}^\top A(\bar{x}-\breve{x}) + \bar{x}^\top A(\breve{x}-\bar{x}))] \\
&\quad + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[(1-\lambda)\breve{x}^\top A\breve{x} + \lambda\bar{x}^\top A\bar{x} + (\lambda-\lambda^2)((\bar{x}^\top - \breve{x}^\top)A(\breve{x}-\bar{x}))] \\
&\quad + a^\top[(1-\lambda)\breve{x} + \lambda\bar{x}] \\
&= \frac{1}{2}[-(\lambda-\lambda^2)(\breve{x}^\top - \bar{x}^\top)A(\breve{x}-\bar{x}) + (1-\lambda)\breve{x}^\top A\breve{x} + \lambda\bar{x}^\top A\bar{x}] \\
&\quad + (1-\lambda)a^\top\breve{x} + \lambda a^\top\bar{x} \\
&= (1-\lambda)[\frac{1}{2}\breve{x}^\top A\breve{x} + a^\top\breve{x}] + \lambda[\frac{1}{2}\bar{x}^\top A\bar{x} + a^\top\bar{x}] \\
&\quad - \frac{1}{2}(\lambda-\lambda^2)(\breve{x}^\top - \bar{x}^\top)A(\breve{x}-\bar{x}) \\
&= (1-\lambda)\phi(\breve{x}) + \lambda\phi(\bar{x}) - \frac{1}{2}(\lambda-\lambda^2)(\breve{x}^\top - \bar{x}^\top)A(\breve{x}-\bar{x}).
\end{aligned}
$$

If we replace κ and $\phi(\bar{x}) = \phi(\breve{x}) - \eta$ in Equation (A.2), we have

$$\phi(x_\lambda) = \phi(\breve{x}) - \lambda\eta - \kappa = \phi(\breve{x}) - (\lambda\eta + \kappa),$$

and since λ and η are strictly greater than zero we obtain

$$\phi(x_\lambda) < \phi(\check{x}),\ \forall\ \lambda\ \in\ (0,1),$$

and, in particular, for $\lambda = \xi$, and this is in contradiction to $\phi(\check{x}) \leqslant \phi(x_\xi)$. Therefore we conclude that $\phi(\check{x}) \leqslant \phi(\bar{x})$ if A is positive semi-definite. In other words, if A is positive semi-definite, then all local minima take the value of the one and only global minimum.

If A is positive definite, then, since $\bar{x} \neq \check{x}$,

$$\frac{1}{2}(\lambda - \lambda^2)(\check{x}^\top - \bar{x}^\top)A(\check{x} - \bar{x}) := \kappa > 0, \quad 0 < \lambda < 1.$$

Supposing that $\phi(\bar{x}) \leqslant \phi(\check{x})$, there exists an $\eta \geqslant 0$ such that

$$\phi(\bar{x}) = \phi(\check{x}) - \eta.$$

If we replace again κ and $\phi(\bar{x}) = \phi(\check{x}) - \eta$ in Equation (A.2), we have

$$\phi(x_\lambda) = \phi(\check{x}) - (\lambda\eta + \kappa),$$

and since κ is strictly greater than zero we obtain

$$\phi(x_\lambda) < \phi(\check{x}), \quad \forall\ \lambda\ \in\ (0,1),$$

and, we again have the same contradiction. Therefore we conclude that $\phi(\check{x}) < \phi(\bar{x})$ for all $\bar{x}\ \in F$ and $\bar{x} \neq \check{x}$ if A is positive definite. In other words, if A is positive definite, then there is only one local minimum, and hence only one global minimum that is reached by one and only one point $\check{x}$.

B Continuity of an integral with respect to a parameter

We present the theorem with a more general statement than what we need. Letting $(E, \mathfrak{A}, \mu)$ be a mesure space, we define

$$\begin{aligned}\mathbb{L}^1[E,\mu] \ := \ \{f : (E,\mathfrak{A}) \to (\mathbb{R}, \mathbb{B}_\mathbb{R}) \mid f \text{ is } \mu-\text{measurable function,} \\ \textstyle\int_E |f|\, d\mu < +\infty\}/\sim\end{aligned}$$

where $\mathbb{B}_\mathbb{R}$ is the Borel σ-algebra on $\mathbb{R}$ and $\sim$ is the equivalence relation defined by

$$f \sim g \Leftrightarrow f - g \in \{f : (E,\mathfrak{A}) \to (\mathbb{R}, \mathbb{B}_\mathbb{R}) \mid f \equiv 0\ \mu-\text{almost everywhere}\}.$$

Theorem 6. *We consider a measure space $(E, \mathfrak{A}, \mu)$, a topological metrizable space X and a function $g : X \times E \longrightarrow \mathbb{R}$. We suppose that*

1. *for $\mu-$almost all $y \in E$, $g(x, y)$ is continuous with respect to $x \in X$*
2. *$\forall\ x \in X, g(x, \cdot) \in \mathbb{L}^1(E, \mathfrak{A}, \mu)$*
3. *$\forall\ x_0 \in X, \exists\ \mathcal{N}(x_0)$ a neighborhood of x_0 and $\exists\ h \in \mathbb{L}^1(E, \mathfrak{A}, \mu)$ such that $\forall\ x \in \mathcal{N}(x_0)$ we have*

$$|g(x, \cdot)| \leqslant h \ \mu - \textit{almost everywhere.}$$

Then the function $x \longmapsto \int\limits_E g(x, y)\, d\mu(y)$ is continuous with respect to x.

