Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Corso di Laurea Magistrale in Ingegneria Matematica

A Stochastic Reachability Approach to Asset Allocation

Relatore: Prof. Emilio BARUCCI
Correlatore: Dott. Gianni POLA

Tesi di Laurea di:
Andrea SCHIAVON
Matr. 852559

Anno Accademico 2016-2017

Abstract

Before the 1950s, managing other people's money was a discipline as far away from being scientifically dictated as it could ever get. For instance, before Markowitz's revolutionary paper *Portfolio Selection*, diversification was not a universally recognized practice in Asset Allocation. Markowitz's work paved the way for new ideas from different scientific and academical fields to influence Finance and Asset Allocation in particular.

Continuing along this line, in this thesis cutting-edge results in Stochastic Reachability (which is a concept belonging to the theory of Control Systems) are employed to tackle the Asset Allocation problem. In particular, once an investor has specified his risk profile and a target return, the model will output an optimal investment strategy having the feature of maximizing the probability of reaching the target return while keeping the risk under control. This strategy will exhibit a *contrarian* behavior, namely it prescribes to buy risky assets (in order to achieve a riskier position) when performance is down and to sell them when performance is up.

What are the drivers that lead a portfolio manager to rebalance portfolio weights? In Part I, the case where time triggers a portfolio rebalancing will be explored. Although this Time-Driven approach is the most intuitive, it might incur in non-negligible transaction costs if the rebalancing frequency is high. On the other hand, in Part II, what causes the portfolio mix to be readjusted will be the fact that the risky asset cumulative return hits a lower or upper barrier. This portfolio rebalancing mechanics leads to the so-called Event-Driven approach to Asset Allocation.

Keywords: Asset Allocation, Stochastic Reachability, Time-Driven approach, Event-Driven approach.

Sommario

Contrariamente a quanto accade oggi per i maggiori investitori istituzionali, prima degli anni '50 chi investiva in borsa lo faceva senza adottare un metodo rigorosamente scientifico. Il cambio di rotta avvenne nel 1952, quando Harry Markowitz, fondatore della moderna teoria del portafoglio, pubblicò *Portfolio Selection*, un articolo che aprì la strada ad un flusso di nuove idee provenienti dall'accademia e da svariate discipline scientifiche, idee che saranno destinate a rivoluzionare il modo di investire nei decenni successivi.

Proseguendo in questa direzione, nel nostro lavoro vengono utilizzati recenti risultati in Stochastic Reachability (concetto sviluppato nella teoria dei Sistemi di Controllo) per riadattarli in un contesto di Asset Allocation. In particolare, una volta individuato un adeguato profilo di rischio e un rendimento da raggiungere, il modello produrrà una strategia di investimento che massimizza la probabilità di raggiungere tale rendimento tenendo allo stesso tempo sotto controllo il rischio. Questa strategia avrà la caratteristica di essere *Contrarian*, cioè prescrive di acquistare titoli rischiosi quando la performance del portafoglio è buona mentre prescriverà di venderli, in favore di titoli privi di rischio, se la performance è bassa.

Con che criterio dunque, un gestore decide di ribilanciare i pesi di portafoglio? Nella prima parte del lavoro verrà presentato l'approccio Time-Driven, in cui è il tempo a dettare la riallocazione dei pesi (e.g. settimanalmente). Nonostante questo sia il metodo più intuitivo, ha come svantaggio gli elevati costi di transazione nel caso di frequenza di riallocazione alta. Nella seconda parte invece, l'approccio che si segue è quello Event-Driven. Ciò significa che una riallocazione di portafoglio viene effettuata solo quando il valore assoluto del rendimento cumulato dell'asset rischioso supera una certa soglia.

Parole chiave: Asset Allocation, Stochastic Reachability, Time-Driven approach, Event-Driven approach.

Contents

List of Tables

List of Figures

List of Algorithms

Acronyms

ATM Air Traffic Management.

BS Black and Scholes.

CAPM Capital Asset Pricing Model.

CDF Cumulative Distribution Function.

CF Characteristic Function.

CPPI Constant-Proportion Portfolio Insurance.

DES Discrete Event System.

DP Dynamic Programming.

ED Event-Driven.

EM Expectation-Maximization.

EMH Efficient-Market Hypotesis.

G Gaussian.

GBM Geometric Brownian Motion.

GH Generalized Hyperbolic.

GIG Generalized Inverse Gaussian.

GM Gaussian Mixture.

MCECM Multi-Cycle Expectation Conditional Maximization.

ML Maximum Likelihood.

MM Method of Moments.

MPT Modern Portfolio Theory.

NIG Normal Inverse Gaussian.

ODAA Optimal Dynamic Asset Allocation.

SDE Stochastic Differential Equations.

TD Time-Driven.

Chapter 1

Introduction

Since the first stock exchange opened for trades, investors have been trying to find ways to allocate their wealth among different securities in order to maximize returns. As a matter of fact, the techniques which are being employed nowadays in the investment industry are quite different from those used in the first half of the 20th century. "Fifty years ago, the business of managing other people's money was very much an art not a science, and was largely a matter of finding someone who was privy to inside information. But during the 1950s, 1960s and 1970s, academics changed the study of what became known as portfolio management. They did so in the face of much initial resistance and scepticism from the industry"[1].

1.1 The academia meets the industry

In those days, any market participant would have been aware that investing was a risky business (nothing ventured, nothing gained). However, a formal and systematic connection between risk and return was still missing. The "annus mirabilis" in asset allocation was 1952, when Harry Markowitz published his pioneering paper *Portfolio Selection* in the *Journal of Finance*, starting the academic invasion of the financial industry. In his paper, which is considered to be the starting point of Modern Portfolio Theory (MPT), Markowitz outlined for the first time how investors should allocate assets so as to achieve the highest returns given a certain level of risk. After having estimated expected returns and the covariance between each security, an investor, according to Markowitz, has to solve a quadratic programming problem for obtaining the so-called *efficient portfolio frontier*. Any portfolio (a mix of securities) belonging to the frontier is efficient in the sense that it

[1] See [2].

provides the highest expected return for a given level of risk (standard deviation). Moreover, each of these portfolios has the feature of being deversified. In Markowitz's words[2]:

> A portfolio with sixty different railway securities, for example, would not be as well diversified as the same size portfolio with some railroad, some public utility, mining, various sort of manufacturing, etc. The reason is that it is generally more likely for firms within the same industry to do poorly at the same time than for firms in dissimilar industries.

Building on Markowitz's work, the second major breakthrough is the Capital Asset Pricing Model (CAPM) and it was made by another University professor, William Sharpe. The CAPM, which appeared in the *Journal of Finance* in 1964, allows one to compute the expected return from an asset in terms of its risk. The risk is divided into two components, namely a *systematic risk* (which is related to the return of the whole market and cannot be eliminated) and a *non-systematic risk* (which is unique to the asset and could be eliminated by diversifying the portfolio) [18]. In spite of the unrealistic assumptions which the CAPM is based on [19], it has proved to be a useful tool for portfolio managers. Other key contributions academia made to the industry are the Efficient-Market Hypotesis (EMH) (Fama, 1970) and the Black and Scholes (BS) model. EMH is a theory that states that security prices perfectly reflect all available information in the market and, consequently, the whole market cannot be beaten. On the other hand, the BS model is a mathematically-rigorous theory for pricing options.

Although all of the models above have shown some shortcomings when applied to the complex reality of capital markets, they have been crucial steps for reaching what Asset Allocation is today. Nonetheless, ideas from other scientific fields have continued to enrich Finance and in particular Asset Allocation. An example of this flourishing contamination is the techniques of Stochastic Reachability presented in this works which have been borrowed from the theory of Control Systems and applied to the asset allocation problem.

1.2 Structure

In this section the structure of the thesis is outlined.

In Chapter 2 we introduce the asset allocation problem and all the quantities related to it, such us the asset class return vector and the portfolio

[2]See [6].

dynamics. Then, the concept of Stochastic Reachability is explored giving an idea about its possible fields of application. We conclude the chapter by discussing the mathematical formulation of the asset allocation problem in a Stochastic Reachability framework. In particular, the ODAA algorithm, which all the thesis rely on, is enunciated.

In Chapter 3 we discuss three models which could be employed to describe the probabilistic properties of the asset class returns vector, namely the Gaussian (G), GM and Generalized Hyperbolic (GH) model. For each of them, the model-related features (*risk* constraint and portfolio value density function) of the ODAA algorithm are obtained.

In Chapter 4 we focus on techniques for calibrating the models presented in Chapter 3 to market data. For the GM case, calibration performance between three calibration methods are compared.

Chapter 5 is dedicated to presenting the numerical results for the time-driven approach. The ODAA asset allocation strategy is compared to the Constant-Mix and the CPPI, which are two benchmark policies in the industry. This will end Part I

Part II begins with Chapter 6, where a brief introduction of the Discrete Event System (DES) theory is given. Afterwards, the asset allocation problem is cast in an Event-Driven (ED) setting, calibration of model parameters is discussed and numerical results are given.

Chapter 7 constitutes the original part of the thesis. In this chapter we attempt to generalize the basic ED model of Chapter 6 in two different ways: first, by modeling the risky asset as a GBM and then by assuming a stochastic dynamics (Vasicek model) for the risk-free interest rate.

Finally, in Chapter 8 we sum up what has been achieved in this thesis and propose future research directions.

The MATLAB code used to implement the models presented in the thesis can be found in the following GitHub repository: `https://github.com/skiamu/Thesis`.

Part I

Time-driven approach

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 [28],[27] and [26].

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 an m-dimensional random vector $\boldsymbol{w}_k = \left[w_k(1), \ldots, w_k(m)\right]^T$ 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 ith asset class and $\{z_k(i)\}_{k \in \mathbb{N}}$ the ith 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 (i.e. correlation between $w_k(i)$ and $w_k(j)$ for $i, j = 1, \ldots, m$)

- *time-lagged* correlation, that is the correlation among different asset class at different time period (i.e. correlation between $w_k(i)$ and $w_{k'}(j)$, with $k \neq k'$ for $i, j = 1, \dots, m$).

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

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

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

An asset allocation at period $k \in \mathbb{N}$ is a vector $\boldsymbol{u}_k \in \mathbb{R}^m$ whose ith 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 goal. The 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 $k \in \mathbb{N}$ (called **rebalancing time**) so as to achieve investor's goal. If the investor is mainly concerned about the final return, the allocation strategy is called **total-return allocation**. On the other hand, if his objective is beating a benchmark (an index created to include multiple securities representing some aspect of the total market), the strategy is called **benchmark allocation**. In the following, we will consider only total-return portfolios.

As well as setting the target return, the investor specifies other requirements that the portfolio manager must take into consideration. This means that 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 set of constraints:

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

The form of the *risk* constraint will actually depend on the model used to describe the probabilistic properties of vector $\boldsymbol{w}_k$. In Chapter 3 we will tackle this issue. Let us now cast the asset allocation problem in a more general mathematical framework.

2.2 Stochastic Reachability Approach

In the previous section the financial setting has been laid out, 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.

2.2.1 The concept of Stochastic Reachability

"In general terms, a reachability problem consists of determining if a given system trajectory will eventually enter a prespecified set starting from some initial state" [10]. For deterministic systems, reachability analysis amounts to compute the set of states that can be reached by system trajectories. However, most of real-life problem are non-deterministic and uncertainty must be taken into account. In these cases, the main concern is determining the probability that the system reaches a prespecified set. "Typically, a certain part of the state space is "unsafe" and the control input of the system has to be chosen so as to keep the state away from it" [10]. One of the most successful application of stochastic reachability techniques has been Air Traffic Management (ATM). "Within the ATM context, safety-critical situation arise during flight when an aircraft comes closer than a minimum allowed distance to another aircraft or enters a forbidden region of the airspace. In the current ATM system, air traffic controllers are in charge of guaranteeing safety by issuing to pilots corrective actions on their flight plans when a safety-critical situation is predicted" [10].

Conversely, when Stochastic Reachability is applied to the financial asset allocation problem, a dual viewpoint is taken. In this context, the focus is on driving the system state (the value of a portfolio of securities) into a "safe" set, and computing the probability that this occurs. The air traffic controller becomes a portfolio manager and signals issued to the pilot turns into orders to traders to buy or sell assets so as to adjust the portfolio mix of securities.

2.2.2 Mathematical Formulation

Let us introduce the following stochastic discrete-time dynamic control system

$$x_{k+1} = f(x_k, \boldsymbol{u}_k, \boldsymbol{w}_{k+1}) = x_k(1 + \boldsymbol{u}_k^T \boldsymbol{w}_{k+1}) \tag{2.2}$$

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

- $x_k \in \mathcal{X} = \mathbb{R}$ is the system state (the portfolio value), $\mathcal{X}$ the system space
- $\boldsymbol{u}_k \in U \subset \mathbb{R}^m$ is the control input (the asset allocation vector), U the control input space
- $\boldsymbol{w}_k$ is a m-dimensional random vector (the 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 we are interested in, namely the time-varying control maps. Any $\mu \in \mathcal{U}$ is a map such that for any $x \in \mathcal{X}$ and any $k \in \mathbb{N}$, it associates an asset allocation vector $\boldsymbol{u}_k \in U$. The control input space U is shaped by the *budget*, *long-only* and *risk* constraint. Given $N \in \mathbb{N}$ we define the set of control sequences as

$$\mathcal{U}_N = \Big\{\pi = \{\mu_k\}_{k=0,\dots,N} : \mu_k \in \mathcal{U}\Big\}$$

and call any $\pi \in \mathcal{U}_N$ a **control policy**. Moreover, let us denote by π^k a control policy starting at period k, that is $\pi^k = \{\mu_k, \dots, \mu_N\}$. We now have all the necessary ingredients to formulate the asset allocation problem in stochastic reachability terms.

***Problem* 2.2.1** (Optimal Dynamic Asset Allocation 1)**:** 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

$$\mathbb{P}\Big(\Big\{\omega \in \Omega : x_0 \in X_0, \dots, x_N \in X_N\Big\}\Big). \tag{2.3}$$

The target sets $\{X_1, \dots, X_N\}$ represent investor's goal and we can think of them as the "safe" states where we want the portfolio value to belong to. For instance, a target set could be $X_k = [\underline{x}_k, \infty)$. Problem (2.2.1) is going to be solved by resorting to Dynamic Programming (DP). However, before doing that, we need to make explicit the dependence in (2.3) from the control policy π. To this end, let $p_{f(x,\boldsymbol{u},\boldsymbol{w}_{k+1})}$ be the density of random variable (2.2), once x_k has been fixed to $x \in \mathcal{X}$, and let us introduce the following function.

***Definition* 2.2.1** (Value function)**:** Given a sequence of target sets $\{X_k\}_{k\geq 0}$, the **value function** associated with Problem 2.2.1 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 (2.3) to the value function in the following way (see [26])

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

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

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

Having restated the Optimal Dynamic Asset Allocation (ODAA) problem in terms of the value function V has been crucial in order to directly apply the powerful technique of DP and solve it [26]. The main result is given in the following theorem, that is the cornerstone on which this work is based on.

