Department of Mathematics and Statistics,
University of Helsinki

Master's thesis

Growth Optimal Portfolio: Analysis and construction on a discrete multi-period market

Viacheslav Bazaliy

Thesis supervisor
Jaakko Lehtomaa

May 12, 2019

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Mathematics and Statistics

Tekijä — Författare — Author
Bazaliy Viacheslav

Työn nimi — Arbetets titel — Title

Growth Optimal Portfolio: Analysis and construction on a discrete multi-period market

Oppiaine — Läroämne — Subject
Mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	May 2019	62 p.

Tiivistelmä — Referat — Abstract

This thesis provides an analysis of *Growth Optimal Portfolio* (GOP) in discrete time. Growth Optimal Portfolio is a portfolio optimization method that aims to maximize expected long-term growth. One of the main properties of GOP is that, as time horizon increases, it outperforms all other trading strategies almost surely. Therefore, when compared with the other common methods of portfolio construction, GOP performs well in the long-term but might provide riskier allocations in the short-term.

The first half of the thesis considers GOP from a theoretical perspective. Connections to the other concepts (*numeraire portfolio, arbitrage freedom*) are examined and derivations of optimal properties are given. Several examples where GOP has explicit solutions are provided and sufficiency and necessity conditions for growth optimality are derived.

Yet, the main focus of this thesis is on the practical aspects of GOP construction. The iterative algorithm for finding GOP weights in the case of independently log-normally distributed growth rates of underlying assets is proposed. Following that, the algorithm is extended to the case with non-diagonal covariance structure and the case with the presence of a risk-free asset on the market. Finally, it is shown how GOP can be implemented as a trading strategy on the market when underlying assets are modelled by ARMA or VAR models. The simulations with assets from the real market are provided for the time period 2014-2019.

Overall, a practical step-by-step procedure for constructing GOP strategies with data from the real market is developed. Given the simplicity of the procedure and appealing properties of GOP, it can be used in practice as well as other common models such as *Markowitz* or *Black-Litterman* model for constructing portfolios.

Avainsanat — Nyckelord — Keywords
Growth Optimal Portfolio, Portfolio Construction, Time Series Models

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpula Campus Library

Muita tietoja — Övriga uppgifter — Additional information

Acknowledgements

I want to thank my supervisor Jaakko Lehtomaa for putting my random flight of fantasy into a nice scientific journey that is resulted in the accurate mathematical text. I am also very grateful to Dario Gasbarra for proposing the most important results of this thesis and for very sapid discussions that often lead to new valuable results or corrections which became part of this thesis.

Contents

1. Introduction

1.1 The history of GOP

It is generally believed that Growth Optimal Portfolio concept was founded by Kelly in his 1956 paper "A New Interpretation of Information Rate" [1], although he considered gambling and did not name his strategy as growth optimal. His main result derived from information theory was that there exists an optimal gambling strategy that accumulates more wealth than any other strategy with probability one. In modern terms, he indeed presented Growth Optimal Portfolio for the class of gambling games.

From the other perspective, GOP's main idea is to maximize the average geometric return of the strategy and this idea was mentioned in different sources earlier. For example, Christensen (2012) [3] refers to William's (1936) paper [2] where it was stated that speculators should focus on the geometric mean instead of arithmetic due to compounding. We examine the optimality properties of GOP, and in particular long-term growth in Section 2.2.

Finally, we can consider the Growth Optimal Portfolio as a particular case of expected utility maximization. Namely, to obtain GOP one can choose logarithm utility function for the future wealth and then use the usual utility maximization framework. The utility theory has deep roots in the history, conventional wisdom holds that it starts from the solution to the famous St. Petersburg gambling paradox that was posed by Nicolas Bernoulli and resolved by Daniel Bernoulli in the 18th century. The logarithm function is commonly used in the utility theory since it is one of the simplest examples of the analytic function that has desired growth and convexity properties.

The GOP discovery was motivated by its appealing growth performance in the long run. However, recent research mainly concentrates on the other aspect of GOP, to wit, its numeraire property. The fact that prices divided by GOP are supermartingales with respect to original probability measure was firstly mentioned by Breiman (1960) [4] and recently was intensively studied by Platen [5], [8], [11]. In Section 2.1 of this thesis, we show that Numeraire Portfolio and GOP are the same concepts.

According to Christensen [3], calculation of GOP weights is generally very difficult in discrete time. Thus, the main focus of this thesis is on the practical side of GOP construction in discrete time. In chapter Chapter 3 we discuss the difficulties related to the optimization problem of GOP, derive necessary and sufficient conditions and show several examples where GOP weights have closed form solution. Following that, in Chapter 4 we consider GOP in the case when growth ratios of underlying assets are log-normally distributed. In Section 4.3 we propose an iterative algorithm for finding GOP's weights, then improve it by variance reduction methods and examine its convergence in Sections 4.4, 4.5. To apply this algorithm on the real market, it is needed to capture the dynamics of underlying assets. For these purposes, we use simple time-series models ARMA and VAR. We show how GOP's optimization problem can be adapted to the time-series model of general form in Section 4.2. Finally, in Chapter 5 we fit ARMA and VAR models for selected assets from the real market, simulate the returns of GOP over 2014-2019 time period and discuss the observed properties of GOP.