A proof of this theorem can be found in [3], page 98.

C Equivalence on the hypothesis

The equivalence

$$\int\limits_{y \in \mathbb{R}^N_{\geqslant 0}} \|y\|_1 \rho(y)\, dy < +\infty \Longleftrightarrow \int\limits_{y \in \mathbb{R}^N_{\geqslant 0}} y_i \rho(y)\, dy < +\infty,\ i = 1, \ldots, N$$

is due to the following theorem that we present with slightly more general hypothesis.

Theorem 7. *Let $\rho : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geqslant 0}$ be a positive function. Then*

$$\int\limits_{y \in \mathbb{R}^N} y_i \rho(y)\, dy \ \textit{converges}\ ,\ i = 1, \ldots, N \Longleftrightarrow \int\limits_{y \in \mathbb{R}^N} |y_i| \rho(y)\, dy \ \textit{converges}\ ,$$

for $i = 1, \ldots, N$.

Proof. We show both sides of the equivalence with the same reasoning.

$$\begin{aligned}\int\limits_{y \in \mathbb{R}^N} y_i \rho(y)\, dy &= \int\limits_{y \in [0, +\infty[^N} y_i \rho(y)\, dy + \int\limits_{y \in]-\infty, 0[^N} y_i \rho(y)\, dy \\ &= \int\limits_{y \in [0, +\infty[^N} y_i \rho(y)\, dy - \int\limits_{y \in]-\infty, 0[^N} |y_i| \rho(y)\, dy.\end{aligned}$$

If we multiply the second integral by -1, it will still converge and so,

$$\int\limits_{y\in[0,+\infty[^N} |y_i|\rho(y)\,dy \qquad + \int\limits_{y\in]-\infty,0[^N} |y_i|\rho(y)\,dy = \int\limits_{y\in\mathbb{R}^N} |y_i|\rho(y)\,dy.$$

$\square$

We conclude the equivalence of the hypothesis with the following obvious equalities:

$$\sum_{i=1}^{N} \int\limits_{y\in\mathbb{R}^N} |y_i|\rho(y)\,dy = \int\limits_{y\in\mathbb{R}^N} \sum_{i=1}^{N} |y_i|\rho(y)\,dy = \int\limits_{y\in\mathbb{R}^N} \|y_i\|_1\rho(y)\,dy.$$

D Dominated Convergence Theorem

We give a particular statement of the Lebesgue Dominated Convergence Theorem. We denote by λ the Lebesgue mesure, that is:

$$\lambda([\alpha,\beta]) = \beta - \alpha, \quad \forall \text{ intervals } [\alpha,\beta] \subset \mathbb{R},$$

and the set of Lebesgue integrable function over $\mathbb{R}$ is

$$\mathbb{L}[\mathbb{R},\lambda] \;:=\; \{f : (\mathbb{R},\mathbb{B}_\mathbb{R}) \to (\mathbb{R},\mathbb{B}_\mathbb{R}) \mid f \text{ is a Borel measurable function, } \textstyle\int_\mathbb{R} |f|\,d\lambda < +\infty\}/\sim$$

where $\mathbb{B}_\mathbb{R}$ is the Borel σ-algebra on $\mathbb{R}$ and $\sim$ is the equivalence relation

$$f \sim g \Leftrightarrow f - g \in \{f : (\mathbb{R},\mathbb{B}_\mathbb{R}) \to (\mathbb{R},\mathbb{B}_\mathbb{R}) \mid f \equiv 0 \ \lambda-\text{almost everywhere}\}$$

Theorem 8. *Let $\{f_n\}_{n=0}^{\infty}$ be a sequence in $\mathbb{L}[\mathbb{R},\lambda]$. We suppose the following:*

- $\forall\ x \in \mathbb{R},\ f(x) = \lim_{n\to\infty} f_n(x)$ *(point by point convergence)*
- $\exists\ g \in \mathbb{L}[\mathbb{R},\lambda]$ *such that*
  - $0 \leqslant g(x)\ \forall\ x \in \mathbb{R}$
  - $|f_n(x)| \leqslant g(x)\ \forall\ n \in \mathbb{N}, \forall\ x \in \mathbb{R}$

We therefore have:

- $f \in \mathbb{L}[\mathbb{R},\lambda]$
- $\lim_{n\to\infty} \int_\mathbb{R} |f_n - f|\,d\lambda = 0$
- $\lim_{n\to\infty} \int_\mathbb{R} f_n\,d\lambda = \int_\mathbb{R} f\,d\lambda$

A proof of this theorem can be found in [12], page 24.