***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

$$\boxed{\begin{aligned} &J_N(x) = \mathbb{1}_{X_N}(x) \\ &J_k(x) = \sup_{\boldsymbol{u}_k \in U_k} \int_{X_{k+1}} J_{k+1}(z) p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}(z) \mathrm{d}z \\ &\qquad k = N-1, \dots, 1, 0. \end{aligned}} \tag{2.4}$$

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 joint probability $p^\star$ of reaching the target sets. It is worth pointing out some interesting features of the ODAA algorithm (2.4):

- $J_k(x)$ 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) of dimension $m \in \mathbb{N}$. At each iteration $k = N-1, \dots, 1, 0$, the optimization has to be repeated for each x belonging to X_k (in practice, this set will be discretized with a fix step length to have a finite number of optimizations).
- The algorithm presented in theorem (2.2.1) does not depend on a particular distribution of random variable $f(x, \boldsymbol{u}_k, \boldsymbol{w}_{k+1})$ as long as its explicit functional form is available. Hence the reason to prefer multivariate distribution closed under linear combination for modeling w_{k+1}. therefore, this distribution-free property gives us enough freedom to look outside the usual Guassian world.
- Given a period $k \in \mathbb{N}$ and a portfolio value realization $x \in \mathcal{X}$, $\mu_k^\star(x) \in U$ 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 in the previous chapter, the ODAA algorithm does not depend on a particular distribution of the asset class returns vector $\boldsymbol{w}_{k+1}$. However, by looking at (2.4) 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, after giving a brief theoretical introduction, we will discuss the model-related features of the ODAA algorithm. That is, first we will derive the portfolio value density function $p_{f(x,\boldsymbol{u}_k,\boldsymbol{w}_{k+1})}$ and secondly, an expression for the *risk* constraint (which also depends on the distribution chosen) will be obtained. Moreover, we assume stationarity, therefore the distribution of $\boldsymbol{w}_{k+1}$ will not depend on k.

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 $\boldsymbol{w} = \left[w_1, \ldots, w_m\right]^T$ 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 $\boldsymbol{\Sigma}$. By definition we have

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

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 *risk* constraint.

***Definition* 3.1.2** (loss function)**:** 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

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

***Definition* 3.1.3** (Value-at-risk)**:** 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 assess the risk they run to carry a portfolio of risky securities for a specified period of time (the portfolio must be kept constant during this time period). For instance, if our portfolio has an (*ex-post*) weekly $V@R_{0.99} = 7\%$, this means that 99% of the times our portfolio did not suffer a loss greater or equal than 7% over the investment period. In our case, we receive the V@R specification as input (*ex-ante* value-at-risk) 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

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

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

where 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 input space U which is the feasible set of the constrained optimization problem given in theorem (2.2.1) at each k.

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 [9].

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 key asset returns features which are observed, on the contrary, in 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 (GM) distribution is a validate alternative to the Gaussian model. Loosely speaking, the pdf of a GM random vector is 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 contours, as can be seen in Figure 3.1. 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).

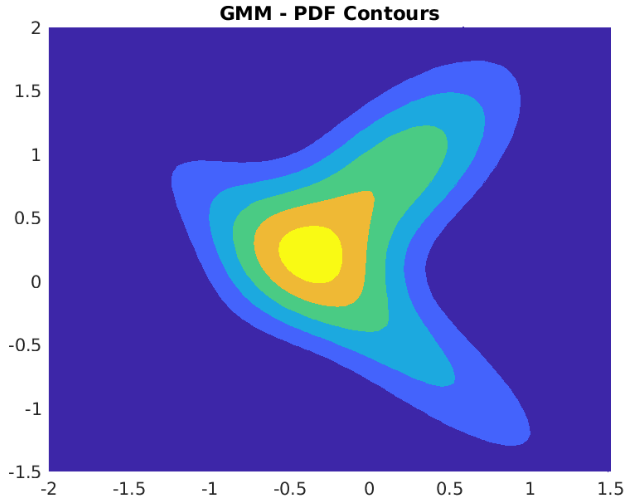

Figure 3.1: Example of a GM density contour plot with two mixing components.

Let us begin the more formal introduction on the GM distribution with 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,

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

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

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

where $\widetilde{X}_i \sim \mathcal{N}\big(\mu_i, \sigma_i^2\big)$. Since the CF completely characterizes the distribution (see [20], theorem 14.1) we have the result. □

Portfolio value density and *risk* constraint Suppose the asset class returns vector $\boldsymbol{w}_{k+1}$ follow a Gaussian Mixture distribution. We want to compute the density of random variable $f(x, \boldsymbol{u}_k, \boldsymbol{w}_{k+1}) = x(1 + \boldsymbol{u}_k^T \boldsymbol{w}_{k+1})$. Thanks to proposition (3.2.1), we know that the random variable $\boldsymbol{u}_k^T \boldsymbol{w}_{k+1}$ follows itself a GM (univariate) distribution. Moreover, by integration we easily derive its Cumulative Distribution Function (CDF). This allows us to write

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

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

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

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

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

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

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 as input by the investor. As far as the former is concerned, the following proposition gives us the answer [9]

***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, we obtain

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

3.3 Generelized Hyperbolic model

The last distribution we propose is the Generalized Hyperbolic (GH). Like the GM, in its general form also the GH presents a non-elliptical behavior 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 expression of the *risk* constraint. References for this section are [7] and [21].

***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

- $\boldsymbol{Z} \sim \mathcal{N}\big(\mathbf{0}, I_d\big)$
- $A \in \mathbb{R}^{m \times d}$ 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 A for the definition of the Generalized Inverse Gaussian (GIG) distribution). W is 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 $\boldsymbol{\gamma} = \mathbf{0}$ the distribution is symmetric around the mean
- $\boldsymbol{X}|W = w \sim \mathcal{N}\big(\boldsymbol{\mu} + w\boldsymbol{\gamma}, w\boldsymbol{\Sigma}\big)$.

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

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

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

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

where

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

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 Definition 3.3.1 and Equation (A.2); in the end we obtain

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

where

$$\begin{aligned}\mathbb{E}[W] &= \left(\frac{\chi}{\psi}\right)^{\frac{1}{2}}\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}\\ \mathrm{Var}\,[W] &= \left(\frac{\chi}{\psi}\right)\frac{1}{K_{\lambda}(\sqrt{\chi\psi})}\left\{K_{\lambda+2}(\sqrt{\chi\psi}) - \frac{K_{\lambda+1}^2(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}\right\}.\end{aligned}$$

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 of equity, bond and cash and a suitable index will be used to represent each of these markets (the dataset will be discussed in Section 5.1). 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, 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 [7]). Different calibration methods are available for the GM model, 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 them will be provided. Finally, in Section 4.2 the GH model will be fitted to data using the Multi-Cycle Expectation Conditional Maximization (MCECM) algorithm.

4.1 GM calibration

The problem of estimating the parameters of a GM distribution dates back to [23] 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 could provide the starting point for the EM algorithm. The main reference for the MM method is [15], for the ML [12] and for EM [21].

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 components. The idea behind MM is basically to match observed and theoretical moments; this translates into a system of polynomial equations that most of the time, for big-sized 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 along with unimodality constraints 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 $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ 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 $\{x_1, \dots, x_n\}$ 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 ith marginal is

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

where μ_{ji} and σ^2_{ji} denote respectively the ith element of the jth mixing component mean vector and the ith diagonal entry of the jth mixing component covariance matrix; in other words, the first subscript indicates the mixing component, the second the dimension. 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 be the one giving the smallest residual. Moreover, in the optimization process we also imposed the following unimodality constraints on each marginal[1]

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

[1]See [14] for the proof of this sufficient condition for unimodality for a 2-mixing-component GM density.

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

4.1.2 The Expectation-Maximization algorithm

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 [12]

***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 univariate GM distribution with 2 mixing components (as presented in [16]). The interested reader can refer to [16] for the general case or [25] for a more thorough 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 $\boldsymbol{x} = \{x_1, \dots, x_N\}$ 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 $i = 1, \dots, N$. Model 1 and model 2 indicate 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

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

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 EM algorithm 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 and also in Python.

4.1.3 MM vs ML vs EM

In this subsection we put the calibration methods into practice to see which one is better at recovering the parameters of a GM distribution. To this end, we simulated 10^4 observations from a GM distribution with the following parameters

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 6.11\text{e}{-4} \\ 1.373\text{e}{-3} \\ 2.34\text{e}{-3} \end{bmatrix} \qquad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 4.761\text{e}{-9} & 2.474\text{e}{-8} & 2.731\text{e}{-8} \\ & 3.21\text{e}{-5} & -2.55\text{e}{-6} \\ & & 3.656\text{e}{-4} \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 6.83\text{e}{-4} \\ -1.61\text{e}{-2} \\ -1.75\text{e}{-2} \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 3.844\text{e}{-9} & 2.42\text{e}{-8} & 6.739\text{e}{-8} \\ & 3.804\text{e}{-5} & -7.644\text{e}{-6} \\ & & 2.757\text{e}{-3} \end{bmatrix}$$

and $\lambda = 0.98$. In order to have a fair comparison, the two Gaussian regimes have a common correlation matrix

$$R = \begin{bmatrix} 1 & 6.33\text{e}{-2} & 2.07\text{e}{-8} \\ & 1 & -2.36\text{e}{-2} \\ & & 1 \end{bmatrix}$$

Then, we applied the three calibration methods introduced in the previous section. The result is summarized in Table 4.1, 4.2 and 4.3.

Parameter	MM	e_{MM} (%)	ML	e_{ML} (%)	EM	e_{EM} (%)
$\widehat{\mu}_1$	6.167e−4	0.94	6.097e−4	0.213	6.11e−4	0.0264
$\widehat{\mu}_2$	1.578e−3	14.98	1.464e−3	6.66	1.368e−3	0.366
$\widehat{\mu}_3$	2.396e−3	2.40	2.209e−3	5.57	2.174e−3	7.066
$\widehat{\Sigma}_{11}$	2.757e−8	479.2	4.632e−9	2.70	4.704e−9	1.189
$\widehat{\Sigma}_{22}$	3.215e−5	0.14	3.071e−5	4.31	3.155e−5	1.699
$\widehat{\Sigma}_{33}$	3.092e−4	15.4	3.532e−4	3.38	3.661e−4	0.146
$\widehat{\Sigma}_{12}$	−3.554e−8	243.6	2.617e−8	5.75	2.249e−8	9.123
$\widehat{\Sigma}_{13}$	−3.1e−8	213.5	3.546e−8	29.86	3.487e−8	27.68
$\widehat{\Sigma}_{23}$	−4.805e−7	81.20	−1.973e−6	22.82	−2.756e−4	7.788

Table 4.1: Estimates for the first mixing component and respective estimation errors

Parameter	MM	e_{MM} (%)	ML	e_{ML} (%)	EM	e_{EM} (%)
$\widehat{\mu}_1$	5.461e−4	20.03	6.772e−4	0.843	6.843e−4	0.193
$\widehat{\mu}_2$	−7.26e−3	54.91	−8.424e−3	47.70	−1.554e−2	3.481
$\widehat{\mu}_3$	−8.11e−3	53.65	−8.365e−3	52.21	−1.953e−2	11.605
$\widehat{\Sigma}_{11}$	1.157e−8	201.07	3.643e−9	5.21	3.223e−9	16.131
$\widehat{\Sigma}_{22}$	8.133e−5	113.79	8.792e−5	131.1	4.156e−5	9.254
$\widehat{\Sigma}_{33}$	2.108e−3	23.52	1.922e−3	30.30	2.941e−3	6.674
$\widehat{\Sigma}_{12}$	−3.662e−8	251.31	3.926e−8	62.21	1.545e−8	36.164
$\widehat{\Sigma}_{13}$	−5.244e−8	177.81	7.336e−8	8.86	2.693e−7	299.6
$\widehat{\Sigma}_{23}$	−1.996e−6	73.88	−7.787e−6	1.87	3.435e−5	549.4

Table 4.2: Estimates for the second mixing component and respective estimation errors

Parameter	MM	e_{MM}	ML	$e_{ML}(\%)$	EM	e_{EM} (%)
$\widetilde{\lambda}$	0.94	4.08	0.95	3.06	0.9812	0.119
$\log L^\star$	1.390e5		1.437e5		1.438e5	

Table 4.3: mixing proportion estimate and log-likelihood

From the tables above we see that the EM method is definitely the most accurate. Therefore, we decide to adopt it as the reference method for calibrating the GM model. As far as the MM method is concerned, the formulation given in Section 4.1.1 relies on the assumption of a common correlation matrix between the two Gaussian regimes. Although this assumption reduces the number of parameter to be estimated, there is empirical evidence (see [11]) that this is not the case in global financial markets where correlation between asset classes is actually increased during bear markets. Nonetheless, even if MM is not as accurate as EM, it is still a valuable method since it does not require full time series but only their sample statistics. This turns out to be particularly useful when distribution parameters are set via market hypothesis and economic views (e.g. bull market in the next investment period) instead of using historical data.

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 (A.2) we have

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

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 defines 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 (A.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, [7])

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

By using Equations (A.2) and (A.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 [7]

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

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

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

Chapter 5

Numerical Results in the Time-Driven Approach

This chapter is dedicated to presenting the results obtained by applying the Stochastic Reachability approach (discussed in Chapter 2) to the asset allocation problem. We recall that the output of the ODAA algorithm (see Theorem (2.2.1)) is a sequence of allocation maps $\pi^\star = \{\mu_0^\star, \dots, \mu_{N-1}^\star\}$. For any portfolio realization $x \in \mathbb{R}$ at time $k \in \mathbb{N}$, the maps $\mu_k^\star$ provides us with the optimal asset allocation $\mu_k^\star(x) = \boldsymbol{u}_k^\star$; for instance, if $\boldsymbol{u}_k^\star = \begin{bmatrix}0.2 & 0.2 & 0.6\end{bmatrix}^T$, this means that 20% of investor's wealth should be allocated to the first asset class, 20% to the second one and the remaining 60% to the third one. Objective of this chapter is to see what form these maps have at different time instants. The chapter unfolds as follows: in Section 5.1 the dataset is presented and summarized by some sample statistics, in Section 5.2 the parameters of the asset allocation problems are set and the allocation maps for the GM model are reported. Moreover, the ODAA strategy will be compared with other famous asset allocation strategies such as the Constant-Mix and the Constant-Proportion Portfolio Insurance (CPPI).

5.1 The Dataset

Our asset class menu consists of cash, bond and equity. To represent these markets we adopt the indexes presented in Table 5.1. The dataset is composed of weekly time series from 23 January 2010 to 15 April 2016. The data is downloaded from Yahoo Finance[1], which is also where the reader is referred for more details of index composition. An overview of the asset classes is given in Figure 5.1 and Table 5.2. By comparing the annualized mean

[1] `https://finance.yahoo.com/`.

Label	Asset Class	Index
C	Money Market	iShares Short Treasury Bond ETF
B	US Bond	Northern US Treasury Index
E	US Equity	S&P 500

Table 5.1: Asset class and relative index

return, it is clear that asset class Equity leads to higher performance than Bond and Bond, in turn, ensures higher performance than Cash. However, the annualized volatility tells us that the same hierarchy holds true also in terms of riskiness, being Equity the riskiest investment and Cash the least risky. Higher sample moments (Skewness and Kurtosis) suggest that the return distribution diverges significantly from a multivariate Gaussian. Indeed, a quantitaive proof of this fact is given us by the Henze-Zirkler[2] multivariate normality test which exhibits a zero p-value.

