Chapter 1

Introduction

Chapter 2

Model Description

In this chapter, first the basic financial quantities are introduced and the asset allocation problem is stated, then the same problem will be embedded in a dynamical control system framework which will allows us to develop the stochastic reachability approach to portfolio construction. We closely follow [14],[13] and [12].

2.1 Portfolio construction

In the financial industry, a group of securities that exhibits similar characteristics in the market place and is subject to the same regulation is called **asset class**. Typical asset classes include stocks, bonds, real estate, cash and commodities. The discipline consisting in allocating investor's wealth among different asset classes is called **asset allocation**. We will now introduce the financial quantities and a formal mathematical setting suitable for describing the asset allocation problem. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space and consider a discrete set of time indexed by $k \in \mathbb{N}$. Moreover, let us consider a universe of $m \in \mathbb{N}$ asset classes. Asset classes' performance at period k is described by a m-dimensional random vector $\boldsymbol{w}_k = [w_k(1), \ldots, w_k(m)]^T$ where

$$w_k(i) = \frac{z_k(i) - z_{k-1}(i)}{z_{k-1}(i)}, \quad i = 1, \ldots, m$$

is the rate of return of the i-th asset class and $\{z_k(i)\}_{k \in \mathbb{N}}$ the i-th asset class price process. In general, the correlation of $\boldsymbol{w}_k$ can be of two kinds:

- *synchronous* correlation, that is the correlation among different asset class at the same time period

- *time-lagged* correlation, that is the correlation among different asset class at different time period.

As the time-lagged correlation is usually negligible for short time period, $\boldsymbol{w}_k$ is a synchronous-correlated random vector. Standard notation is used for Expected Returns and Covariance Matrix

$$\mu_k(i) = \mathbb{E}[w_k(i)], \quad i = 1, \dots, m \quad k \in \mathbb{N}$$

$$\Sigma_k(i,j) = \mathbb{E}\Big[\big(w_k(i) - \mu_k(i)\big)\big(w_k(j) - \mu_k(j)\big)\Big], \quad i,j = 1, \dots, m \quad k \in \mathbb{N}$$

An asset allocation at period $k \in \mathbb{N}$ is a vector $\boldsymbol{u}_k \in \mathbb{R}^m$ whose i-th element indicates the percentage of wealth to be invested in asset class i. This vector is the leverage the asset manager has at his disposal for driving the portfolio value towards his goals. Finally, portfolio performance over the period $[k-1, k]$ is measured by the portfolio return

$$r_{k+1} = \frac{x_{k+1} - x_k}{x_k}$$

where $\{x_k\}_{k\in\mathbb{N}}$ is the portfolio value process. The portfolio return can also be expressed as a weighted average of each asset class return as

$$r_{k+1} = \boldsymbol{u}_k^T \boldsymbol{w}_{k+1}$$

By combining the two previous relations we get the following recursive equation

$$\boxed{x_{k+1} = x_k(1 + \boldsymbol{u}_k^T \boldsymbol{w}_{k+1})} \tag{2.1}$$

which describes the time evolution of portfolio value. In plain words, the **asset allocation problem** consists in choosing the vector $\boldsymbol{u}_k$ at each time period $k \in \mathbb{N}$ so as to achieve investor's goal (i.e. beating a benchmark, maximizing the return and so on). As a matter of fact, the asset allocation vector $\boldsymbol{u}_k$ is bound to stay within a feasible set U_k, for each $k \in \mathbb{N}$. In this work, the feasible set U_k is obtained by imposing the following constraints

- *budget* constraint: $\sum_{i=1}^m u_k(i) = 1$, all the wealth is invested in the portfolio
- *long-only* constraint: $u_k(i) \geq 0, \quad i = 1, \dots, m$, no short-selling are allowed
- *risk* constraint: the metric value-at-risk ($V@R$) is used to limit portfolio risk.

2.2 Stochastic Reachability Approach

In the previous section the financial setting has been laid, now it will be embedded in a more general framework by employing the theory of dynamical systems. We will see that this formalism will allow us to formulate the asset allocation problem as a **stochastic reachability** problem which will be solved by using **dynamic programming** (DP) techniques. Let us introduce the following stochastic discrete-time dynamic control system

$$x_{k+1} = f(x_k, \boldsymbol{u}_k, \boldsymbol{w}_{k+1}) = x_k(1 + \boldsymbol{u}_k^T \boldsymbol{w}_{k+1})$$

where, for any $k \in \mathbb{N}$:

- $x_k \in \mathcal{X} = \mathbb{R}$ is the system state (portfolio value), $\mathcal{X}$ the system space
- $\boldsymbol{u}_k \in U \subset \mathbb{R}^m$ is the control input (asset allocation vector), U the control input space
- $\boldsymbol{w}_k$ is a m-dimensional random vector (asset class returns) with density function $p_{\boldsymbol{w}_k}$

Let $\mathcal{U} = \{\mu : \mathcal{X} \times \mathbb{N} \to U\}$ be the class of controls. In other words, $\mu \in \mathcal{U}$ is a map such that for any $x \in \mathcal{X}$ and any $k \in \mathbb{N}$, associates an asset allocation vector $\boldsymbol{u}_k \in U$. Given $N \in \mathbb{N}$ we define the set of control sequences as $\mathcal{U}_N = \{\pi = \{\mu_k\}_{k=0,\dots,N} : \mu_k \in \mathcal{U}\}$ and call any $\pi \in \mathcal{U}_N$ a **control policy**. Denote by π^k a control policy starting at period k, that is $\pi^k = \{\mu_k, \dots, \mu_N\}$. We are now ready to formulate the asset allocation problem in stochastic reachability terms

***Problem* 2.2.1** (Optimal Dynamic Asset Allocation)**:** Given a finite time horizon $N \in \mathbb{N}$ and a sequence of target sets $\{X_1, \dots, X_N\}$ such that each target set is a subset of the state space $\mathcal{X}$, find the optimal control policy $\pi^\star \in \mathcal{U}_{N-1}$ that maximizes the following objective function

$$\mathbb{P}\big(\{\omega \in \Omega : x_0 \in X_0, \dots, x_N \in X_N\}\big) \tag{2.2}$$

The target sets $\{X_1, \dots, X_N\}$ represent the investor's goal and we can think of them as the "good" states where we want the portfolio value to belongs to. For instance, a target set could be $X_k = [\underline{x}_k, \infty)$. Problem (2.2.1) is solved by resorting to dynamic programming. To this end, let $p_{f(x,\boldsymbol{u},\boldsymbol{w}_{k+1})}$ be the probability density function of random variable $f(x_k, \boldsymbol{u}_k, \boldsymbol{w}_{k+1}) = x_k(1 + \boldsymbol{u}_k^T \boldsymbol{w}_{k+1})$.