E Convex functions and their minimum

Theorem 9. *Let $F : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be a convex function such that F is continuously differentiable with respect to its last variable, i.e. $\forall\, x \in \mathbb{R}^n$ fixed, $\frac{\partial F_x}{\partial \alpha}(\alpha)$ exists and is continuous. We suppose that there exists α_0 such that $\frac{\partial F_x}{\partial \alpha}(\alpha_0) = 0$. Then*

$$\forall\, x \in \mathbb{R}^n fixed,\ F_x(\alpha_0) \text{ is a minimum.}$$

Proof. We suppose that there exists $\alpha \neq \alpha_0$ such that $F_x(\alpha) < F_x(\alpha_0)$. Without loss of generality, we can suppose that $\alpha - \alpha_0 > 0$. Since F_x is convex, then, for all $\lambda \in (0,1)$,

$$\begin{aligned} F_x(\lambda\alpha + (1-\lambda)\alpha_0) &\leqslant \lambda F_x(\alpha) + (1-\lambda)F_x(\alpha_0) \\ F_x(\alpha_0 + \lambda(\alpha - \alpha_0)) - F_x(\alpha_0) &\leqslant \lambda F_x(\alpha) + F_x(\alpha_0) - \lambda F_x(\alpha_0) - F_x(\alpha_0) \end{aligned}$$

We divide both sides by $\frac{1}{\lambda(\alpha-\alpha_0)}$, and we obtain

$$\underbrace{\frac{F_x(\alpha_0 + \overbrace{\lambda(\alpha - \alpha_0)}^{=:h}) - F_x(\alpha_0)}{\underbrace{\lambda(\alpha - \alpha_0)}_{=:h}}}_{\xrightarrow[h\to 0]{} 0} \leqslant \frac{\lambda(F_x(\alpha) - F_x(\alpha_0))}{\lambda(\alpha - \alpha_0)} < 0.$$

The left hand side of the inequality converges to 0 when $h \to 0$, since the function is differentiable, and so we have a contradiction, so $F_x(\alpha) \geqslant F_x(\alpha_0)$. $\square$

Theorem 10. *Let $F : \mathbb{R}^N \longrightarrow \mathbb{R}$ be a convex function. Let $x_0,\ \check{x} \in \mathbb{R},\ x_0 \neq \check{x}$ be such that $F(x_0)$ and $F(\check{x})$ are two local minima. Then*

$$F(x_0) = F(\check{x}).$$

That is, for a convex function, all local minima are the global minimum.

Proof. Suppose that $F(x_0) \neq F(\check{x})$. Without loss of generality, assume that $F(x_0) > F(\check{x})$. Since x_0 is a local minimum, there exists $\epsilon > 0$ such that

$$F(x_0) \leqslant F(x), \quad \forall\, x \in \mathcal{N}_\epsilon(x_0) := \{x \in \mathbb{R}^N |\ |x_i - x_{0_i}| < \epsilon,\ i = 1, \ldots, N\}.$$

Let $\lambda_0 := \frac{\epsilon_0}{\max_i\{|\check{x}_i - x_{0_i}|\}}$ and chose ϵ_0 such that $0 < \epsilon_0 \leqslant \epsilon$ and $\lambda_0 \in (0,1)$. We then obtain

$$\lambda_0 \check{x} + (1-\lambda_0)x_0 \in \mathcal{N}_\epsilon(x_0),$$

since, $\forall\ i = 1, \dots, N$ we have

$$\begin{aligned} |\lambda_0 \check{x}_i + (1-\lambda_0)x_{0_i} - x_{0_i}| &= |\tfrac{\epsilon_0}{\max_i\{|\check{x}_i - x_{0_i}|\}}\check{x}_i - \tfrac{\epsilon_0}{\max_i\{|\check{x}_i - x_{0_i}|\}}x_{0_i}| \\ &= \tfrac{\epsilon_0}{\max_i\{|\check{x}_i - x_{0_i}|\}}|\check{x}_i - x_{0_i}| \leqslant \epsilon_0. \end{aligned}$$

Therefore $F(\lambda_0\check{x} + (1-\lambda_0)x_0) \geqslant F(x_0)$. Since $F(\check{x}) - F(x_0) < 0$, $\lambda_0(F(\check{x}) - F(x_0)) < 0$ and hence

$$\begin{aligned} F(\lambda_0\check{x} + (1-\lambda_0)x_0) &> F(x_0) + \lambda_0(F(\check{x}) - F(x_0)) \\ &> \lambda_0 F(\check{x}) + (1-\lambda_0)F(x_0). \end{aligned}$$

This is a contradiction to the fact that F is convex. Therefore, $F(x_0) = F(\check{x})$. $\square$

F ‘min of min’

Theorem 11. *Let $G : X \times Y \longrightarrow \mathbb{R}$ be a function such that there exists $a \in \mathbb{R}$ such that*

$$a := \min_{x\in X}\{\min_{y\in Y}\{G(x,y)\}\}.$$

Then there exists $\check{a} \in \mathbb{R}$ such that

$$\check{a} = \min_{(x,y)\in X\times Y}\{G(x,y)\}$$

and $a = \check{a}$.

Proof. Let $b := \inf_{(x,y)\in X\times Y}\{G(x,y)\}$ belong to $\bar{\mathbb{R}}$, that is, $-\infty = b$ or $-\infty < b$. We suppose that $b < a$. We consider the two possible cases:

- $-\infty = b$

  $\forall\ c < a,\ \exists\ (x_c, y_c) \in X \times Y$ such that

  $G(x_c, y_c) \leqslant c < a$, and so we have: $\min\limits_{y\in Y}\{G(x_c,y)\} \leqslant G(x_c,y_c) \leqslant c < a$

  However, we also have the following inequalities

  $$\underbrace{\min_{x\in X}\{\min_{y\in Y}\{G(x,y)\}\}}_{=a} \leqslant \min_{y\in Y}\{G(x_c,y)\} \leqslant G(x_c,y_c) \leqslant c < a.$$

  We therefore have a contradiction.