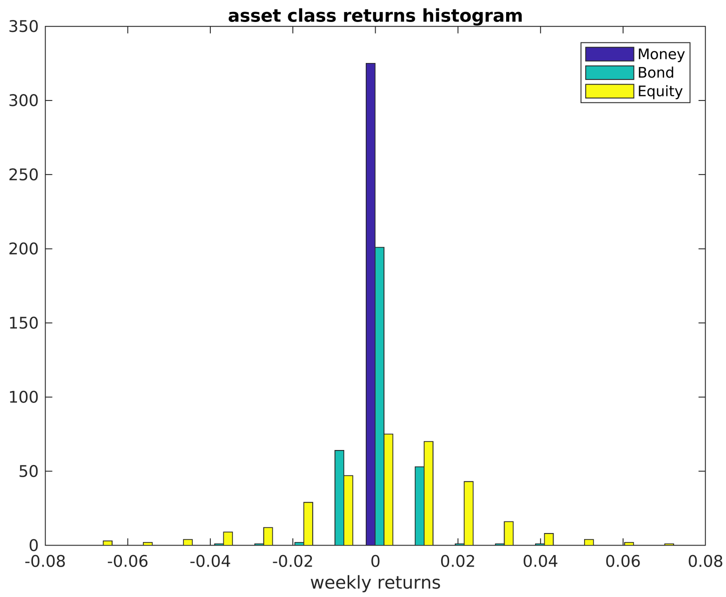

Figure 5.1: Weekly asset class returns histogram.

[2]See [1] for a MATLAB implementation.

Statistic	C	B	E
Mean Return (ann)	0.064%	3.46%	12.11%
Volatility (ann)	0.113%	4.81%	14.81%
Median (ann)	0%	4.58%	17.74%
Skewnwss	0.262	-0.0621	-0.36
Kurtosis	3.90	10.62	4.42
Monthly $V@R_{0.95}$	0.0808%	3.73%	14.95%
Max Drawdown	0.106%	5.87%	23.98%
Mean Drawdown	0.020%	1.5%	4.62%
Sharpe ratio	0	0.692	0.767

Table 5.2: Asset class returns sample statistics

Finally, the sample correlation matrix is

$$\begin{bmatrix} 1 & 0.166 & -0.075 \\ & 1 & -0.454 \\ & & 1 \end{bmatrix}.$$

5.2 Optimal Allocation Maps

Let us consider an investment characterized by the following parameters:

- 2-year investment horizon
- weekly rebalancing frequency, which means $N = 104$ portfolio rebalancings
- monthly (*ex-ante*) value-at-risk equals to 7%
- target return $\theta = 7\%$ per year
- initial wealth $x_0 = 1$.

The target sets we want our portfolio value to stay within are

$$\begin{aligned} X_0 &= \{1\} \\ X_k &= [0, \infty) \quad k = 1, \dots, 103 \\ X_{104} &= [(1+\theta)^2, \infty) = [1.07^2, \infty). \end{aligned}$$

In practice, these sets are discretized with a discretization step of 10^{-3} and truncated where the probability measure is negligible; the actual sets used in the implementation thus are $X_k = [0.5, 1.9]$ $k = 1, \dots, 103$ and $X_{104} = [(1.07)^2, 1.9]$. As stated in Problem 2.2.1, we are looking for a sequence of allocation maps which maximize the following joint probability

$$\mathbb{P}\big(\{\omega \in \Omega : x_0 \in X_0, \dots, x_{104} \in X_N\}\big).$$

The final choice to be made before running the ODAA algorithm is picking a model for the asset class returns. As an example, we opt for the GM model, which has been fitted to data applying the EM method (see Subsection 4.1.2). The parameter estimates are:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1.054\text{e}{-5} \\ 3.713\text{e}{-4} \\ 2.298\text{e}{-3} \end{bmatrix} \qquad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2.437\text{e}{-8} & 1.266\text{e}{-7} & -2.365\text{e}{-7} \\ & 3.596\text{e}{-5} & -5.944\text{e}{-5} \\ & & 4.232\text{e}{-4} \end{bmatrix} \tag{5.1}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 2.115\text{e}{-4} \\ 3.105\text{e}{-2} \\ -8.266\text{e}{-3} \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2.372\text{e}{-8} & -7.961\text{e}{-7} & 1.277\text{e}{-6} \\ & 2.9\text{e}{-5} & -4.411\text{e}{-5} \\ & & 6.949\text{e}{-5} \end{bmatrix} \tag{5.2}$$

and $\lambda = 0.9908$. By applying the backward algorithm enunciated in Theorem 2.2.1, we obtained 103 allocation maps; some of which are reported in Figure 5.2.

Let us now take the time to analyze the kind of investment strategy these maps imply. At the beginning of the investment ($k = 0$), the optimal strategy prescribes that 25% of investor's wealth be invested in Bond and 75% in Equity. After 25 weeks, depending on the realization of portfolio value (x-axis in Figure 5.2), the optimal strategy tells us to allocate wealth as follows: if the portfolio is underperforming (e.g. its value is approximately below 1.029), the optimal allocation is a mix of Equity and Bond, which is the riskiest mix allowed (a 100% allocation in Equity is not permitted due to the *risk* constraint). As soon as performance gets better (i.e. from 1.029 to 1.16) the Equity weight starts decreasing in favor of more Bond and from a certain point on, also Cash. When the portfolio is doing well (i.e. above 1.16), the whole wealth is invested in Cash, namely the least risky of the three asset classes. We could synthesize this behavior by saying that risky positions are taken when the portfolio is doing poorly, whereas conservative positions are taken when the portfolio is doing well. This kind of investing strategy is known in the literature with the name of *contrarian* strategy. The name stems from the fact that contrarian investors bet against the prevailing market trend, namely they try to sell "high" and buy "low". Contrarian

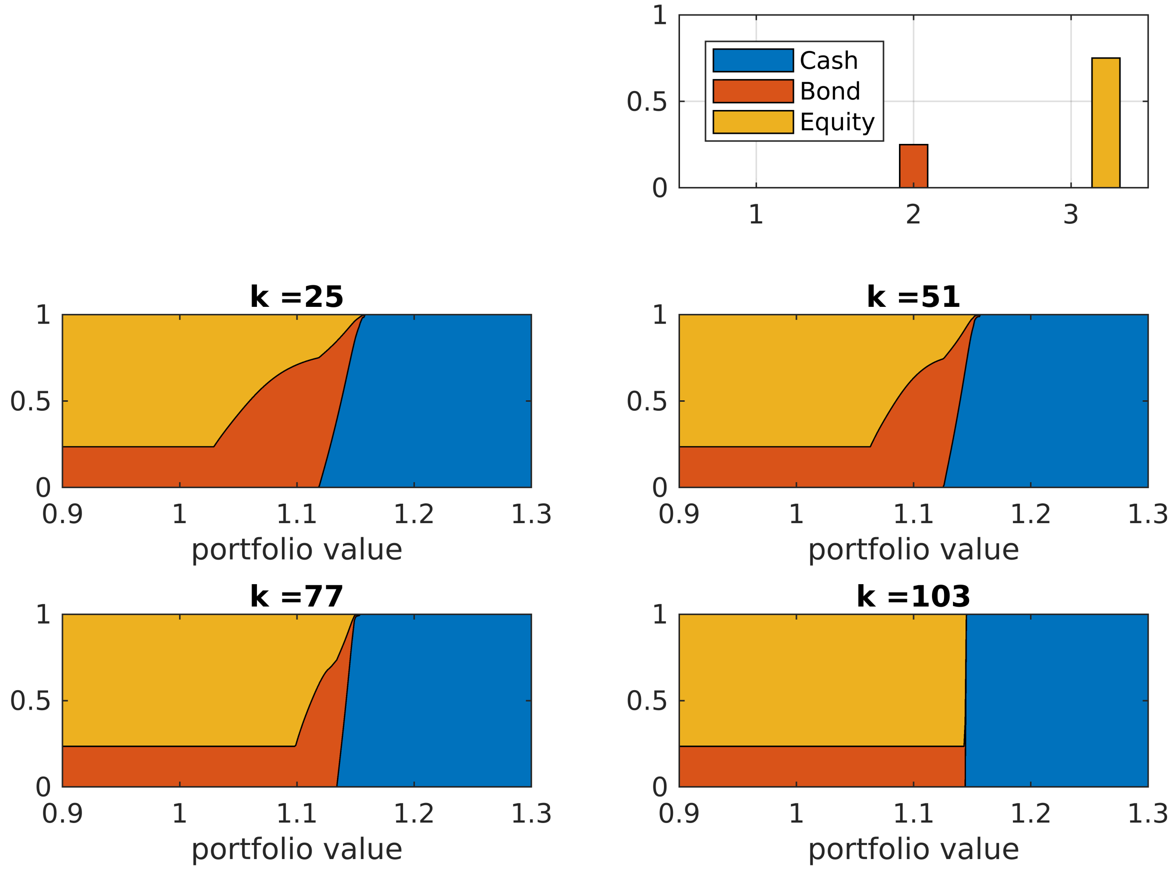

Figure 5.2: Optimal allocation maps, weekly rebalancing, GM model

	G			GM			NIG		
	wk	m	q	wk	m	q	wk	m	q
$p^\star$	79.67%	75.56%	73.26%	78.59%	73.20%	69.44%	78.53%	73.24%	69.47%
p_{MC}	79.77%	75.56%	73.28%	78.82%	73.21%	69.58%	78.76%	73.30%	69.32%
time[h]	0.712	0.157	0.050	0.857	0.316	0.283	6.131	1.467	0.371

Table 5.3: Probability of reaching the target set obtained via ODAA algorithm ($p^\star$) and Monte-Carlo simulation (p_{MC}) for the Gaussian, GM and NIG model and different rebalancing frequencies (weekly, monthly and quarterly). Time is the computational time of the ODAA algorithm, in hours.

strategies perform well in volatile markets and poorly in trending market due to their **concave** nature (see [24]). The optimal strategy obtained by the ODAA algorithm exhibits the same pattern also at successive rebalancing times, the only difference is that it becomes more extreme while approaching the investment end; for instance, at time $k = 103$ there is no transition from the riskiest allocation to the least risky one. In this case, intermediate position are discarded since either the target has already been reached (hence a 100% Cash position) or it has not (hence the riskiest position).

The joint probability of reaching investor's goal is $J(x_0) = p^\star = 78.72\%$. This result is verified by running a Monte-Carlo simulation with 10^5 draws at each rebalancing period from a GM distribution with parameters (5.1) and (5.2). The joint probability obtained is $p_{MC} = 78.73\%$. Another interesting feature of the ODAA strategy is that $p^\star$ increases as the rebalancing frequency decreases. By looking at Table 5.3, it can be seen that as we move from a quarterly rebalancing frequancy[3] to a monthly one the optimal probability goes from 69.44% to 73.20%, and the same happens from monthly to weekly. This fact is rather intuitive since the more rebalancings the more chances to steer the portfolio within the target sets. It should be noted however, that in practice transaction costs have not a negligible impact on portfolio profitability when rebalancing is frequent. This is the reason why investment policies that update portfolio weights only when an "event" occurs are particularly appealing (they will be treated in Part II).

Next, we used also the Gaussian and the NIG model to describe the

[3]the problem of switching from a rebalancing frequency to another has been tackled as follows: the model is calibrated to weekly data, then linear returns are approximated by log-returns enabling us to write additive relations such as $w_{monthly} = w_{wk1} + \ldots + w_{wk4}$. Finally, using the hypothesis of iid returns we analytically derive the distribution of monthly and quarterly returns for the G, GM and NIG model. All this models are closed under convolution.

	G	GM	NIG
$\log L^\star$	4396.2	4471.0	4450.2

Table 5.4: Log-likelihood for G, GM and NIG model

asset class returns distribution. From Table 5.4 we see that the best fitting is provided by the GM model since it exhibits the highest log-likelihood function value; nonetheless, the NIG comes right after it. It is not surprising that the Gaussian model is ranked last as we were well-aware that the data considered deviates from a multivariate Gaussian sample (see Table 5.2).

5.2.1 ODAA vs CPPI vs Constant-Mix

Within the class of asset allocation strategies, the CPPI and the Constant-Mix are among the most popular ones (see [24]). After briefly discussing how they work, we will compare their performance to the ODAA's one.

CPPI The idea behind the CPPI is to maintain the portfolio **exposure** to the risky asset, E_k, equal to a constant multiple m of the portfolio **cushion**, C_k. The risky asset is assume to be a mix of Bond and Equity. The cushion at time k is defined as

$$C_k = \max\left\{x_k - F_k, 0\right\}$$

where x_k is the portfolio value at time k and F_k is the so-called portfolio **floor**. The floor is a value below which the investor does not want the portfolio value to fall. In our case, the floor is a risk-free asset which grows deterministically at the Cash rate. Therefore, once the investor has specified

- an initial allocation $\boldsymbol{u}_0$
- the initial floor F_0
- a cushion multiplier m
- the maximum value-at-risk ($V@R_{1-\alpha}$) according to his risk profile,

he can synthesized the CPPI strategy as follows

$$
\begin{aligned}
\underset{\boldsymbol{u}_k}{\text{maximize}} \quad & A\boldsymbol{u}_k \\
\text{subject to} \quad & \boldsymbol{u}_k^T \mathbf{1} = 1, \\
& u_i \geq 0 \qquad \forall i \in \{1,2,3\}, \\
& \boldsymbol{u}_k \boldsymbol{\Lambda} \boldsymbol{u}_k \leq \sigma_{max}^2, \\
& \underbrace{xA\boldsymbol{u}_k}_{E_k} \leq mC_k.
\end{aligned}
$$

where $A = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$, $\sigma_{max} = \frac{V@R_{1-\alpha}}{z_{1-\alpha}}$, $x \in X_k$ and $k = 1, \dots, N$. From this formulation we see that the investor aims at maximizing the allocation in the risky asset (matrix A selects the allocation in Bond and Equity) while keeping under control the riskiness of the overall allocation and limiting the risky exposure to m times the cushion. The covariance matrix $\boldsymbol{\Lambda}$ depends on the model chosen to describe the asset class returns distribution. In our analysis, we set $m = 6$, $\boldsymbol{u}_0$ equal to the initial ODAA allocation, F_0 is chosen in such a way that also the initial exposure is m times the initial cushion and the asset class return random vector follows a GM distribution with parameters (5.1) and (5.2). The others investment parameters are equal to the ones in the ODAA case. The CPPI allocation maps are reported in Figure 5.3.

Constant-Mix Following a Constant-Mix strategy means maintaining an exposure to the risky asset that is a constant proportion of wealth. For instance, suppose one decides to keep a 60/40 proportion between risky and risk-free asset. After a rebalancing time, asset prices change causing the portfolio proportion to change as well. Let us suppose that the risky asset has increased its price while the risk-free has fall. At the next rebalancing time, the Constant-Mix policy prescribes to sell shares of the risky asset and buy shares of the risk-free in order to recover the 60/40 mix. The constant mix is chosen by solving the following equivalent formulation of the Markowitz problem

$$
\begin{aligned}
\underset{\boldsymbol{u}}{\text{maximize}} \quad & \boldsymbol{u}^T \boldsymbol{\mu} \\
\text{subject to} \quad & \boldsymbol{u}^T \mathbf{1} = 1, \\
& u_i \geq 0 \qquad \forall i \in \{1,2,3\}, \\
& \boldsymbol{u} \boldsymbol{\Lambda} \boldsymbol{u} \leq \sigma_{max}^2.
\end{aligned}
$$

where $\sigma_{max} = \frac{V@R_{1-\alpha}}{z_{1-\alpha}}$ and $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ are the mean and the covariance matrix of the random vector $\boldsymbol{w}_{k+1}$ which represents the asset class returns. The

Statistic	ODAA	CPPI	Constant-mix
p_{MC}	78.93%	52.70%	61.41%
Mean Return (ann)	5.82%	7.55%	10.05%
Volatility (ann)	5.00%	7.93%	13.21%
Median (ann)	7.20%	7.35%	9.41%
Skewnwss	0.700	-0.010	0.0146
Kurtosis	6.700	3.052	3.012
Monthly $V@R_{0.95}$	5.49%	5.51%	9.42%
Max Drawdown	43.39%	25.77%	45.80%
Mean Drawdown	1.97%	2.12%	4.20%
Sharpe ratio	1.163	0.951	0.760

Table 5.5: Investment performance for strategies ODAA, CPPI and Constant-mix obtained via Monte-Carlo simulation (2×10^5 replications).

distribution of $\boldsymbol{w}_{k+1}$ could be any of the ones discussed in Chapter 3. The optimization problem has been solves assuming a GM distribution with parameters (5.1) and (5.2); the optimal constant mix is

$$\boldsymbol{u}^\star = \begin{bmatrix} 0 & 0.2352 & 0.7648 \end{bmatrix}^T .$$

The ODAA and Constant-Mix belongs to the **concave** allocations strategies (see [25]). This means that their policy is to buy risky assets (Equity and Bond) when they fall and sell them when they raise. Concave strategies perform well in oscillating markets. Conversely, the CPPI is an example of a **convex** strategy: risky assets are bought when they raise and sold when they fall. This behavior is clear from the allocation maps in Figure 5.3. When portfolio performance is up, a more risky position is taken, when it is down, risky assets are sold and a more covered position is taken. Convex strategies perform well in trending markets.