1.2 Market model and definitions

Let's consider a multi-period discrete market model. There are $N+1$ primary securities where the first one is a risk-free bond and we assume that transactions are allowed only at times $t \in \{0, 1, ..., T\}, T \in \mathbb{N} \cup \infty$. We denote the price of security i at time t as $\mathbf{S}_{t,i}$ and the price vector $\mathbf{S}_t = (\mathbf{S}_{t,0}, \mathbf{S}_{t,1}, ..., \mathbf{S}_{t,N})$. At each time t, all agents know the full price history $(\mathbf{S}_0, .., \mathbf{S}_t)$ and can make their decisions based on this information. We assume that at time 0 prices are known and deterministic. Mathematically, we can model this information by the sequence of nested sigma algebras $\mathcal{F}_t = \sigma(\mathbf{S_0}, \mathbf{S_1}, ..., \mathbf{S_t})$ called *filtration*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset ... \subset \mathcal{F}_T,$$
$$\mathbb{F} = (\mathcal{F}_0, \mathcal{F}_1, ..., \mathcal{F}_T).$$

Thus, we represent the market by the *stochastic field* $(\Omega, \mathbb{F}, \mathbb{P})$ where all random variables are defined.

Remark 1.1. Securities on the market can generate interest, dividends and other cash flows. Therefore, for rigorous analysis, we have to consider the cumulative price indices that represent a strategy of buying one security and reinvesting all generated cash flow in that security. However, sometimes we can avoid it and use pure prices of securities. For example, an investor can assume that market prices do not incorporate the information regarding other cash flows or those cash flows are negligible like in the case of red-hot growth stocks. For simplicity, we will provide an analysis for security prices but the same theoretical framework can be used for cumulative price indices.

Assumption 1.2.1. $\mathbf{S}_{t,i} > 0$ *for each* $t \in \mathbb{T}$ *and* $i \in \overline{0,N}$ [1] *almost surely.*

Let's denote for convenience $\mathbb{T} = T \in \overline{0,T}$, where $T \in \mathbb{N} \cup \infty$. Using Assumption 1.2.1 we can introduce the notion of *growth ratio.*

Definition 1.2.1 (Growth ratio)**.** The *growth ratio* $\mathbf{h}_{t,i}$ of security $i \in \overline{0,N}$ at time $t \in \mathbb{T}$ is defined as

$$\mathbf{h}_{t,i}(\omega) = \begin{cases} \frac{\mathbf{S}_{t,i}(\omega)}{\mathbf{S}_{t-1,i}(\omega)}, & \text{when } \mathbf{S}_{t-1,i}(\omega) > 0 \\ 0, & \text{otherwise} \end{cases}.$$

Remark 1.2. Growth ratio $\mathbf{h}_{t,i}$ is adapted to filtration $\mathbb{F}$ since $\mathbf{S}_t$ is an adapted process.

Remark 1.3. Note that $\mathbf{h}_{t,i}$ correspond to the usual return rate as $\mathbf{h}_{t,i} = 1 + \mathbf{r}_{t,i}$

We also need to induce some natural restrictions on the growth ratio. Namely, we assume that one-period growth of each security is finite and there is one risk-neutral security that allows us to construct portfolios with strictly positive growth ratios.

Assumption 1.2.2. *We assume that*

1. $\mathbf{h}_{t,i}$ *is almost surely finite. For all* $t \in \mathbb{T}$ *and* $i \in \overline{0,N}$: $\mathbb{P}(\mathbf{h}_{t,i} < \infty) = 1$,
2. $\forall t \in \mathbb{T} : \mathbb{P}(\mathbf{h}_{t,0} > 0) = 1$ *i.e. return of risk-neutral bond is positive almost surely.*

Remark 1.4. Combining assumptions 1.2.1 and 1.2.2 we get that $\forall t \in \mathbb{T} : \mathbf{S}_{t,0} > 0$.

Investor acts on the market by buying or selling financial instruments. Therefore, at each time step, we can characterize his or her decision by the vector where each component represents the number of securities in a positive (long) or negative (short) position. The vector represents the allocation of investor's wealth and we say that an investor holds a *portfolio.* Obviously, gross value and nominal growth of each portfolio depend on the investor's initial capital. However, the growth ratio of the portfolio depends only on the proportions invested in the securities as we will show in Derivation 1.9.

Definition 1.2.2 (Portfolio and Portfolio weights)**.** Let's consider an investor who at time $t \in \mathbb{T}$ has overall wealth W_t. The proportion of wealth invested in security i at that time is

$$w_{t,i} = \frac{\textit{capital invested in security } i \textit{ at time } t}{\mathbf{W}_t}.$$

Note that investment decisions are made given only the past information. Since $\mathbf{S}_t$ is an adapted process, we have $w_{t,i} \in \mathcal{F}_{t-1}$. As in the previous notation, we denote $w_t = (w_{t,0}, \dots, w_{t,N})$. The portfolio V is a combination of initial wealth W_0 and sequence of vectors w_t with corresponding invested portions

$$V = \{(W_0, w_0, ..., w_T) \mid W_0 \in \mathbb{R}^+ \text{ and } w_t \in \mathcal{F}_{t-1}\}. \quad (1.5)$$

[1] We use the following notation $\overline{K,N} = \{K, K+1, ..., N-1, N\}$.

Remark 1.6. For convenience, we will mark the initial wealth or weights that belong to portfolio V by a superscript i.e. W_0^V or $w_{t,i}^V$.

Let V be a portfolio of some investor, so at time $t \in \mathbb{T}$ the proportions vector is w_t and the value is $\mathbf{W}_t^V$. By Definition 1.2.2, we can write

$$\mathbf{W}_t^V = \sum_{i=0}^{N} w_{t,i}\mathbf{S}_{t,i}. \tag{1.7}$$

We restrict our analysis to *self-financing* portfolios i.e portfolios where the value of the portfolio varies only due to changes in primary securities. In other words, we assume that investors do not withdraw or add any cash to their investment accounts.