- $-\infty < b < a$
  $\forall\ c$ such that $b < c < a, \exists\ (x_c, y_c) \in X \times Y$ such that
  $$G(x_c, y_c) \leqslant c < a.$$
  Following the same reasoning as above, we will again encounter the same contradiction. Therefore, $b \geqslant a$. We will show that $b \leqslant a$. By hypothesis, $\exists\ (x_0, y_{x_0})$ such that $G(x_0, y_{x_0}) = a$. By definition, $b = \inf_{(x,y)\in X\times Y}\{G(x, y)\}$, therefore, in particular, it is smaller than a. We have
  $$b \leqslant G(x_0, y_{x_0}) = a, \ \Rightarrow b \leqslant a$$
  Hence $b = a$, and since a exists by hypothesis, then b exists and because
  $$b = \inf_{(x,y)\in X\times Y}\{G(x, y)\} = a = G(x_0, y_{x_0}),$$
  b attains a value of G and so
  $$b = \min_{(x,y)\in X\times Y}\{G(x, y)\}.$$

□

G Equivalence between Problem 2 (with the function $\tilde{F}_\beta$) and Problem 3

The equivalence between Problem 2 (with the function $\tilde{F}_\beta$) and Problem 3 is due to the following arguments.

Let (π^*, α^*) be a solution for Problem 2, then (π^*, α^*, z^*), where $z_j^* := [1 - \langle\pi^*|y_j\rangle - \alpha^*]^+$, is a solution for Problem 3, because if not, then there exist $(\tilde{\pi}, \tilde{\alpha}, \tilde{z})$ such that

$$\tilde{F}_\beta(\pi^*, \alpha^*) = \alpha^* + \mu\sum_{j=1}^{n} z_j^* > \tilde{\alpha} + \mu\sum_{j=1}^{n}\tilde{z}_j,$$

with $\tilde{z}_j \geqslant 0$ and $\tilde{z}_j \geqslant 1 - \langle\tilde{\pi}|y_j\rangle - \tilde{\alpha}$. Defining $\check{z}_j := 1 - \langle\tilde{\pi}|y_j\rangle - \tilde{\alpha}$, the we have $\tilde{z}_j \geqslant [\check{z}_j]^+$, meaning that

$$\tilde{F}_\beta(\pi^*, \alpha^*) > \tilde{\alpha} + \mu\sum_{j=1}^{n}[\check{z}_j]^+ = \tilde{F}_\beta(\tilde{\pi}, \tilde{\alpha}),$$

which is in contradiction to the fact that (π^*, α^*) realizes the minimum of the function $\tilde{F}_\beta$.

Suppose now that (π^*, α^*, z^*) is a solution for Problem 3, then (π^*, α^*) is a solution for Problem 2 (with the function $\tilde{F}_\beta$). If this is not true, then there exist $(\tilde{\pi}, \tilde{\alpha})$ such that

$$\alpha^* + \mu \sum_{j=1}^{n} [1 - \langle \pi^* | y_j \rangle - \alpha^*]^+ > \tilde{\alpha} + \mu \sum_{j=1}^{n} [1 - \langle \tilde{\pi} | y_j \rangle - \tilde{\alpha}]^+$$

By the fifth constraint of the Problem 3, we have

$$\alpha^* + \mu \sum_{j=1}^{n} z_j^* \geqslant \alpha^* + \mu \sum_{j=1}^{n} [1 - \langle \pi^* | y_j \rangle - \alpha^*]^+$$

Defining $\tilde{z}_j := [1 - \langle \tilde{\pi} | y_j \rangle - \tilde{\alpha}]^+$, we have

$$\alpha^* + \mu \sum_{j=1}^{n} z_j^* > \tilde{\alpha} + \mu \sum_{j=1}^{n} \tilde{z}_j,$$

which is in contradiction to the fact that (π^*, α^*, z^*) is a solution for Problem 3.

H Uniqueness of the solution for Problem 2

We recall the function F_β but in a more general setting (where we integrate over all $\mathbb{R}^N$ and not only over $\mathbb{R}^N_{\geqslant 0}$),

$$\begin{array}{rccl} F_\beta : & \Pi \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & (\pi, \alpha) & \longmapsto & \alpha + (1-\beta)^{-1} \displaystyle\int_{y \in \mathbb{R}^N} [f(\pi, y) - \alpha]^+ \rho(y)\, dy, \end{array}$$

where $\Pi \subset \mathbb{R}^N$ and Π is the set of available portfolios.

We show the calculations for the case $N = 2$. In this case, we can substitute for one variable: $\pi_1 + \pi_2 = 1 \Leftrightarrow \pi_2 = 1 - \pi_1$. This substitution leads to

$$1 - \pi_1 y_1 - \pi_2 y_2 - \alpha = 1 - \pi_1 y_1 - (1 - \pi_1) y_2 - \alpha = 1 - \pi_1 (y_1 - y_2) - y_2 - \alpha.$$

Therefore our function is reduced to two variables (π_1, α), and so we will work with (π, α),

$$F_\beta(\pi, \alpha) = \alpha + (1-\beta)^{-1} \int_{y \in \mathbb{R}^2} [1 - \pi(y_1 - y_2) - y_2 - \alpha]^+ \rho(y)\, dy.$$