In order to compare the performance of this three different strategies we ran a Monte-Carlo simulation. Starting from an initial wealth $x_0 = 1$, 2×10^5 portfolio trajectories are drawn assuming a GM distribution with parameters (5.1) and (5.2) for vectors $\boldsymbol{w}_{k+1}$. Performance and risk figures are reported in table 5.5. The empirical density function of the 2-year investment return is reported in Figure 5.4.

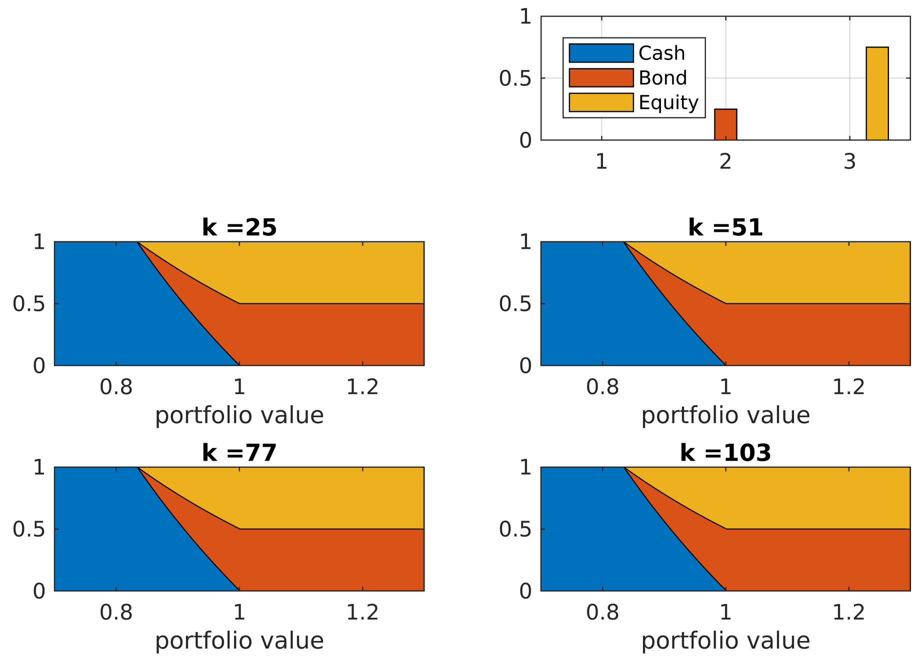

Figure 5.3: Optimal CPPI allocation maps, weekly rebalancing, GM model. An example of a convex strategy.

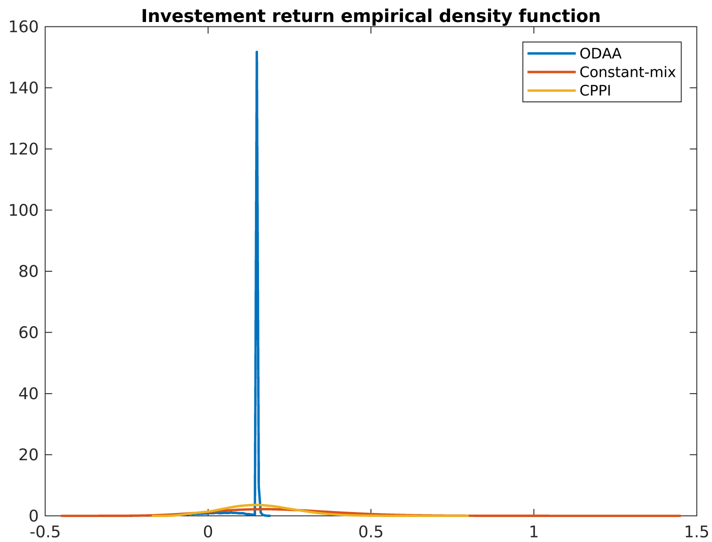

Figure 5.4: Empirical density functions of the 2-year return for the ODAA, CPPI and Constant-mix strategy.

Part II

Event-Driven approach

Chapter 6

Discrete Event Systems and Asset Allocation

From this chapter on, we will adopt a different approach to the asset allocation problem: time will no longer be the driver of portfolio rebalancings but instead, portfolio weights will be updated whenever a predefined **event** occurs. This new point of view stems from the fact that when developing the stochastic reachability approach in Chapter 2, we let the system dynamics be indexed by an independent variable $k \in \mathbb{N}$, which we interpreted as discrete time but, as a matter of fact, the theory did not rely on this particular interpretation. This gives us the freedom to think of k as an abstract index (for instance, it could be an event counter). This observation is the basis for embedding the asset allocation problem in a Discrete Event System (DES) environment.

In this chapter we will present the basics of DES modeling (Section 6.1) which are essential for discussing the Event-Driven (ED) approach to asset allocation (Section 6.2). As far as the Discrete Event Systems are concerned, the main reference is the rich monograph [13], whereas the ED asset allocation model is taken from [29].

6.1 Introduction to DES

From a Control System point of view, the dynamical system introduced in (2.2) can be classified as *continuous-state* (the state space $\mathcal{X}$ is a proper subset of $\mathbb{R}$) and *discrete-time* ($k \in \mathbb{N}$). Informally, if the state space is a discrete set and state transitions are observed whenever an "event" occurs we will talk about Discrete Event Systems. Systems considered so far are time-driven, in that we could imagine the systems being synchronized to a clock and at every clock tick an event e is drawn from an event space E, causing the state to change. However, we could think of a different mechanism governing the state transition: at various time instants (not necessarily known in advanced), some event $e \in E$ occurs, making the state change. The following example will hopefully clarify the difference between Time-Driven (TD) and ED systems. Imagine a particle bound to move on a plane; at every tick of a clock an event is drawn and the particle is allowed to move by a unit step northward, southward, eastward or westward. In this case we have a TD system whose events are $e_1 =$ "one step North", $e_2 =$ "one step South" and so on. On the other hand, suppose there are four players, each of them capable of making the particle move in his direction by issuing a signal. A player issues a signal at random times. The resulting system is ED since it is not synchronized to any clock and state transitions are caused by event like $e_k =$"player 1 issued a signal". The formal definition of a DES reads as follows

Definition **6.1.1** (Discrete Event System)**:** A **Discrete Event System** is a *discrete-state*, *event-driven* system, that is, its state evolution depends entirely on the occurrence of asynchronous discrete events over time.

A DES can be studied from three different levels of abstraction. We will present them in increasing order of complexity

1. *untimed*: the interest is on the sequences of events that the system could execute, without any time information. For instance, an untimed sequence could be
$$e_1, e_2, e_3, e_4, e_5, e_6$$
2. *timed*: in this representation each possible event sequence is coupled with time information, that is, not only the occurrence order is given but also the exact time instant when an event happened. For example, in the *timed* setting, the following could be a system sample path
$$(e_1, t_1), (e_2, t_2), \ldots, (e_6, t_6)$$
3. *stochastic timed*: it is the most detailed description of a DES since it contains event information on all possible orderings, time information about the exact instant at which the event occurs and also statistical information about successive occurrences.

As our modeling purposes required the stochastic timed level of abstraction, we now give the definition of a *Stochastic Clock Structure*, which is the tool used to include time and statistical information to a sequence of events.

***Definition* 6.1.2** (Stochastic Clock Structure)**:** The **Stochastic Clock Structure** associated with an event set E is a set of CDFs

$$\boldsymbol{G} = \{G_i \colon i \in E\}$$

characterizing the stochastic clock sequences

$$\boldsymbol{V}_i = \{V_{i,1}, V_{i,2}, \ldots\} \qquad i \in E$$

where $V_{i,k}$ is a random variable indicating the kth occurrence time of event e_i.

***Remark* 6.1.1:** Sometimes, instead of modeling the exact time when an event occurs, as in the definition above, it is more convenient to model the elapsed time between two events, the so-colled **interevent time**. In this case we will write the Stochastic Clock Sequence in the following way

$$\boldsymbol{T}_i = \{T_{i,1}, T_{i,2}, \ldots\} \qquad i \in E.$$

If a deterministic clock sequence $\boldsymbol{v}_i = \{v_{i,1}, v_{i,2}, \ldots\}$ is given for each event in E, we will talk about a *Clock Structure*. The evolution of a DES needs to be described by a state equation of the form

$$x_{k+1} = f(x_k, e_{k+1}) \qquad k \in \mathbb{N} \tag{6.1}$$

where x_k is the current state and x_{k+1} the state once the event e_{k+1} has occurred. The above recursive equation is the event-driven equivalent of Equation (2.2). However, Equation (6.1) describes only the untimed dynamics, that is no time information is included. Conversely, in asset allocation applications, we are also interested in *when* an event occurs. For this reason, after introducing a Clock Structure $\boldsymbol{v} = \{v_i \colon i \in E\}$ associated with a finite event set $E = \{e_1, \ldots, e_n\}$, we seek a relationship of the form

$$e_{k+1} = h(x_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$$

so that we could replace (6.1) with

$$\begin{cases} x_{k+1} &= f(x_k, e_{k+1}) \\ e_{k+1} &= h(x_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n). \end{cases} \tag{6.2}$$

Equations (6.2) capture the *timed* dynamics of a DES.

As it was mentioned earlier, the Stochastic timed behavior is what interests us; therefore, we conclude this section by giving the definition of a Stochastic Timed Automaton (see [13] for a more detailed treatment)

***Definition* 6.1.3** (Stochastic Timed Automaton)**:** A **Stochastic Timed Automaton** is a six-tuple

$$(\mathcal{E}, \mathcal{X}, \Gamma, p, p_0, \boldsymbol{G})$$

where

$\mathcal{E}$ is a countable event set

$\mathcal{X}$ is a countable state space

$\Gamma(x)$ is the set of feasible events, defined $\forall x \in \mathcal{X}$

$p(x'; x, e')$ is the transition probability from state x to state x' given the occurrence of event e'

$p_0(x)$ is the pmf of the initial state X_0 (which is a random variable)

$\boldsymbol{G} = \{\boldsymbol{T}_i \colon i \in \mathcal{E}\}$ is a Stochastic Clock Time of interevent times.

A Stochastic Timed Automaton, together with the dynamics in Equation (6.2) (where the Clock Structure $\boldsymbol{v}$ is replaced by a Stochastic one $\boldsymbol{V}$) gives the most complete description of a DES. We now move to the asset allocation application.

6.2 Event-Driven Asset Allocation

In this section we present the first event-driven model, having the objective to invest in the derivative market. In fact, we will consider a market consisting of a **risky asset** (a future index) and a **risk-free asset** (a bank account). The event-driven approach aims at modeling the industrial practice of rebalancing the portfolio weights whenever an "event" occurs. In the following, we suppose that an event has occurred every time the absolute value of the risky asset cumulative return hits a threshold (e.g. 7%). This policy could be beneficial in different aspects:

1. in low-volatile markets, when the risky asset price is quite steady, portfolio weights need not to be updated at predefined time instants but only when the market conditions have significantly changed. This cuts down on transaction costs.

2. in high-volatile markets, when the risky asset price repetitively increases and plummets in a short period of time (shorter than the rebalancing frequency), the event-driven policy can swiftly intervene

by changing the portfolio exposure without having to wait the next rebalancing time (which could be after the loss has already become substantial).

The main goal of this section is first to derive a proper event-driven portfolio value dynamics and then find its density function, which will be plugged in the ODAA algorithm (2.2.1). Let us now start off by investigating both the time-driven dynamics ,which will allow us to estimate how long the investment is going to last and the event-driven dynamics.

6.2.1 Time-Driven dynamics

A portfolio rebalancing is performed every time the absolute value of the risky asset cumulative return, starting from an initial baseline, hits a threshold. Let J be this threshold. We suppose the following time-driven discrete dynamics for the risky asset

$$S_{k+1} = S_k(1 + JN_{k+1}^{\Delta t}) \qquad k \in \mathbb{N}. \tag{6.3}$$

The random variable $N_{k+1}^{\Delta t}$ (which is the $(k+1)$th element of a sequence of iid random variables) takes values in the discrete set $\{1, 0, -1\}$ and indicates whether the discrete price process has a positive, negative or null jump at the end of a time interval of length Δt. If it takes value 1, the discrete price process $\{S_k\}$ experiences a positive jump at the end of time period $[t_k, t_{k+1}]$, if the value is -1 then the jump is negative and if the value is 0, the process has no jumps in this time interval. This is the same as saying that when the random variable takes the value 1 then the risky asset return, over the period $[t_k, t_{k+1}]$, is greater than J, when the value is -1 then the cumulative return is smaller than -J and when the value is 0, than it belongs to the interval $[-J, J]$. The superscript Δt indicates the length of the interval $[t_k, t_{k+1}]$. The next step is to find a proper distribution for $N_{k+1}^{\Delta t}$. Let the probability mass function (pmf) of this random variable have the following form

$$f_{N_{k+1}^{\Delta t}}(y) = \begin{cases} \exp\{-\lambda\Delta t\} & \text{if } y = 0 \\ \big(1 - \exp\{-\lambda\Delta t\}\big)p & \text{if } y = 1 \\ \big(1 - \exp\{-\lambda\Delta t\}\big)(1 - p) & \text{if } y = -1 \end{cases} \tag{6.4}$$

where $\lambda \in \mathbb{R}^+$ and $p \in [0, 1]$. This functional form is particularly convenient since it implies an exponential distribution for the interevent times (also called **holding times** in a financial context). This fact is synthesized in the following proposition

***Proposition* 6.2.1:** Given the time-driven dynamics of the risky asset in (6.3) and the pmf (6.4) of random variable $N_{k+1}^{\Delta t}$, let τ_{k+1} be the random variable indicating the holding time between the kth and the $(k+1)$th event. Then $\tau_{k+1} \sim \exp(\lambda)$

Proof. Let t_k be a realization of the random variable V_k, which is an element of a Stochastic Clock Sequence. It indicates when the kth event occurs. From the definition of a Stochastic Clock Structure we have

$$G_{k+1}(t) = \mathbb{P}\big(\tau_{k+1} \leq t\big) = 1 - \mathbb{P}\big(\tau_{k+1} > t\big)$$

but

$$\begin{aligned} \mathbb{P}\big(\tau_{k+1} > t | T_k = t_k\big) &= \mathbb{P}\big(N_{k+1}^{(t_k+t)-t_k} = 0\big) \\ &= \mathbb{P}\big(N_{k+1}^{t} = 0\big) \\ &= \exp\{-\lambda t\} \end{aligned}$$

therefore $\mathbb{P}\big(\tau_{k+1} > t | T_k = t_k\big)$ is independent from t_k. Hence

$$\mathbb{P}\big(\tau_{k+1} > t | T_k = t_k\big) = \mathbb{P}\big(\tau_{k+1} > t\big) = \exp\{-\lambda t\}$$

which implies $G_{k+1}(t) = 1 - \exp\{-\lambda t\}$. Since this is the cdf of an exponential random variable, we have the result. □

***Remark* 6.2.1:** Given that $\mathbb{E}[\tau_{k+1}] = \frac{1}{\lambda}$, the parameter λ acquires the meaning of speed of the price process. The larger λ, the more frequent portfolio rebalancings are executed.