***Definition* 2.2.1** (value function)**:** Given a sequence of target sets $\{X_k\}_{k\geq 0}$, the **value function** is the following real map

$$\begin{aligned} V\colon \mathbb{N}\times\mathcal{X}\times\mathcal{U} &\to [0,1] \\ (k,x,\pi^k) &\mapsto V(k,x,\pi^k) \end{aligned}$$

such that

$$V(k,x,\pi^k)=\begin{cases}\mathbb{1}_{X_N}(x) & \text{if} \quad k=N \\ \int_{X_{k+1}} V(k+1,z,\pi^{k+1})p_{f(x,\boldsymbol{u},\boldsymbol{w}_{k+1})}(z)\mathrm{d}z & \text{if} \quad k=N-1,\dots,0\end{cases}$$

It is now possible to link the objective function (2.2) to the value function (see [12]) in the following way

$$\mathbb{P}\big(\{\omega\in\Omega : x_0\in X_0,\dots,x_N\in X_N\}\big)=V(0,x_0,\pi).$$

This result is extremely important since it allows us to rewrite the ODAA problem in terms of the value function as follows

***Problem* 2.2.2** (Optimal Dynamic Asset Allocation 2)**:** Given a finite time horizon $N\in\mathbb{N}$ and a sequence of target sets $\{X_1,\dots,X_N\}$, find

$$\pi^\star=\underset{\pi\in\mathcal{U}_{N-1}}{\arg\max}\, V(0,x_0,\pi).$$

So far, we have reached an intermediate point where the ODAA problem has been restated in terms of a value function V. The tool of dynamic programming gives us the final step, which is synthesized in the following theorem

***Theorem* 2.2.1** (ODAA algorithm)**:** the optimal value of the ODAA problem (2.2.2) is

$$p^\star=J_0(x_0),$$

where for any $x\in\mathcal{X}$, $J_0(x)$ is the final step of the following algorithm

$$\boxed{\begin{aligned} &J_N(x)=\mathbb{1}_{X_N}(x) \\ &J_k(x)=\sup_{\boldsymbol{u}_k\in U_k}\int_{X_{k+1}} J_{k+1}(z)p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}(z)\mathrm{d}z \\ &\qquad k=N-1,\dots,1,0 \end{aligned}} \tag{2.3}$$

The previous result provides us with a backward procedure (it starts at time N and ends at time 0) whose outputs are the optimal control policy $\pi^\star=\{\mu_0^\star,\dots,\mu_{N-1}^\star\}$ and the optimal probability $p^\star$ of reaching the target sets. It is worth pointing out some interesting features of algorithm in (2.3):

- $J_k(x)$ is a function of portfolio realization $x \in \mathcal{X}$ at time k. This dependence is hidden behind the probability density function $p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}$
- the constrained optimization must be numerically carried out in a space (U_k) of dimension $m \in \mathbb{N}$. At each iteration $k = N-1, \ldots, 1, 0$, the optimization has to be repeated for each x belonging to a suitable discretized grid
- the algorithm presented in theorem (2.2.1) does not depend on the particular distribution of random variable $f(x, \boldsymbol{u}_k, \boldsymbol{w}_{k+1})$. This fact gives us enough freedom to look outside the usual Guassian world
- given a period $k \in \mathbb{N}$ and a portfolio value realization $x \in \mathcal{X}$, $\mu_k^\star(x) \in U_k$ tells us which is the optimal allocation mix of our portfolio.

We now ask ourselves which probability distributions are suitable for vector $\boldsymbol{w}_{k+1}$; the answer to this question is the main objective of the next chapter.

Chapter 3

Asset Class Returns modeling

In this chapter, we address the asset class returns modeling issue. As it was noted above, the ODAA algorithm does not depend on a particular distribution for the asset class returns vector $\boldsymbol{w}_{k+1}$. However, by looking at equation (2.3) we see that we need the explicit analytical form for the density function $p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}$. For this reason, we will be dealing exclusively with probability distributions closed under linear combination. In this work, we propose three such distributions:

- Gaussian
- Gaussian Mixture
- Generelized Hyperbolic

For each of them, first we will give a theoretical introduction, secondly derive the portfolio value density function $p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}$ and an expression for the risk constraint (which depends on the distribution chosen).

3.1 Gaussian model

The first probability distribution we considered is the Gaussian

***Definition* 3.1.1** (Gaussian random vector)**:** A m-dimensional random vector $\boldsymbol{w} = [w_1, \ldots, w_m]^T$ is **Gaussian** if every linear combination $\sum_{i=0}^{m} u_i w_i = \boldsymbol{u}^T \boldsymbol{w}$ has a one-dimensional Gaussian distribution.

Let the asset class returns random vector $\boldsymbol{w}_{k+1}$ follow a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ $\big(\boldsymbol{w}_{k+1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\big)$. By

definition we have

$$x(1+\boldsymbol{u}_k^T\boldsymbol{w}_{k+1}) \sim \mathcal{N}\Big(\underbrace{x(1+\boldsymbol{u}_k^T\boldsymbol{\mu})}_{\tilde{\mu}}, \underbrace{x^2\boldsymbol{u}_k^T\Sigma\boldsymbol{u}_k}_{\tilde{\sigma}^2}\Big)$$

hence

$$\boxed{p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}(z) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\big\{-\frac{1}{2}\frac{(z-\tilde{\mu})^2}{\tilde{\sigma}^2}\big\}, \quad z \in \mathbb{R}.} \tag{3.1}$$

Let us now introduce the two important concepts of loss function and value-at-risk that we will use to derive the portfolio risk constraint

***Definition* 3.1.2** (loss function, $V@R$)**:** Denoting the value of our portfolio at time $k \in \mathbb{N}$ by x_{k+1}, the **loss function** of the portfolio over the period $[k, k+1]$ is given by

$$L_{k+1} := -\frac{(x_{k+1}-x_k)}{x_k} = -r_{k+1} = -\boldsymbol{u}_k^T\boldsymbol{w}_{k+1}.$$

Given some confidence level $1-\alpha \in (0,1)$ the **value-at-risk** ($V@R_{1-\alpha}$) of our portfolio is

$$V@R_{1-\alpha} = \inf\{l \in \mathbb{R} : \mathbb{P}\big(L_{k+1} \leq l\big) \geq 1-\alpha\}.$$

The V@R is a risk measure commonly use by financial institutions to asses the risk they run to carry a portfolio for a specified period of time (the portfolio must be kept constant during this time period). For instance, if our portfolio has a $V@R_{0.99} = 7\%$, this means that with a confidence level of 99% our portfolio does not suffer a loss greater or equal than 7% over per period $[k, k+1]$ (e.g. a month). In our case, we receive the V@R specification ($V@R_{0.99} = 7\%$) in input by the investor (it is an indicator of its risk-aversion) and we will construct an asset allocation $\boldsymbol{u}_k$ that satisfies this risk constraint at each $k \in \mathbb{N}$.

Using definition (3.1.1) we have

$$L_{k+1} \sim \mathcal{N}\big(\underbrace{-\boldsymbol{u}_k^T\boldsymbol{\mu}}_{\mu_p}, \underbrace{\boldsymbol{u}_k^T\boldsymbol{\Sigma}\boldsymbol{u}_k}_{\sigma_p^2}\big)$$

therefore

$$\begin{aligned}\mathbb{P}\big(L_{k+1} \leq V@R_{1-\alpha}\big) &= \mathbb{P}\Big(\boldsymbol{Z} \leq \frac{V@R_{1-\alpha}-\mu_p}{\sigma_p}\Big)\\ &= 1-\alpha\end{aligned}$$

$$\implies \boxed{V@R_{1-\alpha} \geq -\boldsymbol{u}_k^T \boldsymbol{\mu} + z_{1-\alpha}\sqrt{\boldsymbol{u}_k^T \boldsymbol{\Sigma} \boldsymbol{u}_k}} \tag{3.2}$$

where $\boldsymbol{Z}$ is a standard normal random variable and $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. The *risk* constraint in equation (3.2), together with the *budget* and *long-only* constraint, define the control space U_k which is the feasible set of the constrained optimization problem given in theorem (2.2.1).

3.2 Gaussian Mixture model

In this section we present the second asset class returns model, the **Gaussian Mixture model** (GM). After introducing the GM distribution we will derive the density and the *risk* constraint, as we did for the Gaussian model. We closely follow [2].