Assumption 1.2.3 (Self-financing property)**.** *At any time point* $t \in \mathbb{T} \setminus \{0\}$ *an investor can only use the wealth that is generated by his or her portfolio in the previous period* $t-1$

$$\sum_{i=0}^{N} w_{t,i}\mathbf{S}_{t,i} = \sum_{i=0}^{N} w_{t-1,i}\mathbf{S}_{t,i}, \text{ for all } t \in \mathbb{T} \setminus 0. \tag{1.8}$$

Let's now compute the growth ratio of portfolio V at time $t \in \mathbb{N}$

$$\begin{aligned} \mathbf{h}_t^V &= \frac{\sum_{i=0}^{N} w_{t,i}^V \cdot \mathbf{S}_{t,i}}{\sum_{i=0}^{N} w_{t-1,i}^V \cdot \mathbf{S}_{t-1,i}} = \sum_{i=0}^{N} \frac{w_{t-1,i}^V \cdot \mathbf{S}_{t-1,i}}{\sum_{i=0}^{N} w_{t-1,i}^V \cdot \mathbf{S}_{t-1,i}} \cdot \frac{\mathbf{S}_{t,i}}{\mathbf{S}_{t-1,i}} = \\ &= \sum_{i=0}^{N} w_{t-1,i}^V \cdot \frac{\mathbf{S}_{t,i}}{\mathbf{S}_{t-1,i}} = \sum_{i=0}^{N} w_{t-1,i}^V \cdot \mathbf{h}_{t,i}.^2 \end{aligned} \tag{1.9}$$

In the following definitions we will have the logarithmic function applied to the *growth ratio*, thus we can only consider allocations that lead to positive growth ratios.

Definition 1.2.3 (Admissible portfolio)**.** We call a portfolio V admissible, if it is of the form (1.5) and for all $t \in \mathbb{T}$ we have $\mathbf{h}_{t+1}^V > 0$.

Definition 1.2.4 (Set of strictly positive portfolios)**.** The set of all admissible portfolios of the form (1.5) is called the *set of strictly positive portfolios*

$$\mathbb{V} = \{V \;\; is \;\; admissible\}.$$

[2]Note that sometimes we use $\cdot$ to ease the reading of formulas, by this symbol we denote usual scalar multiplication.

Remark 1.10. Note that in the most of realistic scenarios the set $\mathbb{V}$ contains only non-negative weights. For example, whenever we allow unbounded prices and there are no fully correlated assets, weights must be non-negative, for details see Section 4.1. Which means that the initial wealth and all components of the predictable portions process are positive. However, there are some toy examples when we can allow negative weights like markets with finite probability space Ω.

On the set of strictly positive portfolios, we can define one of the central concepts of the growth optimal portfolio theory - *growth rate* $\mathbf{g}_t^V$.

Definition 1.2.5 (Growth rate)**.** At time $t \in \mathbb{T} \setminus T$, the growth rate of portfolio $V \in \mathbb{V}$ is

$$\mathbf{g}_t^V = \mathbb{E}(\log(\mathbf{h}_{t+1}^V)|\mathcal{F}_t).$$

Given the notion of *growth rate* we can define the *optimal growth rate.*

Definition 1.2.6 (Optimal growth rate)**.** At time $t \in \mathbb{T}$ the optimal growth rate is

$$\mathbf{g}_t^*(\omega) = \sup_{V \in \mathbb{V}} \mathbf{g}_t^V(\omega), \text{ for all } \omega \in \Omega.$$

We again narrow down our analysis and try to exclude extreme cases, therefore we have another assumption

Assumption 1.2.4. *The growth rate on the market is finite*

$$\forall t \in \mathbb{T} : \mathbf{g}_t^* < \infty.$$

almost surely.

Now we can define an optimal portfolio in our framework. This corresponds to the choice of investor's *utility function*, i.e. how he or she evaluates returns of his or her portfolio. In our case, we postulate that the growth optimal portfolio coincides with a portfolio with a locally growth optimal rate.

Definition 1.2.7 (Growth optimal portfolio - GOP)**.** We call a portfolio $V^* \in \mathbb{V}$ [3] as growth optimal if

$$\mathbf{W}_0^{V^*} = 1, \tag{1.11}$$

$$\mathbf{g}_t^{V^*} = \mathbf{g}_t^*, \forall t \in \mathbb{T}, \tag{1.12}$$

and for any other portfolio $\tilde{V} \in \mathbb{V}$

$$\mathbb{E}\left(\frac{\mathbf{h}_{t+1}^{\tilde{V}}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) < \infty, \forall t \in \mathbb{T}. \tag{1.13}$$

[3]From here and below we will use star superscript * to demonstrate that variable is associated with GOP.

The first two conditions are natural, they define the scale of the portfolio and its main property. The latter condition is quite artificial and needed for technical reasons. According to Buhmann and Platen, this condition is satisfied in the most common financial models, see [5].

Remark 1.14 (On the construction of GOP on the multi-period market)*.* Note that Definition 1.2.7 involves a sequence of optimization problems, where at each step we identify new *weights* of the portfolio. By *Self-financing* Property 1.8 the wealth at a corresponding time step is reallocated according to new weights.

Our final assumption for the GOP framework is that this portfolio actually exists on the market.

Assumption 1.2.5. *There exist a portfolio $V \in \mathbb{V}$ that is growth optimal in the sense of Definition 1.2.2.*

For the readers who want to understand to which allocations GOP leads in practice, we recommend to go directly to Chapter 3. Examples where GOP has a closed form solution can be found in Sections 3.3, 3.4, 3.5.

2. Growth Optimal Portfolio properties

2.1 Growth Optimal Portfolio as Numeraire portfolio

In this section, we establish the theory that connects GOP to *Numeraire* portfolio. It turns out that with our assumptions these different notions appear to be the same concept. Let's start with the definition of Numeraire portfolio.