We will first calculate the first derivative with respect to π.

$$\begin{aligned}
\tfrac{\partial F_\beta}{\partial \pi}(\pi, \alpha) &= (1-\beta)^{-1} \frac{\partial}{\partial \pi} \Big(\int_{y \in \mathbb{R}^2} [1 - \pi(y_1 - y_2) - y_2 - \alpha]^+ \rho(y)\, dy \Big) \\
&= (1-\beta)^{-1} \frac{\partial}{\partial \pi} \Big(\int_{1-\pi(y_1-y_2)-y_2-\alpha \geqslant 0} (1 - \pi(y_1 - y_2) - y_2 - \alpha) \rho(y)\, dy \Big).
\end{aligned}$$

The set $\{y \in \mathbb{R}^2 | 1 - \pi(y_1 - y_2) - y_2 - \alpha \geqslant 0\}$ is the same as $\{y \in \mathbb{R}^2 | \frac{1-\alpha-\pi y_1}{1-\pi} \geqslant y_2\}$, so our integral becomes

$$= (1-\beta)^{-1} \frac{\partial}{\partial \pi} \Big(\int_{-\infty}^{+\infty} \int_{-\infty}^{\frac{1-\alpha-\pi y_1}{1-\pi}} (1 - \pi(y_1 - y_2) - y_2 - \alpha) \rho(y_1, y_2)\, dy_2\, dy_1 \Big).$$

We switch the first integral with the derivative, and we are interested in calculating

$$\frac{\partial}{\partial \pi} \Big(\int_{-\infty}^{\frac{1-\alpha-\pi y_1}{1-\pi}} (1 - \pi(y_1 - y_2) - y_2 - \alpha) \rho(y_1, y_2)\, dy_2 \Big).$$

We will do this with the help of the Leibniz integral rule (refer to [8, pp 272]),

$$\frac{\partial}{\partial x} \Big(\int_{a(x)}^{b(x)} K(x, \tau)\, d\tau \Big) = \int_{a(x)}^{b(x)} \frac{\partial K}{\partial x}(x, \tau)\, d\tau + \big[K(x, b(x)) b'(x) - K(x, a(x)) a'(x) \big].$$

If we let $K(\pi, y_2) = (1 - \pi(y_1 - y_2) - y_2 - \alpha) \rho(y_1, y_2)$, we obtain

$$\frac{\partial K}{\partial \pi}(\pi, y_2) = (y_2 - y_1) \rho(y_1, y_2).$$

It is obvious that $K(\pi, \frac{1-\alpha-\pi y_1}{1-\pi})$ equals zero. $K(\pi, a(\pi))a'(\pi)$ is zero. Therefore

$$\frac{\partial F_\beta}{\partial \pi}(\pi, \alpha) = (1-\beta)^{-1}\Big(\int_{-\infty}^{+\infty} \int_{-\infty}^{\frac{1-\alpha-\pi y_1}{1-\pi}} (y_2 - y_1)\rho(y_1, y_2)\, dy_2\, dy_1 \Big).$$

In Theorem 1, we have seen that

$$\frac{\partial F_\beta}{\partial \alpha}(\pi, \alpha) = (1-\beta)^{-1}\Big[\int_{f(\pi,y)\leqslant\alpha} \rho(y)\, dy - \beta \Big].$$

Since in this case the derivative was taken with respect to α, we can substitue the constraint (that concerns only the first packet of variables π) in the derivative. Here we see that the inequality in the domain of integration is of opposite sign as the previous one, so the derivative with respect to α is

$$\frac{\partial F_\beta}{\partial \alpha}(\pi, \alpha) = (1-\beta)^{-1}\Big(\int_{-\infty}^{+\infty} \int_{\frac{1-\alpha-\pi y_1}{1-\pi}}^{+\infty} \rho(y_1, y_2)\, dy_2\, dy_1 \Big).$$

We use again the Leibniz integral rule to calculate the second order derivatives.

[$\frac{\partial^2 F_\beta}{\partial \pi^2}$] Let us calculate the second derivative with respect to π for the first derivative with respect to π.

$$\frac{\partial^2 F_\beta}{\partial \pi^2}(\pi, \alpha) = (1-\beta)^{-1}\Big(\int_{-\infty}^{+\infty} \frac{\partial}{\partial \pi} \int_{-\infty}^{\frac{1-\alpha-\pi y_1}{1-\pi}} (y_2 - y_1)\rho(y_1, y_2)\, dy_2\, dy_1 \Big).$$

If we let $K(\pi, y_2) = (y_2 - y_1)\rho(y_1, y_2)$ then

$$\frac{\partial K}{\partial \pi}(\pi, y_2) = 0,$$

and if we let $b(\pi) = \frac{1-\alpha-\pi y_1}{1-\pi}$, then

$$b'(\pi) = \frac{-y_1(1-\pi) - (1-\alpha-\pi y_1)(-1))}{(1-\pi)^2} = \frac{1-\alpha-y_1}{(1-\pi)^2}.$$

We now calculate $K(\pi, b(\pi))$:

$$\begin{aligned}
K(\pi, \frac{1-\alpha-\pi y_1}{1-\pi}) &= (\frac{1-\alpha-\pi y_1}{1-\pi} - y_1)\, \underbrace{\rho(y_1, \frac{1-\alpha-\pi y_1}{1-\pi})}_{=:\bar{\rho}(y_1)} \\
&= \frac{1-\alpha-\pi y_1 - y_1(1-\pi)}{1-\pi}\bar{\rho}(y_1) \\
&= \frac{1-\alpha-y_1}{1-\pi}\bar{\rho}(y_1).
\end{aligned}$$