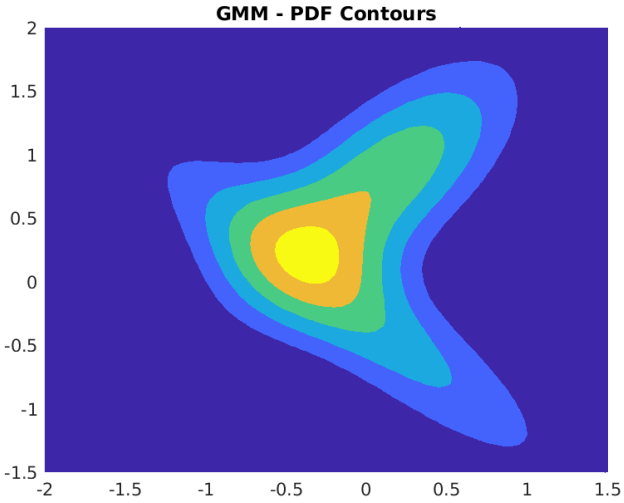
The standard assumption that asset returns have a multivariate Gaussian distribution is a reasonable first approximation to reality and it usually has the big advantage of generating analytically tractable theories (e.g. Markowitz portfolio theory). However, the Gaussian model does not capture two important key features of asset returns which are observed in the market real data:

1. the skewed (asymmetric around the mean) and leptokurtic (more fat-tailed than the Gaussian) nature of marginal probability density function

2. the asymmetric correlation between asset returns, that is the tendency of volatilities and correlations to depend on the prevailing market conditions.

To overcome this shortcomings, the Gaussian Mixture distribution is a validate alternative to the Gaussian one. Loosely speaking, the pdf of a GM random vector id a linear combination of Gaussian pdfs (called Gaussian regimes or mixing components). This closeness to the Normal distribution offers a good trade-off between analytical tractability and parsimony in the number of parameters. By adopting a GM model, it is possible to represent protuberances on the probability iso-density contour, as can be seen in the following figure

Figure 3.1: Example of a GM density contour plot with two mixing components.



To obtain this highly non-linear dependence structure, we would usually need cross-moments of all order; a big advantage of the GM distribution is that its dependence structure is fully and conveniently captured by the means, covariance matrices and weights of each Gaussian regime (as we will see in the following).

We start the more formal introduction on the GM distribution giving its definition

***Definition* 3.2.1** (GM distribution)**:** An m-dimensional random vector $\boldsymbol{Z}$ has a **multivariate GM distribution** if its probability density function is of the form

$$p_{\boldsymbol{Z}}(\boldsymbol{z}) = \sum_{i=1}^{n} \lambda_i \varphi_{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}(\boldsymbol{z}), \quad \boldsymbol{z} \in \mathbb{R}^m,$$

where $\varphi_{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}$ is the multivariate Gaussian density with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$ and λ_i are positive mixing weights summing to one.

The following proposition is crucial for our purposes since it tells us that linear combinations of GM random vector have a one-dimensional GM distribution

***Proposition* 3.2.1:** Linear combinations of GM random vectors follow a univariate GM distribution. In particular, if $\boldsymbol{Z} \sim GM$ then $Y = \boldsymbol{\theta}^T \boldsymbol{Z}$,

$\forall \boldsymbol{\theta} \in \mathbb{R}^m$, has a GM distribution with probability density function

$$p_Y(y) = \sum_{i=1}^{n} \lambda_i \varphi_{(\mu_i, \sigma_i^2)}(y), \quad y \in \mathbb{R}$$

where

$$\begin{cases} \mu_i & = \boldsymbol{\theta}^T \boldsymbol{\mu}_i \quad i = 1, \ldots, m \\ \sigma_i^2 & = \boldsymbol{\theta}^T \boldsymbol{\Sigma}_i \boldsymbol{\theta} \quad i = 1, \ldots, m \end{cases}$$

Proof. The characteristic function (CF) of a GM random vector is the linear combination of the CF of the Gaussian mixing components. Indeed

$$\begin{aligned} \phi_{\boldsymbol{Z}}(\boldsymbol{u}) = \mathbb{E}[\exp\{i\boldsymbol{u}^T \boldsymbol{Z}\}] &= \int_{\mathbb{R}^m} \exp\{i\boldsymbol{u}^T \boldsymbol{z}\} p_{\boldsymbol{Z}}(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} = \\ &= \int_{\mathbb{R}^m} \exp\{i\boldsymbol{u}^T \boldsymbol{z}\} \sum_{i=1}^{n} \lambda_i \varphi_{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} = \\ &= \sum_{i=1}^{n} \lambda_i \phi_{\boldsymbol{X}_i}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^m \end{aligned}$$

where $\boldsymbol{X}_i \sim \mathcal{N}\big(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\big)$. Therefore, $\forall \boldsymbol{\theta} \in \mathbb{R}^m$ we have

$$\begin{aligned} \phi_{\boldsymbol{\theta}^T \boldsymbol{Z}}(u) = \mathbb{E}[\exp\{iu(\boldsymbol{\theta}^T \boldsymbol{Z})\}] = \mathbb{E}[\exp\{i(u\boldsymbol{\theta}^T)\boldsymbol{Z}\}] &= \phi_{\boldsymbol{Z}}(u\boldsymbol{\theta}) = \\ = \sum_{i=1}^{n} \lambda_i \phi_{X_i}(u\boldsymbol{\theta}) = \sum_{i=1}^{n} \lambda_i \exp\{iu \underbrace{\boldsymbol{\theta}^T \boldsymbol{\mu}_i}_{\mu_i} - \frac{1}{2} u^2 \underbrace{\boldsymbol{\theta}^T \boldsymbol{\Sigma}_i \boldsymbol{\theta}}_{\sigma_i^2}\} &= \\ = \sum_{i=1}^{n} \lambda_i \phi_{\widetilde{X}_i}(u), \quad u \in \mathbb{R} \end{aligned}$$

where $\widetilde{X}_i \sim \mathcal{N}\big(\mu_i, \sigma_i^2\big)$. Since the CF completely characterizes the distribution (see theorem 14.1 in [7]) we have the result. □

Suppose the asset class returns vector[1] $\boldsymbol{w}$ follow a Gaussian Mixture distribution ($\boldsymbol{w} \sim GM$). We want to compute the density of random variable $f(x, \boldsymbol{u}, \boldsymbol{w}) = x(1 + \boldsymbol{u}^T \boldsymbol{w})$. Thanks to proposition (3.2.1) we know that the random variable $\boldsymbol{u}^T \boldsymbol{w}$ follows itself a GM (univariate) distribution. Moreover, by integration we easily derive its cumulative distribution function (CDF) allowing us to write

$$\begin{aligned} F_{f(x,\boldsymbol{u},\boldsymbol{w})}(z) = \mathbb{P}\big(x(1 + \boldsymbol{u}^T \boldsymbol{w}) \leq z\big) &= F_{x\boldsymbol{u}^T \boldsymbol{w}}(z - x) \\ &= \sum_{i=1}^{n} \lambda_i \Phi\Big(\frac{z - x(1 + \boldsymbol{u}^T \boldsymbol{\mu}_i)}{\sqrt{x^2 \boldsymbol{u}^T \boldsymbol{\Sigma}_i \boldsymbol{u}}}\Big), \quad z \in \mathbb{R} \end{aligned}$$

[1] for the sake of clarity we drop the subscript $k + 1$ when it is not needed

where Φ is the standard normal CDF. Differentiating with respect to z, we have

$$\boxed{p_{f(x,\boldsymbol{u},\boldsymbol{w})}(z) = \sum_{i=1}^{n} \lambda_i \varphi_{(\mu_i,\sigma_i^2)}(z), \quad z \in \mathbb{R}} \tag{3.3}$$

where

$$\begin{cases} \mu_i &= x(1 + \boldsymbol{u}^T \boldsymbol{\mu}_i) \\ \sigma_i^2 &= x^2 \boldsymbol{u}^T \boldsymbol{\Sigma}_i \boldsymbol{u}. \end{cases}$$