Definition 2.1.1 (Numeraire portfolio)**.** The portfolio V^N satisfying Definition 1.2.2 is called *Numeraire* if for any portfolio V its value process $\mathbf{W}_t^V$ benchmarked by $\mathbf{W}_t^{V^N}$ is a supermartingale

$$\frac{\mathbf{W}_t^V}{\mathbf{W}_t^{V^N}} \geq \mathbb{E}\left(\frac{\mathbf{W}_{t+1}^V}{\mathbf{W}_{t+1}^{V^N}}|\mathcal{F}_t\right).$$

Intuitively we can think of Numeraire as a best locally performing portfolio. Imagine that you invested the same capital in portfolio V and V^N, then according to the Definition 2.1.1 we expect that the value of portfolio V will be smaller compared to the V^N given the available information.

One simple example of the Numeraire portfolio is a portfolio that contains only riskless asset on the arbitrage-free market where risk-neutral martingale measure corresponds to the real-world probability measure. In this case, the benchmarked price process is a martingale and, therefore, it's a supermartingale so the condition in 2.1.1 is satisfied.

For the following proofs, it is convenient to introduce *interpolated* portfolio between GOP and some portfolio $V \in \mathbb{V}$.

Definition 2.1.2 (Interpolated portfolio)**.** Let $\theta \in (0,1)$ and $V \in \mathbb{V}$ such that $W_0^V = 1$. Then, the weights of interpolated portfolio $V^{\theta,V,*}$ at time point $t \in \mathbb{T}$ are

$$w_t^{V^{\theta,V,*}} = \theta \cdot w_t^V + (1-\theta) \cdot w_t^{V^*}.$$

Since the initial wealth is the same, we have similar equation for the wealth at time $t \in \mathbb{T}$

$$\mathbf{W}_t^{\theta,V,V^*} = \theta \cdot \mathbf{W}_t^V + (1-\theta)\cdot \mathbf{W}_t^{V^*}. \tag{2.1}$$

Then we can define the derivative of GOP in the direction to the portfolio V as follows.

Definition 2.1.3.

$$\frac{\partial \mathbf{g}_t^{\theta,V,V^*}}{\partial \theta}\Big|_{\theta=+0} = \lim_{\theta\to+0} \frac{\mathbf{g}^{\theta,V,V^*} - \mathbf{g}_t^{V^*}}{\theta}.$$

Remark 2.2. Note that by Optimal Property of GOP (1.12) we must have

$$\frac{\partial \mathbf{g}_t^{\theta,V,V^*}}{\partial \theta}\Big|_{\theta=+0} \leq 0.$$

The next theorem will present a property of GOP portfolio that will help us to link GOP with Numeraire portfolio.

Theorem 2.1.1. *For any portfolio $V \in \mathbb{V}$ we have*

$$\frac{\partial \mathbf{g}_t^{\theta,V,V^*}}{\partial \theta}\Big|_{\theta=+0} = \mathbb{E}(\frac{\mathbf{h}_{t+1}^V}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t) - 1.$$

Proof. Using the concavity of the logarithm function we obtain the following inequality

$$\forall x \in (0,\infty) : \log(x) \leq x - 1. \tag{2.3}$$

At $x = 0$ the function $f(x) = x$ is tangent to $\log(1 + x)$

$$\log(1+x)\big|_{x=0} = x\big|_{x=0} = 0,$$

$$\frac{\mathrm{d}\log(1+x)}{\mathrm{d}x}\Big|_{x=0} = \frac{\mathrm{d}x}{\mathrm{d}x}\Big|_{x=0} = 1.$$

Since $f(x) = x$ is linear and logarithm is concave ($\frac{\mathrm{d}^2\log(1+x)}{\mathrm{d}x^2} < 0$) we indeed get (2.3).

For fixed $\theta \in (0, \frac{1}{2})$, $V \in \mathbb{V}$ and $t \in \mathbb{T}$ let's consider the expression $\mathbf{G}_{t+1}^{\theta,V,V^*} := \frac{1}{\theta}\log\left(\frac{\mathbf{h}_{t+1}^{\theta,V,V^*}}{\mathbf{h}_{t+1}^{V^*}}\right)$.

Using Inequality (2.3)

$$\mathbf{G}_{t+1}^{\theta,V,V^*} = \frac{1}{\theta}\log\left(\frac{\mathbf{h}_{t+1}^{\theta,V,V^*}}{\mathbf{h}_{t+1}^{V^*}}\right) \leq \frac{1}{\theta}(\frac{\mathbf{h}_{t+1}^{\theta,V,V^*}}{\mathbf{h}_{t+1}V^*} - 1)$$

$$= \frac{1}{\theta}(\frac{(1-\theta)\mathbf{h}^{V^*}_{t+1} + \theta\mathbf{h}^{V}_{t+1} - \mathbf{h}^{V^*}_{t+1}}{\mathbf{h}^{V^*}_{t+1}}) = \frac{\mathbf{h}^{V}_{t+1}}{\mathbf{h}^{V^*}_{t+1}} - 1.$$

and by properties of the logarithm and inequality (2.3)

$$\mathbf{G}^{\theta,V,V^*}_{t+1} = -\frac{1}{\theta}\log\left(\frac{\mathbf{h}^{V^*}_{t+1}}{\mathbf{h}^{\theta,V,V^*}_{t+1}}\right) \geq -\frac{1}{\theta}\left(\frac{\mathbf{h}_{t+1}V^*}{\mathbf{h}^{\theta,V,V^*}_{t+1}} - 1\right)$$
$$= -\frac{1}{\theta}\left(\frac{\mathbf{h}^{V^*}_{t+1} - (1-\theta)\mathbf{h}^{V^*}_{t+1} - \theta\mathbf{h}^{V}_{t+1}}{\mathbf{h}^{\theta,V,V^*}_{t+1}}\right) = \frac{\mathbf{h}^{V}_{t+1} - \mathbf{h}^{V^*}_{t+1}}{\mathbf{h}^{\theta,V,V^*}_{t+1}}.$$

If $\mathbf{h}^{V}_{t+1} - \mathbf{h}^{V^*}_{t+1} > 0$, then $\mathbf{G}^{\theta,V,V^*}_{t+1} > 0$ because interpolated portfolio belongs to the set of strictly positive portfolios.