6.2.2 Event-Driven dynamics

Dynamics (6.3) is still time-driven since the independent variable $k \in \mathbb{N}$ represents discrete time. Instead, in the event-driven framework, we let k indicate the number of events (portfolio rebalancings or trades). In this context, S_{k+1} is the risky asset price after that the $(k+1)$th portfolio rebalancing or trade is performed. The event-driven dynamics of the risky asset reads as follows

$$S_{k+1} = S_k(1 + J\widetilde{N}_{k+1}) \qquad k \in \mathbb{N} \tag{6.5}$$

where $\widetilde{N}_{k+1}$ is distributed according to

$$f_{\widetilde{N}_{k+1}}(y) = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{if } y = -1 \end{cases} \tag{6.6}$$

Let us understand how the pmf (6.6) follows from (6.4). First of all, the random variable $\widetilde{N}_{k+1}$ is Bernoullian. In fact, it models whether the next jump of the risky asset process is positive or negative. Therefore, we are left to compute the probability that the jump is positive (the parameter q of the Bernoulli distribution). By applying the Bayes Theorem, the Law of Total Probability in the continuous case and the fact that τ_{k+1} is exponential with parameter λ, we obtain

$$\begin{aligned}
q &= \mathbb{P}\big(\widetilde{N}_{k+1} = 1\big) = \mathbb{P}\big(N_{k+1}^{\tau_{k+1}} = 1 | (N_{k+1}^{\tau_{k+1}} = 0)^C\big) \\
&= \frac{\mathbb{P}\big(N_{k+1}^{\tau_{k+1}} = 1, (N_{k+1}^{\tau_{k+1}} = 0)^C\big)}{\mathbb{P}\big((N_{k+1}^{\tau_{k+1}} = 0)^C\big)} \\
&= \frac{\mathbb{P}\big(N_{k+1}^{\tau_{k+1}} = 1\big)}{1 - \mathbb{P}\big(N_{k+1}^{\tau_{k+1}} = 0\big)} \\
&= \frac{\int_0^\infty \mathbb{P}\big(N_{k+1}^{\tau_{k+1}} = 1 | \tau_{k+1} = t\big) f_{\tau_{k+1}}(t) \mathrm{d}t}{1 - \int_0^\infty \mathbb{P}\big(N_{k+1}^{\tau_{k+1}} = 0 | \tau_{k+1} = t\big) f_{\tau_{k+1}}(t) \mathrm{d}t} \\
&= \frac{\int_0^\infty (1 - e^{-\lambda t}) p \lambda e^{-\lambda t} \mathrm{d}t}{1 - \int_0^\infty e^{-\lambda t} \lambda e^{-\lambda t} \mathrm{d}t} \\
&= p.
\end{aligned}$$

Parameter p governs the trend of the discrete price process. The greater p, the more likely it is to have positive jumps.

6.2.3 Portfolio dynamics

We recall that the portfolio we are considering consists of a risky and a risk-free asset. The event-driven dynamics of the former has been given in (6.5). In this section, the event-driven dynamics of portfolio value will be derived. For the sake of simplicity, we assume that the risk-free asset evolves in a deterministic way with constant interest rate r (continuously compounded). Throughout this section, let us fix two time instants, t_k and t_{k+1}, which are realizations of random variables V_k and V_{k+1}. These random variables indicate the time when the kth and $(k+1)$th trade take place (or, in other words, when the kth and $(k+1)$th event occur).

In general, the event-driven portfolio dynamics is

$$x_{k+1} = x_k(1 + u_k^C w_{k+1}^C + u_k^S w_{k+1}^S) \qquad k \in \mathbb{N} \tag{6.7}$$

where u_k^C, u_k^S are the portfolio weights of the risk-free and risky asset respectively, w_{k+1}^C and w_{k+1}^S their return over the period $[t_k, t_{k+1}]$. It is important to remark that the length of the time interval $[t_k, t_{k+1}]$ is not deterministic, but it is a random variable exponentially distributed (see Proposition 6.2.1) and denoted by τ_{k+1}. Consequently, w_{k+1}^C and w_{k+1}^S are returns over this stochastic time period. From (6.5) we easily get

$$w_{k+1}^S = J\widetilde{N}_{k+1} \tag{6.8}$$

As far as the risk-free asset in concerned, denoting by C_{k+1} its price after the $(k+1)$th trade, we have

$$\begin{aligned} C_{k+1} &= C_k(1 + w_{k+1}^C) = C_k \exp\{r\tau_{k+1}\} \\ \Longrightarrow \quad w_{k+1}^C &= \exp\{r\tau_{k+1}\} - 1 \end{aligned} \tag{6.9}$$

where r is the deterministic interest rate of the risk-free asset, continuously compounded. By plugging (6.8) and (6.9) into (6.7) the portfolio dynamics becomes

$$\boxed{x_{k+1} = x_k(\exp\{r\tau_{k+1}\} + u_k J\widetilde{N}_{k+1})} \qquad k \in \mathbb{N} \tag{6.10}$$

where we dropped the superscript S from u_k^S and set $u_k^C = 1$. This reflects what is usually done in the derivative trading practice, namely keeping a 100% cash position plus a long or short exposure to the derivative. Consequently, the weight u_k is allowed to take values in the compact set $[-1, 1]$.

6.2.4 The density of x_{k+1}

In order to apply the ODAA algorithm (see Theorem 2.2.1), the explicit form of the density of the random variable x_{k+1} is required. The result is given in the following proposition

***Proposition* 6.2.2:** The probability density function of random variable (6.10) (where x_k has been fixed to $x \in \mathcal{X}$) is

$$f_{x_{k+1}}(z) = \frac{\lambda}{rx}\Big\{p\Big(\frac{z-\xi}{x}\Big)^{-\left(\frac{\lambda+r}{r}\right)}\mathbb{1}_{[x+\xi,\infty)} + (1-p)\Big(\frac{z+\xi}{x}\Big)^{-\left(\frac{\lambda+r}{r}\right)}\mathbb{1}_{[x-\xi,\infty)}\Big\}$$

where $\xi = xJu_{k+1}$ and $z \in \mathbb{R}$.

Proof. Let $F_{\tau_{k+1}}(t) = (1 - e^{-\lambda t})\mathbb{1}_{[0,\infty)}(t)$ be the cdf of τ_{k+1}. The first step consists in finding the cdf F_Y of $Y = x\exp\{r\tau_{k+1}\}$. By simple calculations we obtain

$$F_Y(y) = \Big(1 - \Big(\frac{y}{x}\Big)^{-\frac{\lambda}{r}}\Big)\mathbb{1}_{[x,\infty)}.$$

Let us rewrite the portfolio value at the $(k+1)$th trade in the following way

$$x_{k+1} = Y + xu_k J\widetilde{N}_{k+1} = Y + \xi\widetilde{N}_{k+1}$$

and by using the Law of Total Probability we have

$$\begin{aligned}
F_{x_{k+1}}(z) &= \mathbb{P}\big(Y + \xi\widetilde{N}_{k+1} \leq z\big) \\
&= \mathbb{P}\big(Y + \xi\widetilde{N}_{k+1} \leq z|\widetilde{N}_{k+1} = 1\big)\mathbb{P}\big(\widetilde{N}_{k+1} = 1\big) + \\
&\quad + \mathbb{P}\big(Y + \xi\widetilde{N}_{k+1} \leq z|\widetilde{N}_{k+1} = -1\big)\mathbb{P}\big(\widetilde{N}_{k+1} = -1\big) \\
&= F_Y(z-\xi)p + F_Y(z+\xi)(1-p) \\
&= p\Big\{1 - \Big(\frac{z-\xi}{x}\Big)^{-\lambda/r}\Big\}\mathbb{1}_{[x+\xi,\infty)} + (1-p)\Big\{1 - \Big(\frac{z+\xi}{x}\Big)^{-\lambda/r}\Big\}\mathbb{1}_{[x-\xi,\infty)}
\end{aligned}$$

Differentiating the cdf with respect to z gives us the result. □

6.3 The calibration of p and λ

In this section we calibrate the parameters p and λ of the pmf (6.4) to market data. We recall that parameter p is responsible of the trend of process $\{S_k\}$ whereas λ controls the jump frequency. The time series we are considering is the daily Future S&P 500 from 22 January 2010 to 25 April 2016. In order to find estimates $\widehat{p}$ and $\widehat{\lambda}$ we need to extract a sample of realizations of random variable $N_{k+1}^{\Delta t}$ from the above time series. This could be accomplished by applying Algorithm 3 to the data. Indeed, the algorithm outputs the discrete time series (reported in red in Figure 6.1) and the logical sample $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ (denoted by $\{d_{t_k}\}_{k=1,\ldots,n}$ in the algorithm).

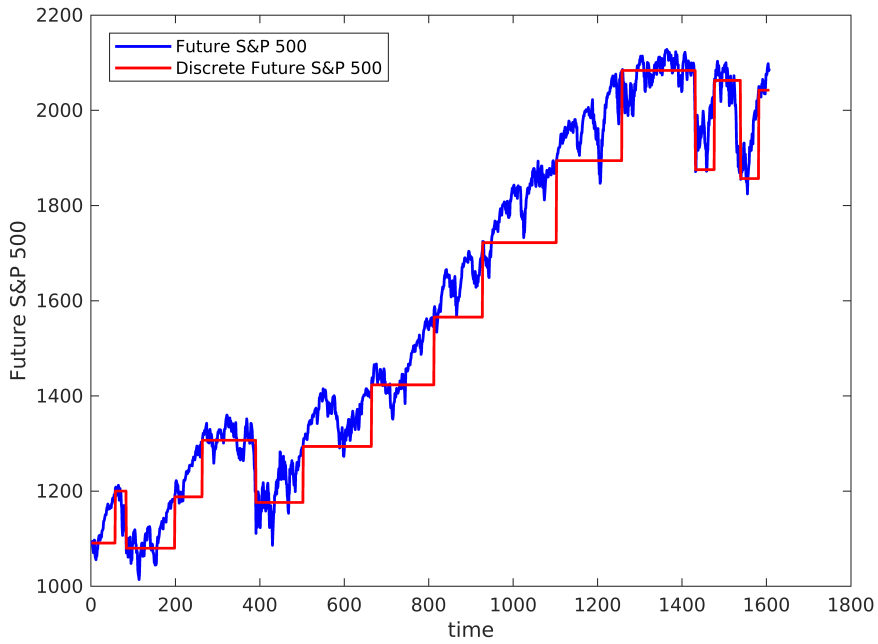

Figure 6.1: Future S&P 500 and its discrete time series. The discrete time series was obtained with a jump size $J = 7\%$

Algorithm 3 Discrete price and logical time series

Input: price time series $\{S_{t_k}\}_{k=0,\ldots,n}$, jump size J

Output: discrete price time series $\{D_{t_k}\}_{k=0,\ldots,n}$, logical time series $\{d_{t_k}\}_{k=1,\ldots,n}$

initialization: $Baseline = S_{t_0}$, $D_{t_0} = S_{t_0}$

for $i = 1, \ldots, n$ **do**

  $R = S_{t_i}/Baseline - 1$

  if $abs(R) > J$ **then**

    $Baseline = Baseline(1 + sign(r)J)$

    $D_{t_i} = Baseline$

    $d_{t_i} = sign(r)$

  else

    $D_{t_i} = D_{t_{i-1}}$

    $d_{t_i} = 0$

  end

end

As estimation method we used maximum likelihood. The likelihood function can be written as follows

$$
\begin{aligned}
L(\lambda, p; \mathbf{y}) &= \prod_{i=1}^{n} f_{N_{k+1}^{\Delta t}}(y_i; \lambda, p) \\
&= \Big(\prod_{i\colon y_i=0} \exp(-\lambda\Delta t) \Big) \Big(\prod_{i\colon y_i=1} \big(1-\exp(-\lambda\Delta t)\big)p \Big) \Big(\prod_{i\colon y_i=-1} \big(1-\exp(-\lambda\Delta t)\big)(1-p) \Big) \\
&= \Big(\exp(-\lambda\Delta t) \Big)^{\alpha} \Big(\big(1-\exp(-\lambda\Delta t)\big)p \Big)^{\beta} \Big(\big(1-\exp(-\lambda\Delta t)\big)(1-p) \Big)^{\gamma}
\end{aligned}
$$

where con

$$
\begin{aligned}
\alpha &= \operatorname{card}\{y_i = 0\colon i = 1, \ldots, n\} \\
\beta &= \operatorname{card}\{y_i = 1\colon i = 1, \ldots, n\} \\
\gamma &= \operatorname{card}\{y_i = -1\colon i = 1, \ldots, n\}
\end{aligned}
$$

and Δt has been set to $1/252$. Imposing the first order optimality condition $\nabla \log(L(\lambda, p; \mathbf{y})) = \mathbf{0}$ and solving with respect to p and λ we obtain the following estimates

$$
\widehat{\lambda} = -\frac{1}{\Delta t} \log\left(\frac{\alpha}{\alpha+\beta+\gamma} \right) \tag{6.11}
$$

$$
\widehat{p} = \frac{\beta}{\beta+\gamma} \tag{6.12}
$$

The point $(\widehat{p}, \widehat{\lambda})$ is actually a maximizer since the Hessian matrix computed in this point is definite negative. Setting $J = 10\%$ and applying the equations above to our data we get the following estimates

$$
\begin{aligned}
\widehat{\lambda} &= 2.363 \\
\widehat{p} &= 0.733.
\end{aligned}
$$

***Remark* 6.3.1:** As $\alpha+\beta+\gamma$ equals n (the sample size), the estimator of λ depends only on the number of interval in which there are no jumps (α). Conversely, the estimator of p depends only on the number of positive and negative jumps, as it was reasonable to expect.

6.4 Numerical Results

Let us consider an investment with the following parameters:

- $N = 10$ portfolio rebalancings

- $J = 10\%$
- $r = 5.5\%$
- The following target sets: $X_0 = \{1\}$, $X_k = [0, \infty)$ for $k = 1, \ldots, 9$ and $X_{10} = [(1 + \theta)^4, \infty)$ with $\theta = 7\%$. In the implementation, they were approximated by $X_k = [0.5, 1.9]$ for $k = 1, \ldots, 9$ and $X_{10} = [(1+\theta)^4, 1.9]$ and discretized with step length 5×10^{-4}.