We now turn to the problem of computing the *risk* constraint under the GM distribution assumption. We will follow two different approaches. Suppose we are given the $V@R_{1-\alpha}$ specification (e.g. 7%); by using definition (3.1.2) we have

$$\mathbb{P}\big(L \leq V@R_{1-\alpha}\big) = F_L(V@R_{1-\alpha}) \geq 1 - \alpha$$

as noted above, the CDF of $L = -\boldsymbol{u}^T \boldsymbol{w}$ is known, therefore

$$\sum_{i=1}^{n} \lambda_i \Phi\Big(\frac{V@R_{1-\alpha} - \mu_i}{\sigma_i}\Big) \geq 1 - \alpha \quad \Longrightarrow$$

$$\boxed{\sum_{i=1}^{n} \lambda_i \Phi\Big(-\Big\{\frac{V@R_{1-\alpha} - \mu_i}{\sigma_i}\Big\}\Big) \leq \alpha} \tag{3.4}$$

where

$$\begin{cases} \mu_i &= -\boldsymbol{u}^T \boldsymbol{\mu}_i \\ \sigma_i^2 &= \boldsymbol{u}^T \boldsymbol{\Sigma}_i \boldsymbol{u}. \end{cases}$$

We present also an alternative method to limit the risk exposure of our portfolio which turns out to be less computationally intensive. The idea is to set an upper bound to portfolio return volatility in the following way

$$(\mathrm{Var}\,[r_{k+1}])^{\frac{1}{2}} = (\boldsymbol{u}_k^T \boldsymbol{\Lambda} \boldsymbol{u}_k)^{\frac{1}{2}} \leq \sigma_{max} \tag{3.5}$$

where $\boldsymbol{\Lambda}$ is the covariance matrix of vector $\boldsymbol{w}_{k+1}$. Two questions are left open: how to compute $\boldsymbol{\Lambda}$ and how to link the upper bound σ_{max} to the $V@R_{1-\alpha}$ specification given in input by the investor. As far as the former is concerned, the following proposition gives us the answer

***Proposition* 3.2.2:** The covariance matrix of a random vector with the GM distribution can be expressed in terms of mean vectors, covariance matrices and weights of the mixing components in the following way

$$\boldsymbol{\Lambda} = \sum_{i=1}^{n} \lambda_i \boldsymbol{\Sigma}_i + \sum_{i=1, j<i}^{n,n} \lambda_i \lambda_j (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T.$$

To answer the latter, we use a Guassian approximation and the fact that if the rebalancing frequency is relatively small (e.g. weekly) the portfolio return mean is negligible. In the end, the obtain

$$\sigma_{max} = \frac{V@R_{1-\alpha}}{z_{1-\alpha}}.$$

3.3 Generelized Hyperbolic model

The last distribution we propose for our asset class returns modeling purposes is the Generalized Hyperbolic (GH). Like the GM, in its general form also the GH presents a non-elliptical behaviour with asymmetric and fat-tailed marginals. We proceed to give the formal definition and then derive the density of $f(x, \boldsymbol{u}_k, \boldsymbol{w}_{k+1})$ and the *risk* constraint.

***Definition* 3.3.1** (GH distribution)**:** A m-dimensional random vector $\boldsymbol{X}$ is said to follow a **multivariate GH distribution** ($\boldsymbol{X} \sim GM_m(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$) if

$$\boldsymbol{X} = \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}A\boldsymbol{Z}$$

where

- $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, I_d)$
- $A \in \mathbb{R}^{m\times d}$ is the Cholesky factor of dispersion matrix $\boldsymbol{\Sigma}$ ($A^T A = \boldsymbol{\Sigma}$)
- $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^m$
- $W \sim \mathcal{N}^-(\lambda, \chi, \psi)$, $W \geq 0$ and $W \perp \boldsymbol{Z}$ (see Appendix B for the definition of the GIG distribution). W is sometimes called mixing random variable.

Remark 3.3.1

- *λ, χ, ψ are shape parameters; the larger these parameters the closer the distribution is to the Gaussian*
- *$\boldsymbol{\gamma}$ is the skewness parameter. If $\boldsymbol{\gamma} = \boldsymbol{0}$ the distribution is symmetric around the mean*
- $\boldsymbol{X}|W = w \sim \mathcal{N}(\boldsymbol{\mu} + w\boldsymbol{\gamma}, w\boldsymbol{\Sigma})$

The GH distribution contains some special cases:

- If $\lambda = \frac{m+1}{2}$ we have a *Hyperbolic* distribution
- If $\lambda = -\frac{1}{2}$ the distribution is called *Normal Inverse Gaussian* (NIG)

- If $\chi = 0$ and $\lambda > 0$ we have the limiting case of the *Variance Gamma* (VG) distribution
- If $\psi = 0$ and $\lambda < 0$ the resulting distribution is called *Student-t*.

The following proposition gives us the closeness under linear transformation that we need for our modeling purposes

***Proposition* 3.3.1:** If $\boldsymbol{X} \sim GH_m(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ and $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$, where $B \in \mathbb{R}^{d\times m}$ and $\boldsymbol{b} \in \mathbb{R}^d$, then

$$\boldsymbol{Y} \sim GH_d(\lambda, \chi, \psi, B\boldsymbol{\mu} + b, B\boldsymbol{\Sigma}B^T, B\boldsymbol{\gamma}).$$

Suppose $\boldsymbol{w}_{k+1} \sim GH_m(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$. Applying the previous result to our case $\big(Y = f(x, \boldsymbol{u}_k, \boldsymbol{w}_{k+1}), \quad B = x\boldsymbol{u}_k^T, \quad b = x\big)$ we have

$$f(x, \boldsymbol{u}_k, \boldsymbol{w}_{k+1}) \sim GH_1(\lambda, \chi, \psi, \underbrace{x(1 + \boldsymbol{u}_k^T\boldsymbol{\mu})}_{\widetilde{\mu}}, \underbrace{x^2\boldsymbol{u}_k^T\boldsymbol{\Sigma}\boldsymbol{u}_k}_{\widetilde{\Sigma}}, \underbrace{x\boldsymbol{u}_k^T\boldsymbol{\gamma}}_{\widetilde{\gamma}})$$

and the density reads as (see Appendix B)

$$\boxed{p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}(z) = c\frac{K_{\lambda-\frac{1}{2}}\Big(\sqrt{\big(\chi + \widetilde{Q}(z)\big)\big(\psi + \widetilde{\gamma}^2/\widetilde{\Sigma}\big)}\exp\big\{(z-\widetilde{\mu})\widetilde{\gamma}/\widetilde{\Sigma}\big\}\Big)}{\Big(\sqrt{\big(\chi + \widetilde{Q}(z)\big)\big(\psi + \widetilde{\gamma}^2/\widetilde{\Sigma}\big)}\Big)^{\frac{1}{2}-\lambda}}} \tag{3.6}$$

where

$$c = \frac{\big(\sqrt{\chi\psi}\big)^{-\lambda}\psi^{\lambda}\big(\psi + \widetilde{\gamma}^2/\widetilde{\Sigma}\big)^{\frac{1}{2}-\lambda}}{(2\pi\widetilde{\Sigma})^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})}$$

and $\widetilde{Q}(z) = (z - \widetilde{\mu})/\widetilde{\Sigma}$.