If $\mathbf{h}^{V}_{t+1} - \mathbf{h}^{V^*}_{t+1} < 0$, then

$$\mathbf{G}^{\theta,V,V^*}_{t+1} \geq -\frac{\mathbf{h}^{V^*}_{t+1}}{\mathbf{h}^{\theta,V,V^*}_{t+1}} = -\frac{1}{1-\theta+\theta\frac{\mathbf{h}^{V}_{t+1}}{\mathbf{h}^{V^*}_{t+1}}} \geq -\frac{1}{1-\theta} \geq -2.$$

Summarizing, we get

$$-2 \leq \mathbf{G}^{\theta,V,V^*}_{t+1} \leq \frac{\mathbf{h}^{V}_{t+1}}{\mathbf{h}^{V^*}_{t+1}} - 1.$$

Now, using Assumpion 1.13 and Definition 2.1, we can apply the Dominated Convergence Theorem for Conditional Expectation on $\mathbf{G}^{\theta,V,V^*}_{t+1}$ to get

$$\frac{\partial \mathbf{g}^{\theta,V,V^*}_{t}}{\partial\theta}\Big|_{\theta=+0} = \lim_{\theta\to+0}\mathbb{E}\left(\mathbf{G}^{\theta,V,V^*}_{t+1}|\mathcal{F}_t\right) = \mathbb{E}\left(\lim_{\theta\to+0}\mathbf{G}^{\theta,V,V^*}_{t+1}|\mathcal{F}_t\right)$$
$$= \mathbb{E}(\frac{\partial}{\partial\theta}\log\left(\frac{\mathbf{h}^{\theta,V,V^*}_{t+1}}{\mathbf{h}^{V^*}_{t+1}}\right)\Big|_{\theta=0}|\mathcal{F}_t) = \mathbb{E}\left(\frac{\mathbf{h}^{V}_{t+1}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t\right) - 1.$$

□

Corollary 2.4. *A portfolio $V^* \in \mathbb{V}$ such that $\mathbf{W}^{V^*}_{t_0} = 1$ is growth optimal if and only if, for any $V \in \mathbb{V}$*

$$\mathbb{E}(\frac{\mathbf{h}^{V}_{t+1}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t) \leq 1. \qquad (2.5)$$

Proof. Suppose that $V^* \in \mathbb{V}$ is GOP. Then, Inequality (2.5) holds by Remark 2.2 and Theorem 2.1.1.
Now assume that Inequality (2.5) holds and consider any $V \in \mathbb{V}$. Then,

$$\mathbf{g}_t^V - \mathbf{g}_t^{V^*} = \mathbb{E}\big(\log\Big(\frac{\mathbf{h}_{t+1}^V}{\mathbf{h}_{t+1}^{V^*}}\Big)|\mathcal{F}_t\big) \leq \mathbb{E}\big(\frac{\mathbf{h}_{t+1}^V}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\big) - 1 = \frac{\partial \mathbf{g}_t^{\theta,V,*}}{\partial \theta}\Big|_{\theta=+0} \leq 0.$$

Thus, by Remark 2.2 Conditions (1.12) and (1.13) are satisfied and V^* is GOP. □

Note that by definition of growth ratio we have

$$\mathbf{W}_t = \mathbf{W}_0 \cdot \mathbf{h}_1^V \cdot \mathbf{h}_2^V \cdot \ldots \cdot \mathbf{h}_t^V. \tag{2.6}$$

Thus, we can multiply (2.5) by $\frac{\mathbf{W}_0 \cdot \mathbf{h}_1^V \cdot \mathbf{h}_2^V \cdot \ldots \cdot \mathbf{h}_t^V}{\mathbf{W}_0^{V^*} \cdot \mathbf{h}_1^{V^*} \cdot \mathbf{h}_2^{V^*} \cdot \ldots \cdot \mathbf{h}_t^{V^*}}$ that is $\mathcal{F}_t$ measurable and get an equivalent expression

$$\mathbb{E}\Big(\frac{\mathbf{W}_{t+1}}{\mathbf{W}_{t+1}^{V^*}}|\mathcal{F}_t\Big) \leq \frac{\mathbf{W}_t}{\mathbf{W}_t^{V^*}}. \tag{2.7}$$

Therefore, by Definition 2.1.1 we see that *Growth Optimal Portfolio* and *Numeraire* portfolio are the same concept.

2.2 Optimality of the Growth Optimal Portfolio

The Growth Optimal Portfolio was defined as a certain supremum over the set of strictly positive portfolios. However, the definition involves an artificial condition expectation and doesn't tell us in what sense this portfolio is optimal. In this section, we will present several optimal properties of GOP.

Definition 2.2.1 (Long term growth rate)**.** Long term growth $\mathbf{L}^V$ of the portfolio $V \in \mathbb{V}$ is defined pathwise for all $\omega \in \Omega$ as

$$\mathbf{L}^V(\omega) = \limsup_{t\to\infty} \frac{1}{t} \log\Big(\frac{\mathbf{W}_t^V(\omega)}{\mathbf{W}_0^V(\omega)}\Big).$$

Theorem 2.2.1. *Long term growth rate of any portfolio $V \in \mathbb{V}$ is almost surely bounded by the long term growth rate of GOP. That is,*

$$\mathbf{L}^V \leq \mathbf{L}^*.$$

For the proof of this theorem, we will need the classical result by Doob for supermartingales.