In Figure 6.2 are reported some allocation maps obtained by running the ODAA algorithm with the investment parameters specified above and model parameters calibrated in the previous section. Each map shows how to allocate the risky asset at portfolio rebalancing number k. The *contrarian* attitude of the strategy is the same as in the time-driven case: when performance is down, the policy prescribes to take a 100% long position in the risky asset. Conversely, when performance is up, a 100% short position on the risky asset in taken. In between, there is a range of long/short exposures which depends on portfolio performance. The joint probability of reaching the target sets is $p^\star = J(x_0) = 86.3\%$. This result is verified by running a Monte Carlo simulation with 10^6 replications at each rebalancing time. Sample statistics regarding this 10-riallocation investment are reported in Table 6.1. Finally, Figure 6.3 shows the empirical density function of the investment total return.

Statistics	value
$p^\star$	0.863
p_{MC}	0.865
Mean Return (ann)	11.45%
Volatility (ann)	10.38%
Median Return (ann)	11.14%
Skewness	-1.55
Kurtosis	5.27
Sharpe Ratio	0 .558
Avg horizon [years]	4.23

Table 6.1: Investment performance obtained via Monte-Carlo simulation with 10^6 replications at each rebalancing time

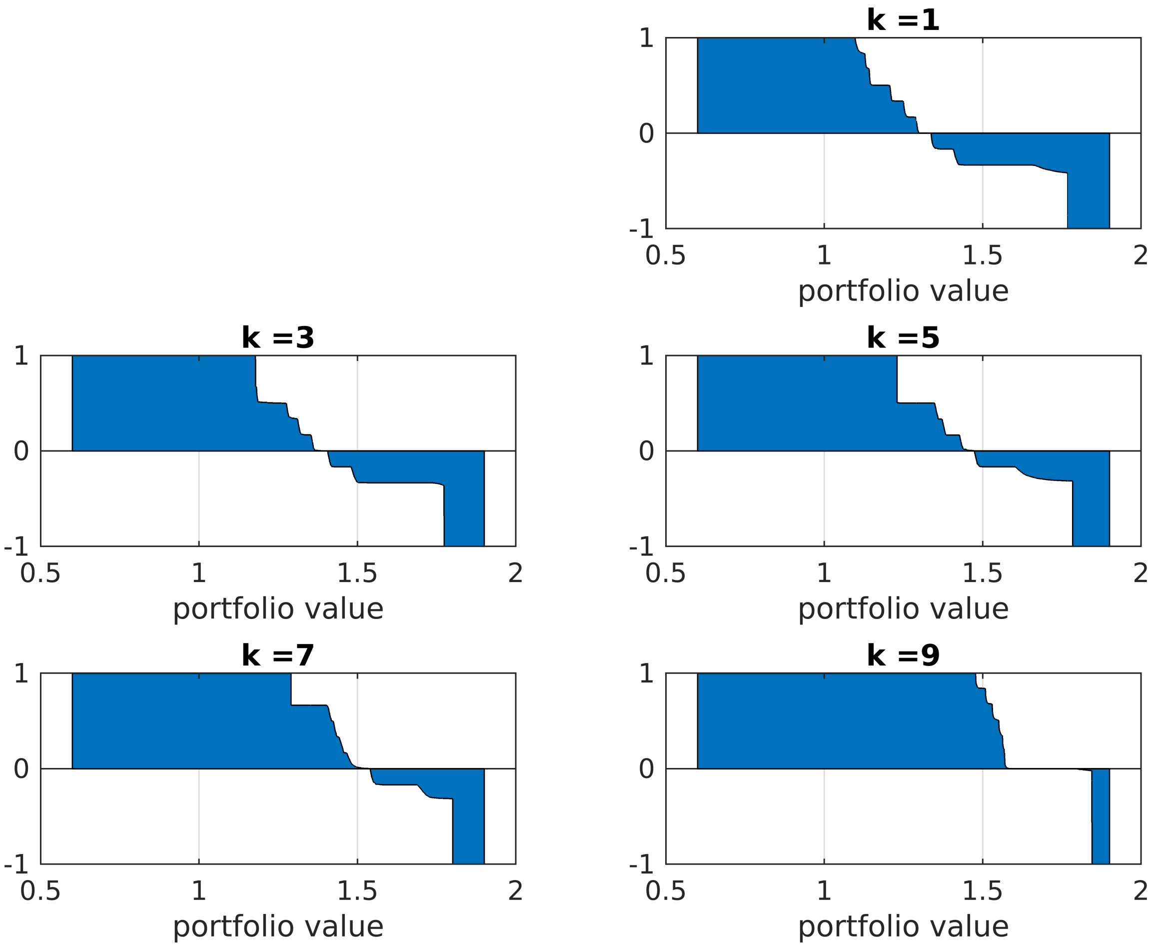

Figure 6.2: Risky asset allocation maps

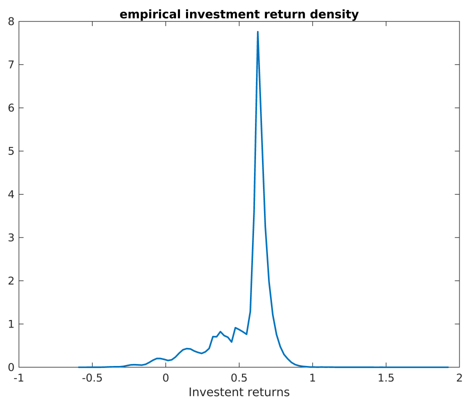

Figure 6.3: Investment return empirical density function

Chapter 7

Model Extensions

The prime objective of this chapter is to extend the event-driven model of Chapter 6 in two different ways. On one hand, in Section 7.1 the risky asset will be modeled in continuous-time as a Geometric Brownian Motion (GBM), on the other hand, in Section 7.2 we no longer let the risk-free asset evolve deterministically but instead its short-rate will evolve according to the Vasicek model.

7.1 A GBM dynamics for the risky asset

Let us consider the same portfolio as the one in Chapter 6, namely a risky and a risk-free asset. As far as the risk-free asset is concerned, the deterministic model of Chapter 6 remains unchanged. On the other hand, the following dynamics for the risky asset is assumed

$$\begin{cases} dS_t = \mu S_t dt + \sigma S_t dW_t \\ S_{t_k} = S_k, \quad t \geq t_k \end{cases} \tag{7.1}$$

where $\{W_t\}_{t \geq t_k}$ is a unidimensional Brownian motion, $\mu \in \mathbb{R}$, $\sigma > 0$ and t_k is the time when the kth event occurs. Solving the above Stochastic Differential Equations (SDE) brings

$$\begin{aligned} S_t &= S_k \exp\left\{(\mu - \sigma^2/2)(t - t_k) + \sigma(W_t - W_{t_k})\right\} \\ &= S_k \exp\left\{\widetilde{\mu}(t - t_k) + \sigma(W_t - W_{t_k})\right\} \end{aligned}$$

where we defined $\widetilde{\mu} = \mu - \sigma^2/2$. The cumulative risky asset log-return (starting from t_k) is denoted by $\{X_t\}_{t \geq t_k}$ and is equal to

$$X_t = \log\left(S_t/S_{t_k}\right) = \widetilde{\mu}(t - t_k) + \sigma(W_t - W_{t_k}). \tag{7.2}$$

Given that in the development of the model it will be more convenient to have the time starting from 0, exploiting the Translation Invariance property of Brownian motion we define the translated log-return process as follows

$$\widetilde{X}_t = X_{t+t_k} = \widetilde{\mu}t + \sigma(W_{t+t_k} - W_{t_k}) = \widetilde{\mu}t + \sigma\widetilde{W}_t \quad t \geq 0. \tag{7.3}$$

In this case t represents the interevent time instead of clock time.

7.1.1 The double exit problem

In the event-driven setting introduced in the previous chapter, what triggers a portfolio rebalancing is the fact that the absolute value of process $\widetilde{X}_t$ exceeds the barrier J. When this happens, we say, in the event-driven jargon, that an event has occurred. Therefore, it is of prime interest modeling the stochastic time instant in which the next event takes place. This could be done by defining the following stopping time

$$\begin{aligned} \tau_{k+1} &= \inf\left\{t \geq 0\colon |\widetilde{X}_t| \geq J\right\} \qquad (7.4) \\ &= \inf\left\{t \geq 0\colon \widetilde{X}_t \notin (-J, J)\right\}. \end{aligned}$$

Given that the density function of x_{k+1} is needed in the ODAA algorithm, we are interested in the distribution of random variable (7.4). This problem is known in literature as *the double exit problem of a Brownian motion with drift*. The central result is given by the following theorem, which covers a more general case in which the upper and lower barrier are different. The theorem is taken from [17] as is.

***Theorem* 7.1.1** (double exit problem)**:** Let $X_t = \mu t + \sigma W_t$ be a Brownian motion with drift, $\mu \in \mathbb{R}$ and $\sigma > 0$. Moreover, assume there are two constant barriers $b < 0 < a$. The distribution of $\tau = \inf\left\{t \geq 0\colon X_t \notin (b, a)\right\}$ is

$$F_\tau(t) = 1 - \left(\exp\left\{\frac{\mu b}{\sigma^2}\right\}K_t^\infty(a) - \exp\left\{\frac{\mu a}{\sigma^2}\right\}K_t^\infty(b)\right)$$

where

$$K_t^N(k) = \frac{\sigma^2\pi}{(a-b)^2}\sum_{n=1}^{N}\frac{n(-1)^{n+1}}{\frac{\mu^2}{2\sigma^2} + \frac{\sigma^2 n^2\pi^2}{2(a-b)^2}}\exp\left\{-\left(\frac{\mu^2}{2\sigma^2} + \frac{\sigma^2 n^2 \pi^2}{2(a-b)^2}\right)t\right\}\sin\left(\frac{n\pi k}{a-b}\right).$$

Applying the theorem to our case ($a = J, b = -J$) we get

$$F_{\tau_{k+1}}(t) = 1 - \left[2\cosh\left(\frac{\widetilde{\mu}J}{\sigma^2}\right)K_t^\infty(J)\right], \tag{7.5}$$

where we used the fact that $K_t^N(k)$ is odd as a function of k.

***Remark* 7.1.1:** Theorem 7.1.1 does not assume the upper and lower barrier to be equal. This generality would allow us to consider a more realistic case in which a portfolio riallocation is triggered, for example, when the cumulative return process is grater than a barrier J_{up} or lower than $-J_{down}$, where $J_{up} > J_{down}$.

7.1.2 Portfolio dynamics and the density of x_{k+1}

Following the same path as in Chapter 6, we are left to compute the event-driven portfolio dynamics and the portfolio value density function. As far as the first issue is concerned, the dynamics can be written as follows

$$\boxed{x_{k+1} = x_k\big(\exp\{r\tau_{k+1}\} + u_k\widetilde{X}_{\tau_{k+1}}\big)} \qquad k \in \mathbb{N} \tag{7.6}$$

where r is the constant risk-free asset return, τ_{k+1} is the stopping time (7.4), u_k is the risky asset portfolio weight and $\widetilde{X}_{\tau_{k+1}}$ is the return process computed at the random time τ_{k+1}. $\widetilde{X}_{\tau_{k+1}}$ can only assume value J or $-J$, therefore it is a Bernoullian random variable. The value of its parameter p is given in the following lemma (which closely follows exercise 5.20, [4])

***Lemma* 7.1.1:** Let $\widetilde{X}_t$ be the return process (7.3) and τ_{k+1} the stopping time (7.4). Then $\widetilde{X}_{\tau_{k+1}} \sim B(p)$ where

$$\begin{aligned} p &= \mathbb{P}\Big(\widetilde{X}_{\tau_{k+1}} = J\Big) \qquad (7.7)\\ &= \frac{1-\exp\{2\widetilde{\mu}J/\sigma^2\}}{\exp\{-2\widetilde{\mu}J/\sigma^2\} - \exp\{2\widetilde{\mu}J/\sigma^2\}}\\ &= \frac{\exp\{2\widetilde{\mu}J/\sigma^2\}-1}{2\sinh(2\widetilde{\mu}J/\sigma^2)}. \end{aligned}$$

Proof. The first step of the proof consists in finding $\xi \in \mathbb{R}\setminus\{0\}$ such that $M = \exp\{\xi\widetilde{X}_t\}$ is a martingale. To this end, we apply Ito's formula ([4], Theorem 8.1) to M_t:

$$\begin{aligned} dM_t &= \xi M_t d\widetilde{X}_t + \frac{1}{2}\xi^2 M_t\sigma^2 dt\\ &= \Big(\widetilde{\mu}\xi + \frac{1}{2}\sigma^2\xi^2\Big)M_t dt + \sigma\xi M_t d\widetilde{W}_t. \qquad (7.8) \end{aligned}$$

If the drift in (7.8) is null then $\{M_t\}_{t\geq 0}$ is a martingale. Therefore, we impose the condition $\widetilde{\mu}\xi + \frac{1}{2}\sigma^2\xi^2 = 0$ which brings $\xi = -2\widetilde{\mu}/\sigma^2$.

The second part of the proof starts by noticing that also the process $\{M_{t\wedge\tau}\}_{t\geq 0}$ is a martingale[1](Proposition 5.6, [4]). Moreover, since $|M_{t\wedge\tau}|\leq J$ for every $t \geq 0$, we can apply the Dominated Convergence Theorem (Proposition 4.2, [4]) in the following way:

$$\begin{aligned}
\mathbb{E}\big[M_\tau\big] &= \mathbb{E}\Big[\lim_{t\to\infty} M_{t\wedge\tau}\Big] = \qquad\qquad \text{(Dominated Conv. Theorem)}\\
&= \lim_{t\to\infty} \underbrace{\mathbb{E}\big[M_{t\wedge\tau}\big]}_{1} =\\
&= 1
\end{aligned}$$

hence

$$\begin{aligned}
\mathbb{E}\big[M_\tau\big] &= \exp\{2\widetilde{\mu}J/\sigma^2\}\mathbb{P}\Big(\widetilde{X}_\tau = -J\Big) + \exp\{-2\widetilde{\mu}J/\sigma^2\}\mathbb{P}\Big(\widetilde{X}_\tau = J\Big)\\
&= \exp\{2\widetilde{\mu}J/\sigma^2\}\Bigg(1 - \underbrace{\mathbb{P}\Big(\widetilde{X}_\tau = J\Big)}_{p}\Bigg) + \exp\{-2\widetilde{\mu}J/\sigma^2\}\underbrace{\mathbb{P}\Big(\widetilde{X}_\tau = J\Big)}_{p}\\
&= 1.
\end{aligned}$$

Finally, solving for p we obtain the result. $\square$

In order to apply the ODAA algorithm we need the probability density function of random variable x_{k+1}. Its explicit form is given in the following proposition.