As far as the *risk* constraint is concerned, we adopt here the alternative approach expressed in Equation (3.5). The covariance matrix $\boldsymbol{\Lambda}$ is easily derived from the definition of a GH random vector (Definition (3.3.1)) and Equation (B.2) in Appendix B; in the end we obtain

$$\boldsymbol{\Lambda} = \mathrm{Var}\,[W]\,\boldsymbol{\gamma}\boldsymbol{\gamma}^T + \mathbb{E}[W]\boldsymbol{\Sigma} \tag{3.7}$$

where

$$\mathbb{E}[W] = \Big(\frac{\chi}{\psi}\Big)^{\frac{1}{2}}\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}$$

$$\mathrm{Var}\,[W] = \Big(\frac{\chi}{\psi}\Big)\frac{1}{K_{\lambda}(\sqrt{\chi\psi})}\Big\{K_{\lambda+2}(\sqrt{\chi\psi}) - \frac{K_{\lambda+1}^2(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}\Big\}.$$

Chapter 4

Model Calibration

In this chapter we show how to calibrate the models introduced in Chapter 3 to market data. The asset class menu we will consider consists in equity, bond and cash and a suitable index will be used to represent each of these markets (the dataset is discussed in Section **??**). In this work, we set the number of mixing Gaussian components to 2. In financial terms, the two mixing components could be interpreted as economic regimes, namely a *tranquil* regime and a distressed one (see [1]). We focus our attention only on GM and GH since calibrating the Gaussian model is trivial (it amounts to compute the sample mean and covariance matrix). As far as the GM model is concerned, different calibration methods are available, namely the **Method of Moments** (MM), **Maximum Likelihood** (ML) estimation and the **Expectation-Maximization** (EM) algorithm. Each of them will be discussed in Section 4.1 and also a comparison between the three will be provided. Finally, in Section 4.2 the GH model will be fitted to data using the multi-cycle expectation conditional estimation (MCECM) algorithm.

4.1 GM calibration

The problem of estimating the parameters of a Gaussian Mixture distribution dates back to [10] and still nowadays it raises in a wide spectrum of different disciplines (Finance and Classification just to name a few). Thanks to the computational power available today, the EM algorithm is considered to be the state-of-the-art method for fitting the GM distribution. Nevertheless, MM and ML are worth studying as they shed light on different aspect of the problem at hand and they could provide the starting point for the EM algorithm. The main reference for the MM method is [5], for the ML [3] and for EM [8].

4.1.1 Method of Moments

In this subsection we present the Method of Moments for calibrating a 3-dimensional Gaussian Mixture distribution with $n = 2$ mixing component. The idea behind MM is to match observed and theoretical moments; this translates into a system of polynomial equations that most of the times, for real-size problems, has to be solved numerically. Since we need to fit a 3-dimensional distribution, we will work component-wise: moment-matching equations will be written for each component together with unimodality on each marginal. In order to keep the number of parameters to a reasonable degree, we will suppose a common correlation matrix between the two Gaussian mixing components.

Let $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ be a random sample from a GM distribution whose density function is

$$f(\boldsymbol{z}) = \lambda\varphi_{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)}(\boldsymbol{z}) + (1-\lambda)\varphi_{(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)}(\boldsymbol{z}), \quad \boldsymbol{z} \in \mathbb{R}^3 \tag{4.1}$$

Our goal is to estimate $\{\lambda, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\}$ from the random sample. Due to the assumption of a shared correlation matrix, the number of actual parameters to estimate is 16: λ, 6 means, 6 standard deviations and 3 correlations. To set the notation we give the following definition

***Definition* 4.1.1** (theoretical and sample moments)**:** Let X be a random variable and $\{x_1, \ldots, x_n\}$ a realization of a random sample. The first four theoretical and sample moments are:

$$\mu_X = \mathbb{E}[X] \qquad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

$$\sigma_X^2 = \mathbb{E}\big[(X-\mu_X)^2\big] \qquad s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

$$\gamma_X = \frac{1}{\sigma_X^3}\mathbb{E}\big[(X-\mu_X)^3\big] \qquad \widehat{\gamma} = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^2}\right)^3}$$

$$\kappa_X = \frac{1}{\sigma_X^4}\mathbb{E}\big[(X-\mu_X)^4\big] \qquad \widehat{\kappa} = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^4}{s^4}$$

Let $\boldsymbol{X}$ be a random vector with density (4.1), its i-th marginal is

$$f_{X_i}(z) = \varphi_{(\mu_{1i}, \sigma_{1i}^2)}(z) + (1-\lambda)\varphi_{(\mu_{2i}, \sigma_{2i}^2)}(z), \quad z \in \mathbb{R} \quad i \in \{1,2,3\}$$

where μ_{ji} and σ_{ji}^2 denote respectively the j-th element of the i-th mixing component mean vector and the j-th diagonal entry of the i-th mixing component covariance matrix, $i \in \{1,2,3\}$, $j \in \{1,2\}$ (namely the first subscripts

indicates the dimension, the second the mixing component). Computing explicitly the theoretical moments we obtain

$$\mu_{X_i} = \lambda\mu_{1i} + (1-\lambda)\mu_{2i}$$

$$\sigma^2_{X_i} = \lambda(\sigma^2_{1i} + \mu^2_{1i}) + (1-\lambda)(\sigma^2_{2i} + \mu^2_{2i})$$

$$\gamma_{X_i} = \frac{1}{\sigma^3_{X_i}}\Big\{\big[\lambda(\mu^3_{1i} + 3\mu_{1i}\sigma^2_{1i}) + (1-\lambda)(\mu^3_{2i} + 3\mu_{2i}\sigma^2_{2i})\big] - 3\mu_{X_i}\sigma^2_{X_i} - \mu^3_{X_i}\Big\}$$

$$\begin{aligned}\kappa_{X_i} = \frac{1}{\sigma^4_{X_i}}\Big\{&\big[\lambda(\mu^4_{1i} + 6\mu^2_{1i}\sigma^2_{1i} + 3\sigma^4_{1i}) + (1-\lambda)(\mu^4_{2i} + 6\mu^2_{2i}\sigma^2_{2i} + 3\sigma^4_{2i})\big] + \\ &- \mu^4_{X_i} - 6\mu^2_{X_i}\sigma^2_{X_i} - 4\gamma_{X_i}\sigma^3_{X_i}\mu_{X_i}\Big\}\end{aligned}$$

where $i \in \{1,2,3\}$. Equating them with their sample counterparts gives us the first twelve moment equations. The three correlation equations are derived equating the theoretical covariances (written as a function of correlation coefficients ρ_{ij})

$$\sigma_{X_iX_j} = \lambda\rho_{ij}\sigma_{1i}\sigma_{1j} + (1-\lambda)\rho_{ij}\sigma_{2i}\sigma_{2j} + \lambda(1-\lambda)(\mu_{1i} - \mu_{2i})(\mu_{1j} - \mu_{2j})$$

and the sample ones

$$\widehat{\sigma}_{X_iX_j} = \frac{1}{n}\sum_{s=1,t=1}^{n,n}(x_s - \bar{x})(x_t - \bar{x})$$

$i \in \{1,2,3\} \quad j < i$. So far, we have derived 15 equations in 16 unknown parameters. In order to have as many equations as unknown parameters, we solve the moment equation system by numerically minimizing the sum of square differences between theoretical and sample moments for different values of λ in a discretized grid of the interval $[0,1]$. The optimal λ will the one giving the smallest residual. Moreover, in the optimization process we also imposed the following uni-modality constraints on each marginal[1]

$$(\mu_{2i} - \mu_{1i})^2 \leq \frac{27}{4}(\sigma^2_{2i}\sigma^2_{1i})/(\sigma^2_{1i} + \sigma^2_{2i}) \quad i \in \{1,2,3\}$$

and positive-definiteness constraints on the standard deviation and correlation parameters. The uni-modality constraint is required since bi-modal return distributions are not observed in the market.