Lemma 2.8 (Doob's inequality)**.** *Let* $\mathbf{X}_n$ *be a non-negative supermartingale. Then, for any* $a > 0$

$$a\mathbb{P}\big(\sup_{k\geq n} \mathbf{X}_k > a\big) \leq \mathbb{E}(\mathbf{X}_n).$$

Proof of Theorem 2.2.1. Let's consider a $V \in \mathbb{V}$ such that $\mathbf{W}_0^V = 1$. By Doob's inequality for any $\epsilon \in (0,1)$ and $k \in \mathbb{T}$

$$\exp(\epsilon k) \cdot \mathbb{P}\left(\sup_{k\leq t} \frac{\mathbf{W}_t^V}{\mathbf{W}_t^*} > \exp(\epsilon k)\right) \leq \mathbb{E}\left(\frac{\mathbf{W}_k^V}{\mathbf{W}_k^*}\right) \leq \frac{\mathbf{W}_0^V}{\mathbf{W}_0^*} = 1.$$

Which implies that

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\sup_{k\leq t}\left(\log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_t^*}\right)\right) > \epsilon k\right) \leq \sum_{k=1}^{\infty} \exp(-\epsilon k) < \infty.$$

Now by Borel-Cantelli lemma we get that

$$P\left(\limsup_{k\to\infty}\left(\sup_{k\leq t}\left(\log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_t^*}\right)\right) > \epsilon k\right)\right)$$
$$= P\left(\lim_{k\to\infty}\left(\sup_{k\leq t}\left(\log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_t^*}\right)\right) > \epsilon k\right)\right) = 0.$$

Thus, there exists k_ϵ such that for all $k > k_\epsilon$ and $t \geq k$ almost surely,

$$\log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_t^*}\right) \leq \epsilon k \leq \epsilon t.$$

Therefore,

$$\sup_{t\geq k} \frac{1}{t} \log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_t^*}\right) \leq \epsilon,$$

and

$$\sup_{t\geq k} \frac{1}{t} \log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_t^*}\frac{\mathbf{W}_0^*}{\mathbf{W}_0^V}\right) = \sup_{t\geq k}\left(\frac{1}{t}\log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_0^V}\right) - \frac{1}{t}\log\left(\frac{\mathbf{W}_t^*}{\mathbf{W}_0^*}\right)\right) \leq \epsilon.$$

Finally we get that

$$\limsup_{t\to\infty} \frac{1}{t} \log\left(\frac{\mathbf{W}_t^V}{\mathbf{W}_0^V}\right) \leq \limsup_{t\to\infty} \frac{1}{t} \log\left(\frac{\mathbf{W}_t^*}{\mathbf{W}_0^*}\right) + \epsilon.$$

Since the inequality holds for any $\epsilon \in (0,1)$ we have that

$$\mathbf{L}^V \leq \mathbf{L}^*.$$

□

The next optimal property of GOP ensures that no other portfolio can systemically outperform GOP. The systematic outperformance is closely connected to the concept of stochastic dominance, namely the next theorem actually tells us that once GOP and some other portfolio started with the same initial capital, then there is no time point where the other portfolio's value stochastically dominates GOP's value.

Definition 2.2.2 (Systematic Outperformance)**.** A portfolio $V^1 \in \mathbb{V}$ systematically outperforms a portfolio $V^2 \in \mathbb{V}$ if

1. $\mathbf{W}_0^{V^1} = \mathbf{W}_0^{V^2}$ and,
2. $\exists t : \mathbb{P}(\mathbf{W}_t^{V^1} \geq \mathbf{W}_t^{V^2}) = 1$ and $\mathbb{P}(\mathbf{W}_t^{V^1} > \mathbf{W}_t^{V^2}) > 0$.

Theorem 2.2.2. *Growth Optimal Portfolio cannot be systematically outperformed by any portfolio from the set of strictly positive portfolios.*

Proof. Let's consider a portfolio $V \in \mathbb{V}$. By Inequality (2.7), we see that the benchmarked (i.e. divided by the value of GOP) value process satisfies the supermartingale property, therefore for $s \in \mathbb{T} \setminus 0$ we can write

$$\mathbb{E}\left(\frac{\mathbf{W}_s^V}{\mathbf{W}_s^*} - \frac{\mathbf{W}_0^V}{\mathbf{W}_0^*}|\mathcal{F}_0\right) = \mathbb{E}\left(\frac{\mathbf{W}_s^V}{\mathbf{W}_s^*} - 1\right) \leq 0. \tag{2.9}$$

On the other hand, Condition 2 of systematic outperformance implies that

$$\mathbb{P}(\frac{\mathbf{W}_s^V}{\mathbf{W}_s^*} - 1 \geq 0) = 1$$

and

$$\mathbb{P}(\frac{\mathbf{W}_s^V}{\mathbf{W}_s^*} - 1 > 0) > 0.$$

Therefore we have a lower bound estimate

$$\mathbb{E}(\frac{\mathbf{W}_s^V}{\mathbf{W}_s^*} - 1) = \int_\omega (\frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1) dP(\omega)$$

$$= \int_{\omega: \frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1 > 0} \big(\frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1\big) dP(\omega) + \int_{\omega: \frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1 \leq 0} \big(\frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1\big) dP(\omega) =$$

$$= \int_{\omega: \frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1 > 0} \big(\frac{\mathbf{W}_s^V(\omega)}{\mathbf{W}_s^*(\omega)} - 1\big) dP(\omega) > 0$$

that contradicts Inequality(2.9).

□

There are many other optimality properties of GOP. For example, that the ruin probability of GOP is zero or that GOP needs the shortest expected time to reach a given wealth level. More properties of GOP can be found in [3] and [8].

2.3 Connection between GOP and arbitrage freedom

In this section, we will examine a very general case when GOP exists on the market. Let's start with the usual definition of arbitrage.