***Proposition* 7.1.1:** Let x_{k+1} be the random variable (7.6) (where x_k has been fixed to $x \in \mathcal{X}$). Its density function is

$$f_{x_{k+1}}(z) = \frac{2\cosh\left(\frac{\tilde{\mu}J}{\sigma^2}\right)}{rx}\Big[p\Gamma^\infty_{\frac{z-\xi}{x}}(J)\mathbb{1}_{(x+\xi,\infty)} + (1-p)\Gamma^\infty_{\frac{z+\xi}{x}}(J)\mathbb{1}_{(x-\xi,\infty)}\Big] \quad (7.9)$$

where $\xi = xu_kJ$, p is the probability given by Lemma 7.1.1 and

$$\Gamma^\infty_z(J) = \frac{\sigma^2\pi}{4J^2}\sum_{n=1}^{\infty} n(-1)^{n+1}z^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2}+\frac{\sigma^2n^2\pi^2}{8J^2}\right)-1}\sin(\frac{\pi}{2}n) \quad (7.10)$$

Proof. The scheme of the proof is the same as the one in Proposition 6.2.2. Let us rewriting the portfolio dynamics as

$$x_{k+1} = x\exp\{r\tau_{k+1}\} + xu_k\widetilde{X}_{\tau_{k+1}} = Y + xu_k\widetilde{X}_{\tau_{k+1}}.$$

[1]for the sake of clarity, we dropped the subscript $k+1$ from τ_{k+1}

The first step is to find the cdf of Y.Thanks to Theorem 7.1.1 we have

$$F_Y(y) = \mathbb{P}\Big(x\exp\{r\tau_{k+1}\} \le y\Big) = F_{\tau_{k+1}}\Big(\frac{1}{r}\log\big(y/x\big)\Big)$$

$$= \Big\{1 - \Big[2\cosh(\widetilde{\mu}J/\sigma^2)K^{\infty}_{\frac{1}{r}\log(\frac{y}{x})}(J)\Big]\Big\}\mathbb{1}_{[x,\infty)}.$$

By invoking the Law of Total Probability we can write

$$\begin{aligned}
F_{x_{k+1}}(z) &= \mathbb{P}\big(x_{k+1} \le z\big) \\
&= \mathbb{P}\Big(Y \le z - xu_k\widetilde{X}_{\tau_{k+1}}|\widetilde{X}_{\tau_{k+1}} = J\Big)\mathbb{P}\Big(\widetilde{X}_{\tau_{k+1}} = J\Big) + \\
&\quad + \mathbb{P}\Big(Y \le z - xu_k\widetilde{X}_{\tau_{k+1}}|\widetilde{X}_{\tau_{k+1}} = -J\Big)\mathbb{P}\Big(\widetilde{X}_{\tau_{k+1}} = -J\Big) \\
&= F_Y(z-\xi)p + F_Y(z+\xi)(1-p) \\
&= p\Big\{1 - \Big[2\cosh(\widetilde{\mu}J/\sigma^2)K^{\infty}_{\frac{1}{r}\log(\frac{z-\xi}{x})}(J)\Big]\Big\}\mathbb{1}_{[x+\xi,\infty)} + \\
&\quad + (1-p)\Big\{1 - \Big[2\cosh(\widetilde{\mu}J/\sigma^2)K^{\infty}_{\frac{1}{r}\log(\frac{z+\xi}{x})}(J)\Big]\Big\}\mathbb{1}_{[x-\xi,\infty)}.
\end{aligned}$$

where $\xi = xu_kJ$. Now, the density is obtained differentiating the cdf above. This amounts to compute $\frac{d}{dz}K^{\infty}_{\frac{1}{r}\log(\frac{z-\xi}{x})}(J)$ and $\frac{d}{dz}K^{\infty}_{\frac{1}{r}\log(\frac{z+\xi}{x})}(J)$. As an example, let us compute the first derivative:

$$\begin{aligned}
\frac{d}{dz}K^{\infty}_{\frac{1}{r}\log(\frac{z-\xi}{x})}(J) &= \frac{d}{dz}\Bigg(\frac{\sigma^2\pi}{4J^2}\sum_{n=1}^{\infty}\frac{n(-1)^{n+1}}{\frac{\tilde{\mu}^2}{2\sigma^2} + \frac{\sigma^2n^2\pi^2}{8J^2}}\Big(\frac{z-\xi}{x}\Big)^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2} + \frac{\sigma^2n^2\pi^2}{8J^2}\right)}\sin\big(\frac{\pi}{2}n\big)\Bigg) \\
&= \big(-\frac{1}{rx}\big)\frac{\sigma^2\pi}{4J^2}\sum_{n=1}^{\infty}n(-1)^{n+1}\Big(\frac{z-\xi}{x}\Big)^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2} + \frac{\sigma^2n^2\pi^2}{8J^2}\right)-1}\sin\big(\frac{\pi}{2}n\big) \\
&= \big(-\frac{1}{rx}\big)\Gamma^{\infty}_{\frac{z-\xi}{x}}(J).
\end{aligned}$$

Substituting into

$$f_{x_{k+1}}(z) = 2\cosh\Big(\frac{\widetilde{\mu}J}{\sigma^2}\Big)\Big[-p\frac{d}{dz}K^{\infty}_{\frac{1}{r}\log(\frac{z-\xi}{x})}(J)\mathbb{1}_{(x+\xi,\infty)} - (1-p)\frac{d}{dz}K^{\infty}_{\frac{1}{r}\log(\frac{z+\xi}{x})}(J)\mathbb{1}_{(x-\xi,\infty)}\Big]$$

and rearranging, gives us the result. □

Density (7.9) is plotted in Figure 7.1 for different values of the risky asset portfolio weight u_k.

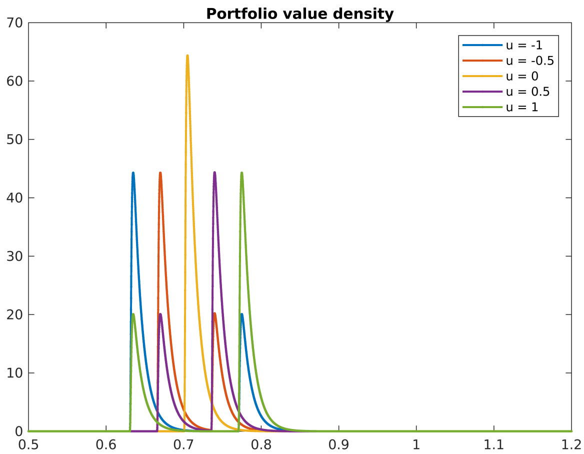

Figure 7.1: Probability density function of random variable variable x_{k+1} for different value of the risky-asset weight u_k and the following parameters: $x = 0.7$, $r = 5.5\%$, $\mu = 0.114$, $\sigma = 0.1602$ and $J = 10\%$

7.1.3 Numerical results

In this section we report the results obtained by applying the ODAA algorithm in an event-driven framework and modeling the risky asset as a Geometric Brownian Motion. Before showing the results however, we need to discuss the calibration of parameter μ and σ of the GBM dynamics to market data and the truncation of the series in (7.10). As far as data is concerned, we use the same time series presented in Section 6.3 (observations from the Future S&P 500 index from 22 January 2010 to 25 April 2016).

GBM calibration Let us consider the discretized solution of SDE (7.1)

$$\begin{aligned} S_{t_{k+1}} &= S_{t_k} \exp\left\{(\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma(W_{t_{k+1}} - W_{t_k})\right\} \\ &= S_{t_k} \exp\left\{(\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t} Z\right\} \end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$. In terms of log-return, we have

$$X_{t_{k+1}} = (\mu - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t} Z.$$

Let $x_1, \ldots, x_n$ be a random sample of log-returns from a normal population with mean $(\mu - \frac{1}{2}\sigma^2)\Delta t$ and variance $\sigma^2\Delta t$. From this random sample we obtain the following estimates

$$\widehat{\sigma}^2 = \frac{s^2}{\Delta t}$$

$$\widehat{\mu} = \frac{\bar{x}}{\Delta t} + \frac{1}{2}\widehat{\sigma}^2$$

where s^2 is the sample variance and $\bar{x}$ the sample mean. Setting $\Delta t = 1/252$ and applying the equations above to our data we get $\widehat{\mu} = 0.0594$ and $\widehat{\sigma} = 0.1923$.

Series truncation In practice, the series in (7.10) has to be truncated. Let $\Gamma_z^N(J)$ be the function in (7.10) when the series is truncated at the Nth term. The residual could be bounded in the following way

$$\begin{aligned}
\left|\Gamma_z^{\infty}(J) - \Gamma_z^{N}(J)\right| &\leq \frac{\sigma^2\pi}{4J^2}\left|\sum_{n=N+1}^{\infty} n(-1)^{n+1} z^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2}+\frac{\sigma^2 n^2\pi^2}{8J^2}\right)-1}\sin(\frac{\pi}{2}n)\right| \\
&\leq \frac{\sigma^2\pi}{4J^2}\sum_{n=N+1}^{\infty} n z^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2}+\frac{\sigma^2 n^2\pi^2}{8J^2}\right)-1} \\
&\leq \frac{\sigma^2\pi}{4J^2}\int_{N}^{\infty} n z^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2}+\frac{\sigma^2 n^2\pi^2}{8J^2}\right)-1}\mathrm{d}n \\
&\leq \frac{r}{\pi\log z} z^{-\frac{1}{r}\left(\frac{\tilde{\mu}^2}{2\sigma^2}+\frac{\sigma^2 N^2\pi^2}{8J^2}\right)-1}
\end{aligned}$$

therefore, denoting by ϵ the maximum error tolerance, we obtain

$$N(z) \geq \sqrt{-\frac{8rJ^2}{\sigma^2\pi^2}\left\{\frac{\widetilde{\mu}^2}{2r\sigma^2} + \log_z\left(\frac{\epsilon\pi z\log z}{r}\right)\right\}}. \qquad (7.11)$$

which is the minimum number of terms in the series in order to have the specified accuracy.

We considered an investment characterized by the following parameters:

- Initial wealth $x_0 = 1$.
- $J = 10\%$.
- $N = 10$ portfolio rebalancings.
- The following target sets: $X_0 = \{1\}$, $X_k = [0, \infty)$ for $k = 1, \ldots, 9$ and $X_{10} = [(1+\theta)^3, \infty)$ with $\theta = 7\%$. In the implementation, they were approximated by $X_k = [0.5, 1.9]$ for $k = 1, \ldots, 9$ and $X_{10} = [(1+\theta)^3, 1.9]$ and discretized with a step length of 5×10^{-4}.

- $r = 5.5\%$ (risk-free rate of return).

Some of the allocation maps obtained via ODAA algorithm are reported in Figure 7.2. They exhibit the *contatrian* attitude typical of all the ODAA strategies presented so far. The joint probability of reaching the target sets is $p^{\star} = J(x_0) = 0.9727$. This result is verified by running a Monte Carlo simulation with 10^6 replications at each rebalancing period, the outcome, in terms of joint probability and other investment statistics, is reported in Table 7.1. Finally, in Figure 7.3 we show the empirical density function of the investment return.

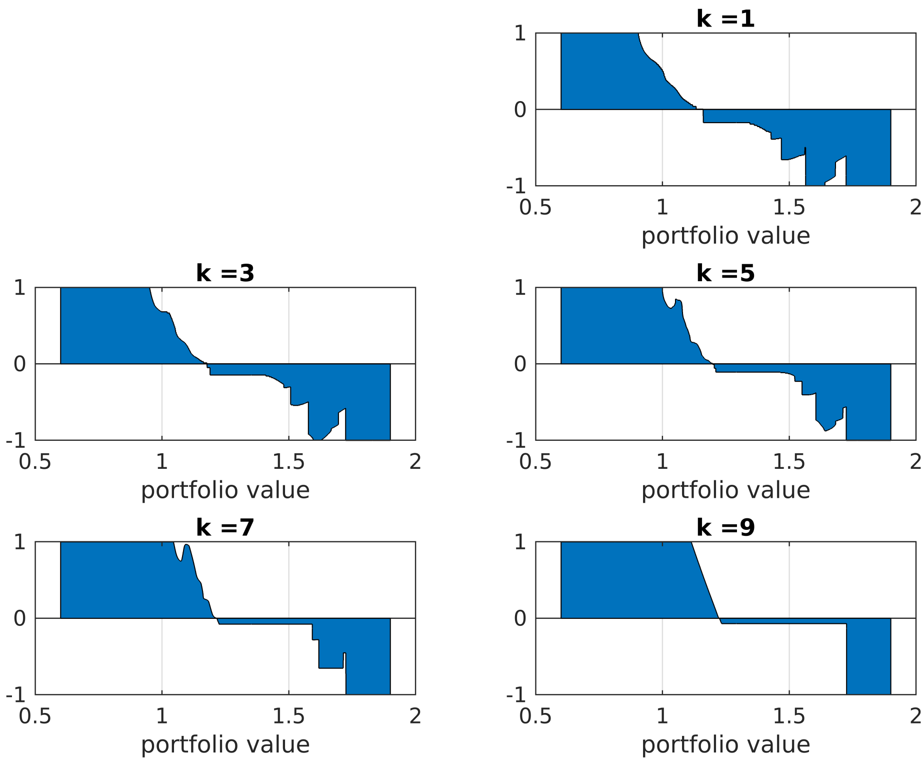

Figure 7.2: allocation maps for the risky asset, which is modeled according to a Geometric Brownian Motion

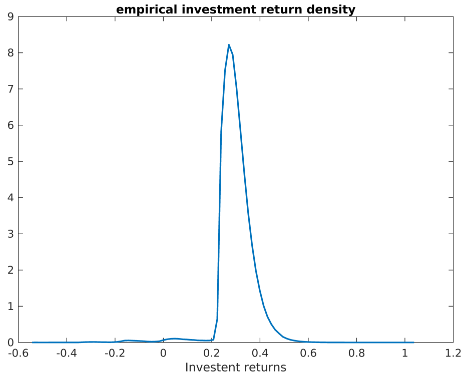

Figure 7.3: Empirical density function of the investment return when the risky asset is modeled as a GBM.

Statistics	value
$p^\star$	0.9727
p_{MC}	0.9728
Mean Return (ann)	7.50%
Volatility (ann)	3.93%
Median Return (ann)	7.46%
Skewness	-1.81
Kurtosis	16.61
Sharpe Ratio	0.471
Avg horizon [years]	3.70

Table 7.1: Investment performance obtained via Monte-Carlo simulation with 10^6 replication at each rebalancing time.

7.2 Interest rate dynamics for the risk-free asset

In this section we try to extend the model presented in Chapter 6 by assuming a dynamics for the bank account interest rate. Starting from the portfolio dynamics (6.10), our aim is replacing it with

$$\boxed{x_{k+1} = x_k\Big(\exp\Big\{\int_{t_k}^{t_k+\tau_{k+1}} r_s \mathrm{d}s\Big\} + u_k J\widetilde{N}_{k+1}\Big)} \qquad k \in \mathbb{N} \tag{7.12}$$

where $\{r_t\}_{t\geq t_k}$ is the short-rate process and everything else remains unchanged from Chapter 6. Due to its analytical tractability, we decided to model the short-rate according to the Vasicek Model (see [8]). The following SDE provides the dynamics of the short-rate

$$\begin{cases} dr_t = a(b - r_t)dt + \sigma dW_t \\ r_{t_k} = r_k, \qquad t \geq t_k \end{cases} \tag{7.13}$$

where a, b, σ and r_k are positive constants and $\{W_t\}_{t\geq t_k}$ is a unidimensional Brownian motion. The main feature of the Vasicek model is the *mean reversion* property: the process will tend to move to its average over time. Moreover, the process has a non-null probability of becoming negative. This is no longer a taboo since negative interest rates are observed in the market.