[1] see [4] for the proof of this sufficient condition for uni-modality for a 2-mixing-component GM density

4.1.2 Expectation-Maximization

In this section we introduce the EM algorithm for calibrating a GM model. Before diving into it, we need to define the maximum-likelihood estimator since the EM algorithm comes into play to solve difficulties in the ML method.

***Definition* 4.1.2** (Likelihood function)**:** Let $\boldsymbol{x} = \{x_1, \ldots, x_N\}$ be a realization of a random sample from a population with pdf $f(x|\boldsymbol{\theta})$ parametrized by $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_k]^T$. The **likelihood function** is defined by

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = L(\theta_1, \ldots, \theta_N|x_1, \ldots, x_k) = \prod_{i=1}^{N} f(x_i|\boldsymbol{\theta}).$$

The following definition of a maximum likelihood estimator is taken from [3]

***Definition* 4.1.3** (maximum-likelihood estimator)**:** For each sample point $\boldsymbol{x}$, let $\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$ be the parameters value at which $L(\boldsymbol{\theta}|\boldsymbol{x})$ attains its maximum as a function of $\boldsymbol{\theta}$, with $\boldsymbol{x}$ held fixed. A **maximum-likelihood estimator** (MLE) of the parameters vector $\boldsymbol{\theta}$ based on a random sample $\boldsymbol{X}$ is $\widehat{\boldsymbol{\theta}}(\boldsymbol{X})$

Intuitively, the MLE is a reasonable estimator since is the parameter point for which the observed sample is most likely. However, its main drawback is that finding the maximum of the likelihood function (or its logarithmic transformation) might be difficult both analytically and numerically. Consequently, the idea is to adopt an iterative procedure that converges to a local maximum. In order to focus on the idea behind the EM algorithm and not on technical details, we will present it in the simpler case of a uni-variate GM distribution with 2 mixing components (as presented in [6]). The interested reader can refer to [6] for the general case or [11] for a more throughout discussion.

Consider a mixture of two Gaussian random variables

$$X = (1 - \Delta)X_1 + \Delta X_2$$

where $X_1 \sim \mathcal{N}\big(\mu_1, \sigma_1^2\big)$, $X_2 \sim \mathcal{N}\big(\mu_2, \sigma_2^2\big)$ and $\Delta \sim B(\lambda)$ is the mixing random variable. The density function of X, parametrized by $\boldsymbol{\theta} = [\lambda, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2]^T$, is

$$f_X(x) = (1 - \lambda)\varphi_{(\mu_1, \sigma_1^2)}(x) + \lambda\varphi_{(\mu_2, \sigma_2^2)}(x), \quad x \in \mathbb{R}.$$

Our objective is to find an estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Let $\boldsymbol{x} = \{x_1, \ldots, x_N\}$ be a realization of a random sample (our data at hand), the log-likelihood function

is

$$l(\boldsymbol{\theta};\boldsymbol{x}) = \sum_{i=1}^{N} \log\left[(1-\lambda)\varphi_{(\mu_1,\sigma_1^2)}(x_i) + \lambda\varphi_{(\mu_2,\sigma_2^2)}(x_i)\right] \tag{4.2}$$

In higher dimensions, the direct maximization of (4.2) is difficult and prevent the ML method from being successful. Let us suppose to know the following latent random variables

$$\Delta_i = \begin{cases} 1 & \text{if } X_i \text{ comes from model 2} \\ 0 & \text{if } X_i \text{ comes from model 1} \end{cases}$$

for $i = 1, \ldots, N$. Model 1 or 2 is intended the population whose density is the first or second Gaussian component. In this hypothetical case, the log-likelihood function would be

$$\begin{aligned} l_0(\boldsymbol{\theta};\boldsymbol{x},\boldsymbol{\Delta}) = &\sum_{i=1}^{N} \left[(1-\Delta_i)\log\left(\varphi_{(\mu_1,\sigma_1^2)}(x_i)\right) + \Delta_i \log\left(\varphi_{(\mu_2,\sigma_2^2)}(x_i)\right)\right] + \\ &+ \sum_{i=1}^{N} \left[(1-\Delta_i)\log(1-\lambda) + \Delta_i \log(\lambda)\right]. \end{aligned}$$

If the Δ_i's were known, the maximum-likelihood estimate for μ_1 and σ_1^2 would be the sample mean and sample variance from the observations with $\Delta_i = 0$. The same holds true for μ_2, σ_2^2 and $\Delta_i = 1$. The estimate for λ would be the proportion of $\Delta_i = 1$. However, as the Δ_i's are not known, we use as their surrogates the conditional expectations

$$\gamma_i(\boldsymbol{\theta}) = \mathbb{E}[\Delta_i|\boldsymbol{\theta},\boldsymbol{x}] = \mathbb{P}\big(\Delta_i = 1|\boldsymbol{\theta},\boldsymbol{x}\big) \quad i = 1, \ldots, N$$

called *responsability* of model 2 for observation i. The iterative procedure called EM algorithm consists in alternating an *expectation* step in which we assign to each observation the probability to come from each model, and a *maximization* step where these responsabilities are used to update ML estimates.

Algorithm 1 Expectation-Maximization (EM) for 2-component GM

1: take initial guesses for parameters $\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\lambda}$

2: *Expectation* step: compute responsabilities

$$\widehat{\gamma}_i = \frac{\widehat{\lambda}\varphi_{(\widehat{\mu}_2,\widehat{\sigma}_2^2)}(x_i)}{(1-\widehat{\lambda})\varphi_{(\widehat{\mu}_1,\widehat{\sigma}_1^2)}(x_i) + \widehat{\lambda}\varphi_{(\widehat{\mu}_2,\widehat{\sigma}_2^2)}(x_i)}, \quad i = 1, \dots, N$$

3: *Maximization* step: compute weighted means and standard deviations

$$\widehat{\mu}_1 = \frac{\sum_{i=1}^N (1-\widehat{\gamma}_i)x_i}{\sum_{i=1}^N (1-\widehat{\gamma}_i)}, \qquad \widehat{\sigma}_1^2 = \frac{\sum_{i=1}^N (1-\widehat{\gamma}_i)(x_i-\widehat{\mu}_1)^2}{\sum_{i=1}^N (1-\widehat{\gamma}_i)}$$

$$\widehat{\mu}_2 = \frac{\sum_{i=1}^N \widehat{\gamma}_i x_i}{\sum_{i=1}^N \widehat{\gamma}_i}, \qquad \widehat{\sigma}_2^2 = \frac{\sum_{i=1}^N \widehat{\gamma}_i (x_i-\widehat{\mu}_2)^2}{\sum_{i=1}^N \widehat{\gamma}_i}$$

4: Iterate 2 and 3 until convergence.

A reasonable starting value for $\widehat{\mu}_1$ and $\widehat{\mu}_2$ is a random sample point x_i, both $\widehat{\sigma}_1, \widehat{\sigma}_2$ can be set equal to the sample variance and $\widehat{\lambda} = 0.5$. A full implementation of the EM algorithm is available in MATLAB.