In addition, $K(\pi, a(\pi))a'(\pi)$ is zero. Therefore

$$\frac{\partial^2 F_\beta}{\partial \pi^2}(\pi, \alpha) = \frac{(1-\beta)^{-1}}{(1-\pi)^3} \int_{-\infty}^{+\infty} (1-\alpha-y_1)^2 \bar{\rho}(y_1)\, dy_1.$$

[$\frac{\partial^2 F_\beta}{\partial \pi \partial \alpha}$] Let us calculate the second derivative with respect to α for the first derivative with respect to π.

$$\frac{\partial^2 F_\beta}{\partial \pi \partial \alpha}(\pi, \alpha) = (1-\beta)^{-1}\Big(\int_{-\infty}^{+\infty} \frac{\partial}{\partial \alpha} \int_{-\infty}^{\frac{1-\alpha-\pi y_1}{1-\pi}} (y_2 - y_1)\rho(y_1, y_2)\, dy_2\, dy_1 \Big).$$

If we let $K(\alpha, y_2) = (y_2 - y_1)\rho(y_1, y_2)$ then

$$\frac{\partial K}{\partial \alpha}(\alpha, y_2) = 0,$$

and if we let $b(\alpha) = \frac{1-\alpha-\pi y_1}{1-\pi}$ then

$$b'(\alpha) = \frac{-1}{1-x}.$$

We now calculate $K(\alpha, b(\alpha))$.

$$\begin{aligned}
K(\alpha, \frac{1-\alpha-\pi y_1}{1-\pi}) &= (\frac{1-\alpha-\pi y_1}{1-\pi} - y_1)\bar{\rho}(y_1) \\
&= \frac{1-\alpha-y_1}{1-\pi}\bar{\rho}(y_1),
\end{aligned}$$

In addition, $K(\alpha, a(\alpha))a'(\alpha)$ is zero. Therefore

$$\frac{\partial^2 F_\beta}{\partial \pi^2}(\pi, \alpha) = \frac{-(1-\beta)^{-1}}{(1-\pi)^2} \int_{-\infty}^{+\infty} (1-\alpha-y_1)\bar{\rho}(y_1)\, dy_1.$$

[$\frac{\partial^2 F_\beta}{\partial\alpha\partial\pi}$] Let us calculate the second derivative with respect to π for the first derivative with respect to α. We should get the same result as above which serves as a verification,

$$\frac{\partial^2 F_\beta}{\partial\alpha\partial\pi}(\pi,\alpha) = (1-\beta)^{-1}\Big(\int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial\pi} \int\limits_{\frac{1-\alpha-\pi y_1}{1-\pi}}^{+\infty} \rho(y_1,y_2)\,dy_2\,dy_1\Big).$$

If we let $K(\pi, y_2) = \rho(y_1, y_2)$ then

$$\frac{\partial K}{\partial\pi}(\pi, y_2) = 0,$$

and if we let $a(\pi) = \frac{1-\alpha-\pi y_1}{1-\pi}$ then

$$a'(\pi) = \frac{1-\alpha-y_1}{(1-\pi)^2}.$$

We now calculate $K(\pi, a(\pi))$:

$$K(\pi, a(\pi)) = \bar{\rho}(y_1).$$

In addition, $K(\pi, b(\pi))b'(\pi)$ is zero. Therefore

$$\frac{\partial^2 F_\beta}{\partial\alpha\partial\pi}(\pi,\alpha) = \frac{-(1-\beta)^{-1}}{(1-\pi)^2} \int\limits_{-\infty}^{+\infty} (1-\alpha-y_1)\bar{\rho}(y_1)\,dy_1.$$

[$\frac{\partial^2 F_\beta}{\partial\alpha^2}$] Let us calculate the second derivative with respect to α for the first derivative with respect to α.

$$\frac{\partial^2 F_\beta}{\partial\alpha^2}(\pi,\alpha) = (1-\beta)^{-1}\Big(\int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial\alpha} \int\limits_{\frac{1-\alpha-\pi y_1}{1-\pi}}^{+\infty} \rho(y_1,y_2)\,dy_2\,dy_1\Big).$$

If we let $K(\alpha, y_2) = \rho(y_1, y_2)$ then

$$\frac{\partial K}{\partial\alpha}(\alpha, y_2) = 0,$$

and if we let $a(\alpha) = \frac{1-\alpha-\pi y_1}{1-\pi}$ then

$$a'(\alpha) = \frac{-1}{(1-\pi)}.$$

We now calculate $K(\alpha, a(\alpha))$.

$$K(\alpha, a(\alpha)) = \bar{\rho}(y_1)$$

$K(\alpha, b(\alpha))b'(\alpha)$ is zero. Therefore

$$\frac{\partial^2 F_\beta}{\partial \alpha^2}(\pi, \alpha) = \frac{(1-\beta)^{-1}}{1-\pi} \int\limits_{-\infty}^{+\infty} \bar{\rho}(y_1)\, dy_1.$$