The solution of SDE (7.13) is the following Ornstein–Uhlenbeck process

$$r_t = r_k e^{-a(t-t_k)} + b(1 - e^{-a(t-t_k)}) + \sigma e^{-a(t-t_k)} \int_{t_k}^{t} e^{a(s-t_k)} dW_s. \tag{7.14}$$

However, we are interested in an explicit expression of the integral of process r_t since it appears in the portfolio dynamics (7.12). Therefore, let us define the integrated short-rate process by $v_t = \int_{t_k}^{t} r_s \mathrm{d}s$. In order to find its explicit form, we integrate (7.13) from t_k to a generic instant t, obtaining

$$r_t = r_k + a\big(b(t - t_k) - v_t\big) + \sigma(W_t - W_{t_k}). \tag{7.15}$$

After equating (7.15) and (7.14) and solving for v_t we get

$$v_t = \frac{1}{a}\Big[(r_k - b)(1 - e^{-a(t-t_k)}) + ab(t - t_k) + \sigma \int_{t_k}^{t} (1 - e^{-a(t-s)}) dW_s\Big]. \tag{7.16}$$

From Appendix B and (Proposition 7.1, [4]) we have that

$$v_t \sim \mathcal{N}\Big(\eta(t-t_k), \zeta^2(t-t_k)\Big)$$

$$\eta(t-t_k) = \mathbb{E}[v_t] = \frac{1}{a}\Big[(r_k - b)\big(1 - e^{-a(t-t_k)}\big) + ab(t-t_k)\Big] \tag{7.17}$$

$$\zeta^2(t-t_k) = \text{Var}\,[v_t] = \frac{\sigma^2}{2a^3}\Big[2a(t-t_k) + 4e^{-a(t-t_k)} - e^{-2a(t-t_k)} - 3\Big] \tag{7.18}$$

where $t \in \mathbb{R}^+$.

***Remark* 7.2.1:** The parameters $\eta(t-t_k)$ and $\zeta(t-t_k)$ depends only on the difference $t - t_k$, therefore the distribution of v_t is stationary. This fact will be particularly useful in the following since the distribution of v_{t_k+t} (which is the one we are interested in) depends only on t.

7.2.1 The density of x_{k+1}

The last step is finding the probability density function of the random variable x_{k+1}.

***Proposition* 7.2.1:** Let x_{k+1} be the random variable (7.12). Then, its probability density function is

$$\begin{aligned} f_{x_{k+1}}(z) = & \frac{p}{(z-\xi)}\left\{\int_0^\infty \varphi\Big(\frac{\log\big(\frac{z-\xi}{x}\big) - \eta(t)}{\zeta(t)}\Big)\Big(\frac{1}{\zeta(t)}\Big) f_{\tau_{k+1}}(t)\mathrm{d}t\right\}\mathbb{1}_{[x+\xi,\infty)} + \\ & + \frac{(1-p)}{(z+\xi)}\left\{\int_0^\infty \varphi\Big(\frac{\log\big(\frac{z+\xi}{x}\big) - \eta(t)}{\zeta(t)}\Big)\Big(\frac{1}{\zeta(t)}\Big) f_{\tau_{k+1}}(t)\mathrm{d}t\right\}\mathbb{1}_{[x-\xi,\infty)} \end{aligned}$$

where $\xi = xJu_k$, $f_{\tau_{k+1}} = \big\{\lambda e^{-\lambda t}\big\}\mathbb{1}_{[0,\infty)}$ is the density of random variable τ_{k+1}, φ is the density of a standard normal and $\eta(t), \zeta(t)$ are function (7.17) and the square root of (7.18) respectively, computed in $t + t_k$.

Proof. Let us rewrite the portfolio dynamics in the following way

$$x_{k+1} = x\exp\{v_{t_k+\tau_{k+1}}\} + xJu_k\widetilde{N}_{k+1} = Y + \xi\widetilde{N}_{k+1}.$$

The first step of the proof is finding the cdf of Y:

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}\big(x\exp\{v_{t_k+\tau_{k+1}}\} \le y\big) \\
&= \int_0^\infty \mathbb{P}\Big(v_{t_k+\tau_{k+1}} \le \log(y/x)\big|\tau_{k+1} = t\Big) f_{\tau_{k+1}}(t)\mathrm{d}t \\
&= \int_0^\infty \mathbb{P}\Big(v_{t_k+t} \le \log(y/x)\Big) f_{\tau_{k+1}}(t)\mathrm{d}t \\
&= \bigg\{\int_0^\infty \Phi\Big(\frac{\log(y/x)-\eta(t)}{\zeta(t)}\Big) f_{\tau_{k+1}}(t)\mathrm{d}t\bigg\}\mathbb{1}_{[x,\infty)}
\end{aligned}
$$

where we used the Law of Total Probability in the continuous case and the fact that v_{t_k+t} is Gaussian with parameters $\eta(t)$ and $\zeta(t)$ (which are the functions (7.17) and the square root of (7.18) respectively, computed in $t+t_k$). Φ is the cdf of a standard normal.

Repeating the usual steps, we are able to find the cdf of x_{k+1}:

$$
\begin{aligned}
F_{x_{k+1}}(z) &= \mathbb{P}\big(Y + \xi\widetilde{N}_{k+1} \le z\big) \\
&= F_Y(z-\xi)p + F_Y(z+\xi)(1-p) \\
&= p\bigg\{\int_0^\infty \Phi\Big(\frac{\log\left(\frac{z-\xi}{x}\right)-\eta(t)}{\zeta(t)}\Big) f_{\tau_{k+1}}(t)\mathrm{d}t\bigg\}\mathbb{1}_{[x+\xi,\infty)} + \\
&+ (1-p)\bigg\{\int_0^\infty \Phi\Big(\frac{\log\left(\frac{z+\xi}{x}\right)-\eta(t)}{\zeta(t)}\Big) f_{\tau_{k+1}}(t)\mathrm{d}t\bigg\}\mathbb{1}_{[x-\xi,\infty)}.
\end{aligned}
$$

where p is the probability that $\widetilde{N}_{k+1}$ is equal to 1. Finally, deriving with respect to z under the integral sign gives us the result. □

7.2.2 The calibration of the Vasicek model

In this section we see how parameters a, b and σ of the Vasicek model can be calibrated to data using the linear regression method. We closely follow [5].

The first step is to discretize Equation (7.13) in order to find a linear relation between the short-rate process r_t at time instants t and $t+\Delta t$:

$$
\begin{aligned}
r_{t+\Delta t} &= r_t + a(b-r_t)\Delta t + \sigma\sqrt{\Delta t}Z \\
&= (1-a\Delta t)r_t + ab\Delta t + \sigma\sqrt{\Delta t}Z \qquad (7.19)
\end{aligned}
$$

where $Z \sim \mathcal{N}(0,1)$. From (7.19) is clear the linear relation between $r_{t+\Delta t}$ and r_t. Therefore, the above equation can be rewritten in a fashion suitable for the linear regression estimation as follows

$$
r_{t+\Delta t} = \alpha r_t + \beta + \varepsilon
$$

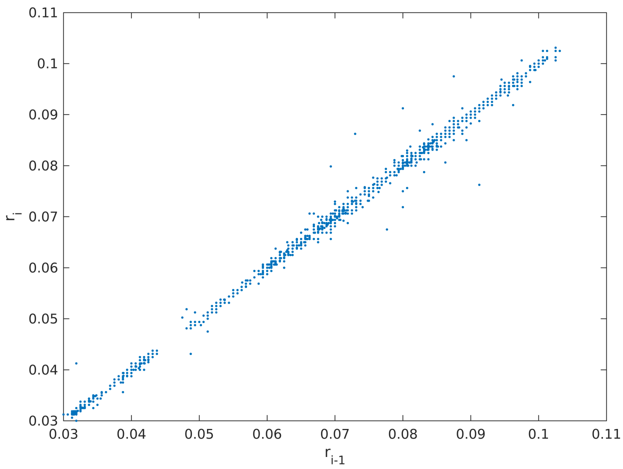

Figure 7.4: Linear relation between consecutive daily 3-month LIBOR rate quotes.

where

$$\begin{cases} \alpha = 1 - a\Delta t \\ \beta = ab\Delta t \end{cases}$$

and $\varepsilon \sim \mathcal{N}(0, \sigma^2 \Delta t)$ is the error term. Let us suppose to have a random sample $\{r_1, \ldots, r_n\}$, for example daily 3-month LIBOR quotes. The independent variable and the response are

$$\boldsymbol{y} = \begin{bmatrix} r_2 \\ \vdots \\ r_n \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} r_1 \\ \vdots \\ r_{n-1} \end{bmatrix}.$$

The linear relation between $\boldsymbol{y}$ and $\boldsymbol{x}$ is reported in Figure (7.4).

Once the least square estimates $\widehat{\alpha}$ and $\widehat{\beta}$ have been found, we can recover the estimates of the Vasicek model parameters in the following way

$$\begin{cases} \widehat{a} = \frac{1}{\Delta t}(1 - \widehat{\alpha}) \\ \\ \widehat{b} = \widehat{\beta}/(\widehat{a}\Delta t) \\ \\ \widehat{\sigma} = \frac{RSE}{\sqrt{\Delta t}} \end{cases}$$

where RSE is the residual standard error of the fitting.

Applying this method to daily ($\Delta t = 1/252$) quotes of 3-month LIBOR starting from 22 January 2010 to 25 April 2016 brings the following estimates

$$\begin{cases} \widehat{a} = 0.1670 \\ \widehat{b} = 0.0280 \\ \widehat{\sigma} = 0.0160. \end{cases}$$

Chapter 8

Conclusions

In this final chapter we summarize what has been achieved in this thesis and then suggest some directions for future research.

8.1 The thesis in a nutshell

Recent results in Stochastic Reachability have been the basis for this thesis. These results (e.g. the ODAA theorem) allowed us to cast the asset allocation problem in a Control System setting. The generality of this approach permitted portfolio rebalancings to be driven first by time and then by the occurrence of discrete events. In the time-driven case, a universe of three asset classes (Cash, Bond and Equity) was considered and the allocation maps for a 2-year investment were obtained. These maps exhibited a *contrarian* attitude: the higher the portfolio performance the less risky the asset class mix, the lower the performance the riskier the mix. In the 2-year investment example, the ODAA strategy outperformed both the CPPI and the constant-mix in terms of annualized returns while showing a comparable risk profile. In the event-driven case, the portfolio consisted of a risk-free asset (a bank account) and a risky asset (an exposure to a future index). In this context, the risky asset allocation maps, obtained via ODAA algorithm, showed a *contrarian* behavior in the sense that when portfolio performance is up the risky asset is shorted, when it is down a long position is taken instead.

As far as the time-driven model is concerned, a small contribution to the literature has been the use of Generalized Hyperbolic distributions, as an alternative to the Gaussian Mixture model proposed in [28], to describe the statistical properties of asset class returns. The personal contribution to the event-driven literature has been the extension of the model introduced in [29] in two innovative ways. In particular, first the risky asset has been modeled as a Geometric Brownian Motion, then an interest rate dynamics as been considered for the bank account. In the first case, the usual *contrarian* allocation maps for the risky asset were obtained. In the second case, the complexity of the formulas made the implementation quite difficult. In particular,

8.2 Further developments

The first idea could be casting the asset allocation problem in a stochastic hybrid system setting. A stochastic hybrid system is a control system whose evolution has a continuous and a discrete component. For instance, the problem of maintaining the temperature of $r \in \mathbb{N}$ different rooms within a certain range over time could be modeled as a stochastic hybrid system. The discrete state space being the mode (ON or OFF) of the heater switches and the continuous state space the room temperatures. In an asset allocation problem, the discrete state space could be the states of the economy. Main references are [3] for the discrete-time case and [10] for the continuous-time case.

Appendix A

Probability Distributions

In this appendix we give further details about the probability distributions used in part I. Main references are [7], [22] and [21].

A.1 Generalized Inverse Gaussian

Definition **A.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 **A.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

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

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:

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

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

A.2 Density Functions

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

A.2.1 GH

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

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

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

A.2.2 Student-t

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

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

where

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

A.2.3 VG

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

where

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

Appendix B

Mean and Variance of an Integrated Ornstein–Uhlenbeck process

In this appendix we compute $\mathbb{E}[v_t]$ and $\mathrm{Var}\,[v_t]$ where v_t is the random variable (7.16).

The mean Given that the mean of a stochastic integral in 0 (see [4]), the mean function is simply

$$\mathbb{E}[v_t] = \frac{1}{a}\Big[(r_k - b)\big(1 - e^{-a(t-t_k)}\big) + ab(t - t_k)\Big]. \tag{B.1}$$

The variance

$$\begin{aligned}
\mathrm{Var}\,[v_t] &= \mathrm{Var}\left[\frac{\sigma}{a}\int_{t_k}^{t}(1 - e^{-a(t-s)})dW_s\right] \\
&= \Big(\frac{\sigma}{a}\Big)^2 \mathbb{E}\Big[\Big(W_t - W_{t_k} - e^{-at}\int_{t_k}^{t} e^{as}dW_s\Big)^2\Big] \\
&= \Big(\frac{\sigma}{a}\Big)^2 \left[(t - t_k) + e^{-2at}\mathbb{E}\Big[\big(\int_{t_k}^{t} e^{as}dW_s\big)^2\Big] - 2e^{-at}\mathbb{E}\Big[\int_{t_k}^{t} dW_s \int_{t_k}^{t} e^{as}dW_s\Big]\right] \\
&= \Big(\frac{\sigma}{a}\Big)^2 \left[(t - t_k) + e^{-2at}\int_{t_k}^{t} e^{2as}\mathrm{d}s - 2e^{-at}\int_{t_k}^{t} e^{as}\mathrm{d}s\right] \\
&= \Big(\frac{\sigma}{a}\Big)^2 \left[(t - t_k) + \frac{1}{2a}(1 - e^{-2a(t-t_k)}) - \frac{2}{a}(1 - e^{-a(t-t_k)})\right] \\
&= \frac{\sigma^2}{2a^3}\left[2a(t - t_k) - e^{-2a(t-t_k)} + 4e^{-a(t-t_k)} - 3\right]
\end{aligned}$$

where in the fourth equality we used Ito's lemma ([4],Theorem 7.1) and ([4], Remark 7.1)

Bibliography

[1] `https://la.mathworks.com/matlabcentral/fileexchange/17931-hzmvntest`.

[2] The asset-minded professors. `http://www.economist.com/node/9333696`, June 2007.

[3] Abate, A., Prandini, M., Lygeros, J., and Sastry, S. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica 44*, 11 (2008), 2724 – 2734.

[4] Baldi, P. *Stochastic Calculus: An Introduction Through Theory and Exercises*. Universitext. Springer International Publishing, 2017.

[5] Bernal, V. Calibration of the vasicek model: An step by step guide. `https://victor-bernal.weebly.com/uploads/5/3/6/9/53696137/projectcalibration.pdf`, 2016.

[6] Bernstein, P. *Capital Ideas: The Improbable Origins of Modern Wall Street*. Wiley, 2012.

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

[8] Brigo, D., and Mercurio, F. *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. Springer Finance. Springer Berlin Heidelberg, 2007.

[9] 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.

[10] Bujorianu, L. M. *Stochastic Reachability Analysis of Hybrid Systems*. Communications and Control Engineering. Springer-Verlag London, 2012.

[11] Campbell, R., Koedijk, K., and Kofman, P. Increased correlation in bear markets. *Financial Analysts Journal 58*, 1 (2002), 87–94.

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

[13] Cassandras, C., and Lafortune, S. *Introduction to Discrete Event Systems*. SpringerLink Engineering. Springer US, 2009.

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

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

[16] 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.

[17] Hieber, P., and Scherer, M. A note on first-passage times of continuously time-changed brownian motion. *Statistics & Probability Letters 82*, 1 (2012), 165–172.

[18] Hull, J. *Options, Futures, and Other Derivatives*. Pearson Education, 2014.

[19] Hull, J. *Risk Management and Financial Institutions*. Wiley Finance. Wiley, 2015.

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

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

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

[23] 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.

[24] Perold, A. F., and Sharpe, W. F. Dynamic strategies for asset allocation. *Financial Analysts Journal 44*, 1 (1988), 16–27.

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

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

[27] 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.

[28] 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.

[29] Specchio, R. An event driven/stochastic reachability approach to portfolio construction. final project at SCHOOL OF MANAGEMENT, Politecnico di Milano, 2011.