Definition 2.3.1 (Arbitrage strategy)**.** Let's consider a portfolio $V \in \mathbb{V}$. An arbitrage opportunity exists if the initial wealth is zero, and at some point we can make money out of nothing

$$\mathbf{W}_0^V = 0 \tag{2.10}$$

and

$$\exists T > 0 : \mathbb{P}(\mathbf{W}_T^V \geq 0) = 1, \quad \mathbb{P}(\mathbf{W}_T^V > 0) > 0. \tag{2.11}$$

It turns out that the absence of the arbitrage is equivalent to existence of GOP under certain conditions.

Theorem 2.3.1. *If GOP exists on the market, then the market is arbitrage free.*

Proof. Let's assume that there exists a growth optimal portfolio $V^* \in \mathbb{V}$ but there exists an arbitrage opportunity which can be achieved by portfolio V^a. Note that for V^a we can't apply previous theorems because from the definition of arbitrage it does not belong to the set of strictly positive portfolios.
To fix this let's consider a portfolio $V^o = V^* + V^a$. V^o also belongs to the set $\mathbb{V}$ because V^* is admissible and the growth ratio of the sum is a sum of growth ratios by Property (1.9).

Then by the arbitrage property we have

$$\mathbf{W}^o_0 = \mathbf{W}^*_0,$$

and there exist T such that:

$$1 = \mathbb{P}(\mathbf{W}^a_T \geq 0) = \mathbb{P}(\mathbf{W}^a_T + \mathbf{W}^*_T \geq \mathbf{W}^*_T) = \mathbb{P}(\mathbf{W}^o_T \geq \mathbf{W}^*_T)$$

Which means that

$$0 < \mathbb{P}(\mathbf{W}^a_T > 0) = \mathbb{P}(\mathbf{W}^o_T > \mathbf{W}^*_T).$$

Which contradicts to Theorem 2.2.2 that states the absence of systematic outperformance on the market. □

In Christensen's survey [3] Theorem 1.1 states that the existence of GOP is equivalent to the absence of arbitrage. Unfortunately, there is no detailed proof but it is mentioned in the text that portfolio with Numeraire property 2.1.1 can be somehow constructed from the martingale risk-neutral measure. I could not find any exhaustive proofs for sufficiency in the discrete time setting. A comprehensive proof of this linkage is given in [9], however only the one-period market with finite probability space is considered. The proof can be easily extended to the multiperiod market since the absence of arbitrage on the multiperiod is equivalent to the absence of arbitrage in all consecutive sub-one-period markets and the construction of GOP is done for each time step independently. I suppose that in the discrete time more assumptions on the market structure are needed in order to achieve the existence of GOP. For example, the proof can be completed assuming that the market is *complete*.

Theorem 2.3.2. *GOP exists on the market if the market is arbitrage free and complete.*

Proof. Let's consider an asset $i \in \overline{0, N}$. Absence of arbitrage guarantees us that there exists an equivalent risk-neutral measure Q to real-world probability measure P, under which the discounted price process is a martingale, i.e. for any $t \in \mathbb{T}$

$$\mathbb{E}_Q\left(\frac{\mathbf{S}_{t+1,i}}{\mathbf{S}_{t+1,0}}|\mathcal{F}_t\right) = \frac{\mathbf{S}_{t,i}}{\mathbf{S}_{t,0}}. \tag{2.12}$$

Let's denote $\mathbf{Z}_t := \mathbb{E}_P\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}|\mathcal{F}_t\right)$, using Abstract Bayes' Theorem [7] (Prop. B.41) and properties of conditional expectation we can rewrite the right hand side as

$$\mathbb{E}_Q\left(\frac{\mathbf{S}_{t+1,i}}{\mathbf{S}_{t+1,0}}|\mathcal{F}_t\right) = \frac{\mathbb{E}_P\left(\frac{\mathbf{S}_{t+1,i}}{\mathbf{S}_{t+1,0}}\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}|\mathcal{F}_t\right)}{\mathbb{E}_P\left(\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}|\mathcal{F}_t\right)}$$

$$= \frac{\mathbb{E}_P \left(\frac{\mathbf{S}_{t+1,i}}{\mathbf{S}_{t+1,0}} \mathbf{Z}_{t+1} | \mathcal{F}_t \right)}{\mathbf{Z}_t} = \mathbb{E}_P \left(\frac{\mathbf{S}_{t+1,i}}{\mathbf{S}_{t+1,0}} \frac{\mathbf{Z}_{t+1}}{\mathbf{Z}_t} | \mathcal{F}_t \right).$$

So that Equation (2.12) becomes

$$\mathbb{E}_P \left(\mathbf{S}_{t+1,i} \frac{\mathbf{S}_{t,0}}{\mathbf{S}_{t+1,0}} \frac{\mathbf{Z}_{t+1}}{\mathbf{Z}_t} | \mathcal{F}_t \right) = \mathbf{S}_{t,i}. \tag{2.13}$$

Since the market is complete we can find the replicating portfolio for the claim $\frac{\mathbf{S}_{t,0}}{\mathbf{Z}_t}$, i.e. there exists a sequence of portfolio weights $(w_0^V, w_1^V, ..., w_T^V)$ such that at any $t \in \mathbb{T}$

$$\frac{\mathbf{S}_{t,0}}{\mathbf{Z}_t} = \mathbf{W}_t^V = \sum_{i=0}^{N} w_{t,i}^V \mathbf{S}_{t,i} \tag{2.14}$$

Therefore, we can rewrite (2.13) as

$$\mathbb{E}_P \left(\mathbf{S}_{t+1,i} \frac{\mathbf{W}_t^V}{\mathbf{W}_{t+1}^V} | \mathcal{F}_t \right) = \mathbf{S}_{t,i}$$

and finally

$$\mathbb{E}_P \left(\frac{\mathbf{S}_{t+1,i}}{\mathbf{W}_{t+1}^V} | \mathcal{F}_t \right) = \frac{\mathbf{S}_{t,i}}{\mathbf{W}_t^V}. \tag{2.15}$$

The replicating portfolio does not depend on the selected asset, so Equation (2.15) holds for all $i \in \overline{0, N}$. By linearity of conditional expectation we have the same martingale equation for any portfolio $\tilde{V} \in \mathbf{V}$

$$\mathbb{E}_P \left(\frac{\mathbf{W}_{t+1}^{\tilde{V}}}{\mathbf{W}_{t+1}^V} | \mathcal{F}_t \right) = \frac{\mathbf{W}_t^{\tilde{V}}}{\mathbf{W}_t^V}. \tag{2.16}$$

Thus, GOP in this case exists as a replicating portfolio (2.14) and it is indeed growth optimal by martingale property (2.16) and Corollary 2.4.