4.2 GH calibration

In this section we present a modified EM scheme (the MCECM algorithm) for fitting a GH model to data. In Definition (3.3.1) we introduced the GH distribution using the so-called $(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$-parametrization. Although this is the most convenient one from a modeling perspective, it comes with an identification issue: the distributions $GH(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ and $GH(\lambda, \chi/k, k\psi, \boldsymbol{\mu}, k\boldsymbol{\Sigma}, k\boldsymbol{\gamma})$ are the same (it is easily seen by writing the density (3.6) in the two cases). To solve this problem, we require the mixing random variable W (see Definition (3.3.1)) to have expectation equal to 1. From Equation (B.2) we have

$$\mathbb{E}[W] = \sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}\big(\sqrt{\chi\psi}\big)}{K_\lambda\big(\sqrt{\chi\psi}\big)} = 1$$

and if we set $\bar{\alpha} = \sqrt{\chi\psi}$ it follows that

$$\psi = \bar{\alpha}\frac{K_{\lambda+1}(\bar{\alpha})}{K_\lambda(\bar{\alpha})}, \qquad \chi = \frac{\bar{\alpha}^2}{\psi} = \bar{\alpha}\frac{K_\lambda(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})} \tag{4.3}$$

The relations above define the $(\lambda, \bar{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$-parametrization, which will be used in the MCECM algorithm.

Let $\boldsymbol{X} \sim GH_m(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ and $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a realization of an iid random sample. Our objective is to find an estimate of the parameters represented by $\boldsymbol{\theta} = [\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}]^T$. The log-likelihood function to be maximized is

$$\log L(\boldsymbol{\theta}; \boldsymbol{x}) = \log L(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \log f_{\boldsymbol{X}}(\boldsymbol{x}_i; \theta) \tag{4.4}$$

where $f_{\boldsymbol{X}}$ is the function in (3.6). It well-known that finding a maximizer of (4.4) might be difficult, therefore we resort to a different approach. The situation would look much better if we could observe the latent mixing variables $W_1, \ldots, W_n$. Let us suppose to be in this fortunate situation and define the augmented log-likelihood function

$$\begin{aligned} \log \widetilde{L}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, W_1, \ldots, W_n) = &\sum_{i=1}^{n} \log f_{\boldsymbol{X}|W}(\boldsymbol{x}_i | W_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}) + \\ &+ \sum_{i=1}^{n} \log h_W(W_i; \lambda, \chi, \psi) \end{aligned} \tag{4.5}$$

where we used the fact that $f_{(\boldsymbol{X}_i, W_i)}(\boldsymbol{x}, w; \boldsymbol{\theta}) = f_{\boldsymbol{X}_i|W_i}(\boldsymbol{x}|w; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}) h_{W_i}(w; \lambda, \chi, \psi)$ and h_{W_i} is the density in (B.1). The advantage of this augmented formulation is that the two terms in (4.5) can be maximized separately. Although counter-intuitive, the first term involving the difficult parameters (e.g. a matrix), is the easiest to maximize and it is done analytically; the second term has to be treated numerically instead. To overcome the latency of the mixing variables W_i's, the MCECM algorithm is used. The algorithm consists in alternating an *expectation* step (in which the W_i's are replaced by an estimate deducted from the data and the current parameters estimate) and a *maximization* step (where parameters estimates are updated). Suppose we are at iteration k and $\boldsymbol{\theta}^{(k)}$ is the current parameters estimate, the two steps are as follows

- **E-step**: compute the conditional expectation of the augmented log-likelihood function given the data and the current parameters estimate

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = \mathbb{E}[\log \widetilde{L}(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{W}) | \boldsymbol{x}, \boldsymbol{\theta}^{(k)}] \tag{4.6}$$

- **M-step**: maximize $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$ to get $\boldsymbol{\theta}^{(k+1)}$.

In practice, the E-step amounts to numerically maximize the second term in (4.6), which is

$$\mathbb{E}\Big[\sum_{i=1}^{n}\log h_{W_i}(W_i;\lambda,\chi,\psi)\Big|\boldsymbol{x},\boldsymbol{\theta}\Big]=\sum_{i=1}^{n}-\lambda\log\chi+\lambda\log\sqrt{\chi\psi}+ \tag{4.7}$$
$$-\log 2K_\lambda(\sqrt{\chi\psi})+(\lambda-1)\underbrace{\mathbb{E}\big[\log W_i|\boldsymbol{x},\boldsymbol{\theta}^{(k)}\big]}_{\xi_i}-\tfrac{1}{2}\chi\underbrace{\mathbb{E}\big[W_i^{-1}|\boldsymbol{x},\boldsymbol{\theta}^{(k)}\big]}_{\delta_i}+$$
$$-\tfrac{1}{2}\psi\underbrace{\mathbb{E}\big[W_i|\boldsymbol{x},\boldsymbol{\theta}^{(k)}\big]}_{\eta_i}=n\big(-\lambda\log\chi+\lambda\log\sqrt{\chi\psi}-\log 2K_\lambda(\sqrt{\chi\psi})\big)+$$
$$+(\lambda-1)\sum_{i=1}^{n}\xi_i-\tfrac{1}{2}\chi\sum_{i=1}^{n}\delta_i-\tfrac{1}{2}\sum_{i=1}^{n}\eta_i.$$

In order to proceed further, we need to compute the conditional expectations ξ_i,δ_i and η_i. Thankfully, the following results holds (see Appendix E.1 in [1])

$$W_i|\boldsymbol{x}_i\sim\mathcal{N}^-\big(\underbrace{\lambda-\tfrac{1}{2}d}_{\widetilde{\lambda}},\underbrace{\chi+(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})}_{\widetilde{\chi}},\underbrace{\psi+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}_{\widetilde{\psi}}\big).$$

By using Equations (B.2) and (B.3) we end up with

$$\delta_i=\mathbb{E}[W_i^{-1}|\boldsymbol{x},\boldsymbol{\theta}^{(k)}]=\Big(\frac{\widetilde{\chi}}{\widetilde{\psi}}\Big)^{-\frac{1}{2}}\frac{K_{\lambda-1}(\sqrt{\widetilde{\chi}\widetilde{\psi}})}{K_\lambda(\sqrt{\widetilde{\chi}\widetilde{\psi}})} \tag{4.8}$$

$$\eta_i=\mathbb{E}[W_i|\boldsymbol{x},\boldsymbol{\theta}^{(k)}]=\Big(\frac{\widetilde{\chi}}{\widetilde{\psi}}\Big)^{\frac{1}{2}}\frac{K_{\lambda+1}(\sqrt{\widetilde{\chi}\widetilde{\psi}})}{K_\lambda(\sqrt{\widetilde{\chi}\widetilde{\psi}})} \tag{4.9}$$

$$\xi_i=\mathbb{E}[\log W_i|\boldsymbol{x},\boldsymbol{\theta}^{(k)}]=\frac{\mathrm{d}}{\mathrm{d}\alpha}\left\{\Big(\frac{\widetilde{\chi}}{\widetilde{\psi}}\Big)^{\frac{\alpha}{2}}\frac{K_{\lambda+\alpha}(\sqrt{\widetilde{\chi}\widetilde{\psi}})}{K_\lambda(\sqrt{\widetilde{\chi}\widetilde{\psi}})}\right\}_{\alpha=0} \tag{4.10}$$

We have now all the ingredients to present the MCECM algorithm as exposed in [1]

Algorithm 2 MCECM

1: Select reasonable starting points. For instance $\lambda^{(1)} = 1, \bar{\alpha}^{(1)} = 1$, $\boldsymbol{\mu}^{(1)} =$ sample mean, $\boldsymbol{\Sigma}^{(1)} =$ sample covariance and $\boldsymbol{\gamma}^{(1)} = \mathbf{0}$

2: Compute $\chi^{(k)}$ and $\psi^{(k)}$ using (4.3)