The Hessian matrix $D[\nabla(F_\beta)]_{(\pi,\alpha)}$ is therefore given by

$$\begin{pmatrix} \frac{(1-\beta)^{-1}}{(1-\pi)^3} \int\limits_{-\infty}^{+\infty} (1-\alpha-y_1)^2 \bar{\rho}(y_1)\, dy_1 & \frac{-(1-\beta)^{-1}}{(1-\pi)^2} \int\limits_{-\infty}^{+\infty} (1-\alpha-y_1) \bar{\rho}(y_1)\, dy_1 \\ \frac{-(1-\beta)^{-1}}{(1-\pi)^2} \int\limits_{-\infty}^{+\infty} (1-\alpha-y_1) \bar{\rho}(y_1)\, dy_1 & \frac{(1-\beta)^{-1}}{1-\pi} \int\limits_{-\infty}^{+\infty} \bar{\rho}(y_1)\, dy_1 \end{pmatrix}.$$

The first element in this matrix (first row, first column) is strictly positive and if we show that the determinant of the Hessian is strictly positive, then by [4, Thm.4.7], the Hessian matrix $D[\nabla(F_\beta)]_{(\pi,\alpha)}$ is positive definite. We know that there exist a point $(\breve{\pi}, \breve{\alpha})$ such that $\nabla F_\beta(\breve{\pi}, \breve{\alpha}) = 0$. For any point $(\pi, \alpha) \neq (\breve{\pi}, \breve{\alpha})$, and if $D[\nabla(F_\beta)]_{(\pi,\alpha)}$ is positive definite for all points, then, by a Taylor expansion (neglecting the terms of order 3 and more) we have

$$F_\beta(\pi, \alpha) - F_\beta(\pi_0, \alpha_0) \simeq \frac{1}{2} \underbrace{((\pi, \alpha) - (\pi_0, \alpha_0))^\top D[\nabla(F_\beta)]_{(\pi,\alpha)}((\pi, \alpha) - (\pi_0, \alpha_0))}_{>0}$$

Therefore, there is one and only one point that minimizes the function F_β.

Let us show that the determinant of the Hessian is strictly positive for all points. Let V denote the vector space of affine function of one variable, i.e.

$$V := \{\gamma : \mathbb{R} \longrightarrow \mathbb{R} \,|\gamma(y) = \eta + y, \text{ for a certain } \eta \in \mathbb{R}\}.$$

On V, we define a scalar product

$$\begin{array}{rccc} \langle\cdot|\cdot\rangle_{\bar{\rho}} : & V \times V & \longrightarrow & \mathbb{R} \\ & (\gamma_1, \gamma_2) & \longmapsto & \int\limits_{\mathbb{R}} \gamma_1(y)\gamma_2(y)\bar{\rho}(y)\, dy \end{array}$$

Then, functional analysis confirms that we have the well known Cauchy-Schwarz inequality

$$|\langle\gamma_1|\gamma_2\rangle_{\bar\rho}|^2 \leqslant \langle\gamma_1|\gamma_1\rangle_{\bar\rho}\langle\gamma_2|\gamma_2\rangle_{\bar\rho}$$

and

$$\|\gamma\|_{\bar\rho} := \sqrt{\langle\gamma|\gamma\rangle_{\bar\rho}} = \sqrt{\int\limits_{\mathbb{R}} \gamma(y)^2\bar\rho(y)\,dy}$$

is a norm.

We define $\gamma_1(y) := 1-\alpha-y$ and $\gamma_2(y) := 1$. Since γ_1 and γ_2 are not co-linear, we have the strict inequality

$$\Big(\int\limits_{\mathbb{R}} \gamma_1(y)\bar\rho(y)\,dy\Big)^2 < \int\limits_{\mathbb{R}} \gamma_1(y)^2\bar\rho(y)\,dy \int\limits_{\mathbb{R}} \bar\rho(y)\,dy.$$

Therefore, we have

$$\frac{(1-\beta)^{-1}}{(1-\pi)^4}\int\limits_{\mathbb{R}} \gamma_1(y)^2\bar\rho(y)\,dy \int\limits_{\mathbb{R}} \bar\rho(y)\,dy - \frac{(1-\beta)^{-1}}{(1-\pi)^4}\Big(\int\limits_{\mathbb{R}} \gamma_1(y)\bar\rho(y)\,dy\Big)^2 > 0,$$

which is the determinant of $D[\nabla(F_\beta)]_{(\pi,\alpha)}$, and so the Hessian is positive definite for all points.

I Uniqueness of the solution for Problem 4

Let $\Pi = \{\pi \in \mathbb{R}^N | \pi_i \geqslant 0, i = 1,\dots,N \text{ and} \sum\limits_{i=1}^{N}\pi_i = 1\} \subset [0,1]^N$. We define

$$\begin{array}{rccl} H_- : & \Pi & \longrightarrow & \mathbb{R} \\ & \pi & \longmapsto & \sum\limits_{i=1}^{N}\pi_i\log(\pi_i). \end{array}$$

For all i, we define the function

$$\begin{array}{rccl} \phi_i : & [0,1] & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & x\log(x). \end{array}$$

We know that, for all $i \in \{1,\dots,N\}$, ϕ_i is continuous and it is strictly convex (i.e. $\forall\, \lambda \in]0,1[,\ \phi(\lambda x+(1-\lambda)y) < \lambda\phi(x)+(1-\lambda)\phi(y)$), because of the following lemma.