3: Compute the weights η_i and δ_i using (4.8) and (4.9). Average the weights to get

$$\bar{\eta}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\eta_i \qquad \bar{\delta}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\delta_i$$

4: If a symmetric model is to be fitted set $\boldsymbol{\gamma} = \mathbf{0}$, else

$$\boldsymbol{\gamma}^{(k+1)} = \frac{1}{n}\frac{\sum_{i=1}^{n}(\delta_i^{(k)}(\bar{\boldsymbol{x}} - \boldsymbol{x}_i)}{\bar{\eta}^{(k)}\bar{\delta}^{(k)} - 1}$$

5: Update $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\Sigma}^{(k)}$

$$\boldsymbol{\mu}^{(k+1)} = \frac{1}{n}\frac{\sum_{i=1}^{n}\delta_i^{(k)}(\boldsymbol{x}_i - \boldsymbol{\gamma}^{(k+1)})}{\bar{\delta}^{(k)}}$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{(k)}(\boldsymbol{x}_i - \boldsymbol{\mu}^{(k+1)})(\boldsymbol{x}_i - \boldsymbol{\mu}^{(k+1)})^T - \bar{\eta}^{(k)}\boldsymbol{\gamma}^{(k+1)}\boldsymbol{\gamma}^{(k+1)T}$$

6: Set $\boldsymbol{\theta}^{(k,2)} = [\lambda^{(k)}, \bar{\alpha}^{(k)}\boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}, \boldsymbol{\gamma}^{(k+1)}]$ and compute $\eta_i^{(k,2)}, \delta_i^{(k,2)}$ and $\xi_i^{(k,2)}$ using (4.9),(4.8) and (4.10)

7: Maximize (4.7) with respect to λ and $\bar{\alpha}$ (using relation (4.3)) to complete the calculation of $\boldsymbol{\theta}^{(k,2)}$. Go to step 2

Appendix A

proofs

this is the appendix

Appendix B

Probability Distributions

In this appendix we give further details about the probability distributions used in part I. Main references are [1], [9] and [8].

B.1 Generalized Inverse Gaussian

***Definition* B.1.1** (Bessel function)**:** The modified Bessel function of the third kind (simply called **Bessel function**) is defined as

$$K_\nu(x) = \frac{1}{2}\int_0^\infty t^{\nu-1}\exp\big\{-\frac{1}{2}x(t+t^{-1})\big\}\mathrm{d}t, \quad x>0.$$

***Definition* B.1.2** (Generalized Inverse Gaussian)**:** the density of a **Generalized Inverse Gaussian** (GIG) random variable W ($W\sim\mathcal{N}^-\big(\lambda,\chi,\psi\big)$) is

$$f_{GIG}(w) = \Big(\frac{\psi}{\chi}\Big)^{\frac{\lambda}{2}}\frac{w^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})}\exp\Big\{-\frac{1}{2}\big(\frac{\chi}{w}+\psi w\big)\Big\} \tag{B.1}$$

with parameters satisfying

$$\begin{cases}\chi>0,\psi\geq 0, & \text{if} \quad \lambda<0\\ \chi>0,\psi>0, & \text{if} \quad \lambda=0\\ \chi\geq 0,\psi>0, & \text{if} \quad \lambda>0\end{cases}$$

Useful formulas The following formulas are used in the text:

$$\mathbb{E}[W^n] = \Big(\frac{\chi}{\psi}\Big)^{\frac{n}{2}}\frac{K_{\lambda+n}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} \tag{B.2}$$

$$\mathbb{E}[\log W] = \Big\{\frac{\mathrm{d}\mathbb{E}[X^\alpha]}{\mathrm{d}\alpha}\Big\}_{\alpha=0} \tag{B.3}$$

B.2 Density Functions

We give here the probability density function for the general multivariate GH distribution and same special cases

B.2.1 GH

$$f(\boldsymbol{x}) = c\frac{K_{\lambda-\frac{m}{2}}\left(\sqrt{\big(\chi+Q(\boldsymbol{x})\big)\big(\psi+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)}\right)\exp\left\{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}\boldsymbol{\gamma}\right\}}{\left(\sqrt{\big(\chi+Q(\boldsymbol{x})\big)\big(\psi+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)}\right)^{\frac{m}{2}-\lambda}} \tag{B.4}$$

where $Q(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ and

$$c = \frac{\big(\sqrt{\chi\psi}\big)^{-\lambda}\psi^{\lambda}\big(\psi+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)^{\frac{m}{2}-\lambda}}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})}$$

B.2.2 Student-t

Setting the degree of freedom $\nu = -2\lambda$ the density reads

$$f(\boldsymbol{x}) = c\frac{K_{\frac{\nu+m}{2}}\left(\sqrt{\big(\nu-2+Q(\boldsymbol{x})\big)\big(\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)}\right)\exp\left\{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}\boldsymbol{\gamma}\right\}}{\left(\sqrt{\big(\nu-2+Q(\boldsymbol{x})\big)\big(\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)}\right)^{\frac{\nu+m}{2}}} \tag{B.5}$$

where

$$c = \frac{\big(\nu-2\big)^{\frac{\nu}{2}}\big(\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)^{\frac{\nu+m}{2}}}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}-1}}$$

B.2.3 VG

$$f(\boldsymbol{x}) = c\frac{K_{\lambda-\frac{m}{2}}\left(\sqrt{Q(\boldsymbol{x})\big(2\lambda+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)}\right)\exp\left\{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}\boldsymbol{\gamma}\right\}}{\left(\sqrt{Q(\boldsymbol{x})\big(2\lambda+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)}\right)^{\frac{m}{2}-\lambda}} \tag{B.6}$$

where

$$c = \frac{2\lambda^{\lambda}\big(2\lambda+\boldsymbol{\gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\big)^{\frac{m}{2}-\lambda}}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\Gamma(\lambda)}$$

Bibliography

[1] Breymann, W., and Luthi, D. ghyp: A package on generelized hyperbolic distributions.

[2] Buckley, I., Saunders, D., and Seco, L. Portfolio optimization when asset returns have the gaussian mixture distribution. *European Journal of Operational Research 185*, 3 (2008), 1434 – 1461.

[3] Casella, G., and Berger, R. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.

[4] Eisenberger, I. Genesis of bimodal distributions. *Technometrics 6*, 4 (1964), 357–363.

[5] Everitt, B., and Hand, D. *Finite Mixture Distributions*. Monographs on applied probability and statistics. Chapman and Hall, 1981.

[6] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer New York, 2009.

[7] Jacod, J., and Protter, P. *Probability Essentials*. Hochschultext / Universitext. Springer, 2000.

[8] McNeil, A. J., Frey, R., and Embrechts, P. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ, USA, 2005.

[9] Paolella, M. *Intermediate Probability: A Computational Approach*. Wiley, 2007.

[10] Pearson, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 185* (1894), 71–110.

[11] Plasse, J. H. The em algorithm in multivariate gaussian mixture models using anderson acceleration. https://pdfs.semanticscholar.org/d8c1/650966042e2a58fedf6ce04ef002dd6fb3a8.pdf, 2013.

[12] Pola, G., Lygeros, J., Domenica, M., and Benedetto, D. Invariance in stochastic dynamical control systems.

[13] Pola, G., and Pola, G. Optimal dynamic asset allocation: A stochastic invariance approach. In *Proceedings of the 45th IEEE Conference on Decision and Control* (Dec 2006), pp. 2589–2594.

[14] Pola, G., and Pola, G. A stochastic reachability approach to portfolio construction in finance industry. *IEEE Transactions on Control Systems Technology 20*, 1 (Jan 2012), 189–195.