Lemma 2. *Let $f : [a,b] \longrightarrow \mathbb{R}$ be a continuous function such that $f'(x)$ exists for all $x \in (a,b)$. If $f'(x)$ is monotonically increasing, then f is a convex function.*

Proof. Let $x, y \in [a,b]$ and define $k := \lambda x + (1-\lambda)y$ with $\lambda \in]0,1[$. By the Mean Value Theorem, we have

$$f(x) - f(k) = f'(\xi_1)(x-k) \text{ and } f(y) - f(k) = f'(\xi_2)(y-k),$$

and by hypothesis, we have $f'(\xi_1) \leqslant f'(\xi_2)$, which implies

$$\begin{aligned}
\frac{f(x)-f(k)}{x-k} &\leqslant \frac{f(y)-f(k)}{y-k} \\
\frac{f(x)-f(\lambda x+(1-\lambda)y)}{(1-\lambda)(x-y)} &\leqslant \frac{f(y)-f(\lambda x+(1-\lambda)y)}{\lambda(y-x)} \\
\frac{f(\lambda x+(1-\lambda)y)}{(y-x)} &\leqslant \frac{(1-\lambda)(f(y)-f(\lambda x+(1-\lambda)y))+\lambda f(x)}{\lambda(y-x)} \\
0 &\leqslant \frac{-f(\lambda x+(1-\lambda)y)+\lambda f(x)+(1-\lambda)f(y)}{\lambda} \\
f(\lambda x+(1-\lambda)y) &\leqslant \lambda f(x)+(1-\lambda)f(y),
\end{aligned}$$

and hence shows that f is convex. □

It is easy to see from the proof that if $f'(x)$ is strictly monotonically increasing, then f is strictly convex. In our case, $\phi_i'(x) = \log(x) + 1$ which is well defined on $(0,1)$ and is strictly increasing and therefore ϕ_i is strictly convex.

We can now define

$$\begin{aligned}
\Phi :\ [0,1]^N &\longrightarrow \mathbb{R} \\
x &\longmapsto \sum_{i=1}^{N} \phi_i(x).
\end{aligned}$$

We know that Φ is continuous and strictly convex (since it is the sum of continuous and strictly convex functions). If we restrict Φ to the convex set Π, then obviously it is still continuous and strictly convex, and because $H_- \equiv \Phi_{|\Pi}$, H_- is continuous and strictly convex on Π and in particular, on any closed, convex subset F of Π (where F is seen as the feasible set of Problem 4). Therefore, there exists a unique $\pi^* \in F$ such that $H_-(\pi^*) = \min_{\pi \in F}\{H_-(\pi)\}$. This is because if we suppose that $\exists\, \pi^*, \check{\pi}^* \in F$ such that $H(\pi^*) = \min\limits_{\pi \in F}\{H_-(\pi)\}$ and $H(\check{\pi}^*) = \min\limits_{\pi \in F}\{H_-(\pi)\}$, and because H_- is strictly convex, we have $H_-(\pi^*) = H_-(\check{\pi}^*)$ (refer to Appendix E) and

$$
\begin{aligned}
H_-(\lambda\pi^* + (1-\lambda)\check{\pi}^*) &< \lambda H_-(\pi^*) + (1-\lambda)H_-(\check{\pi}^*) \\
H_-(\lambda\pi^* + (1-\lambda)\check{\pi}^*) &< H_-(\check{\pi}^*).
\end{aligned}
$$

This is a contradiction to the fact that $\check{\pi}^*$ is a minimum of the function H_-.

References

[1] J.C.G. Boot, *Programmation Quadratique: Algorithmes, Anomalies, Applications*, Dunod, Paris, 1968.

[2] G.H. Choe, *Computational Ergodic Theory*, Algorithms and Computation in Mathematics, Vol.13, *Springer*, Berlin, 2005.

[3] J. Gapaillard, *Intgration pour la licence*, Masson, Paris, 1997.

[4] E. Hairer and G. Wanner, *L'analyse au fil de l'histoire*, Springer, Berlin, 2001.

[5] J. Jacod and P. Protter, *Probability Essentials*, Springer, Berlin, 2004.

[6] R. Korn and E. Korn, *Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics*, Graduate Studies in Mathematics, Vol. 3, *American Mathematical Society*, Providence, Rhode Island, 2001.

[7] H. Markowitz, *Portfolio Selection*, The Journal of Finance, Vol.7, Issue 1, 1952, pp 77-91.

[8] A.P. Prudnikov, Yu.A. Brychkov, O.I. Marichev, *Integrals and Series: Volume I, Elementary Functions*, Translated from the Russian by N.M. Queen, Gordon and Breach, New York, 1986.

[9] R.T. Rockafellar, *Convex Analysis.*, Princton Landmarks in Mathematics and Physics, Princeton University Press, Princeton, New Jersey, 1970.

[10] R.T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, Journal of Banking and Finance, Vol. 26, No. 7, 2002, pp 1443-1471.

[11] R.T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Journal of Risk, Vol. 2, No. 3, 2000, pp 21-41.

[12] W. Rudin, *Analyse relle et complexe*, Dunod, Paris, 1998.

[13] M. Zinsmeister, *Formalisme thermodynamique et systmes dynamiques holomorphes*, Panoramas et Synthses 4, *Socit Mathmatique de France*, Paris, 1996.