□

3. Growth Optimal Portfolio examples

Even though the discrete market model and the optimization problem stated in Definition 1.2.7 may sound simple, the problem of finding the actual weights of GOP is often unfeasible and surprisingly more complicated than in the continuous setting, see [3]. On the other hand, once we are given a portfolio candidate, we can easily check if it is growth optimal by the *Numeraire* Property (2.7). In this Chapter, I will consider different examples where GOP can be found explicitly and then provide a framework for empirical construction of GOP for a broad class of growth ratios' time series models.

3.1 Domain restrictions for portfolio weights

The heart of the problem of the actual construction of GOP lies in the logarithmic function, that allows us to work only with positive arguments. From the first perspective, the restriction of analysis to the set of strictly positive portfolios does not affect our solution at all because if the solution for maximization problem exists, ${\mathbf{h_t}}^{V^*}$ is positive and the corresponding restrictions on the portfolio weights are not active. However, in most of the cases, we cannot easily map the set of strictly positive portfolios to the respective set of feasible portfolio weights. This map actually heavily depends on the distribution assumptions of the ${\mathbf{h_t}}^{V^*}$.

For instance, we can consider two independent assets. Whenever we allow unbounded distributions for $\mathbf{h}_{t,i}$ our strictly positive restriction implies that possible weights are always positive. To see this, let's consider a one period market with two securities. Let $w_{t,0} < 0$ then we can find $\omega = \{\omega_1, \omega_2\}$ such that $w_{t,0}\mathbf{h}_{t,0}(\omega_1) + w_{t,1}\mathbf{h}_{t,1}(\omega_2)$ is negative. Therefore, we cannot compute the logarithm for this element of probability space and the expectation operation is undefined as well.

The set of possible portfolio weights can be identified only when we are given exact distributions of $\mathbf{h}_{t,i}$. While without those restrictions we cannot apply the Karush-Kuhn-Tucker theorem [10] in order to find necessary conditions for optimality.

3.2 Necessary and Sufficient condition for the case of unrestricted portfolio weights

Let's consider Optimization problem 1.2.6 and restate it in terms of portfolio weights

$$\mathbf{g}_t^* = \sup_{V \in \mathbb{V}} \mathbb{E}\left(\log\left(\mathbf{h}_{t+1}^V\right)|\mathcal{F}_t\right) = \sup_{w_t^V \in \mathbb{W}_t} \mathbb{E}\left(\log\left(\sum_{i=0}^{N} w_{t,i}^V \mathbf{h}_{t+1,i}\right)|\mathcal{F}_t\right), \tag{3.1}$$

where the set of possible weights $\mathbb{W}_t$ is given by

$$\mathbb{W}_t = \{w_t^V \in \mathcal{F}_{t-1} | \mathbf{h}_t^V > 0 \text{ and } \sum_{i=0}^{N} w_{t,i}^V = 1\}. \tag{3.2}$$

Assumption 3.2.1. *Let's assume that the restriction $\mathbf{h}_t^V > 0$ does not affect the solution of the optimization problem* (3.1)*, so we can instead optimize over*

$$\mathbb{W}_t = \{w_t^V \in \mathcal{F}_{t-1} | \sum_{i=0}^{N} w_{t,i}^V = 1\}.$$

Let's denote $\tilde{w}_t^V = (w_{t,1}^V, w_{t,2}^V, ..., w_{t,N}^V) \in \mathcal{F}_{t-1}$, then for any $\omega \in \Omega$ we can rewrite (3.1) as

$$\begin{aligned}
\mathbf{g}_t^*(\omega) &= \sup_{w_t^V \in \mathbb{W}_t} \mathbb{E}\left(\log\left(\sum_{i=0}^{N} w_{t,i}^V \mathbf{h}_{t+1,i}\right)|\mathcal{F}_t\right)(\omega) \\
&= \sup_{\tilde{w}_t^V \in \mathbb{R}^N} \mathbb{E}\left(\log\left(\left(1 - \sum_{i=1}^{N} w_{t,i}^V\right)\mathbf{h}_{t+1,0} + \sum_{i=1}^{N} w_{t,i}^V \mathbf{h}_{t+1,i}\right)|\mathcal{F}_t\right)(\omega).
\end{aligned}$$

Theorem 3.2.1 (Optimality condition in the case of unrestricted domain)**.** *Under Assumption 3.2.1, the portfolio $V^* \in \mathbb{V}$ is growth optimal if and only if for any $t \in \mathbb{N}$*

$$\forall i : \mathbb{E}\left(\frac{\mathbf{h}_{t+1,i}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) = 1 \quad a.s. \tag{3.3}$$

Proof. The logarithm is a concave function and inside the logarithm, we have a linear function of weights, which means that the composition of the functions is concave too. Thus, for every fixed $\omega \in \Omega$ we have a unique maximum.

The function inside the conditional expectation has the following properties.

1. For all $w_t^V \in \mathcal{F}_{t-1}$ we have a $\mathcal{F}_{t+1}$ measurable r.v. inside the expectation.
2. By Inequality 2.3 and Jensen's inequality we have

$$\mathbb{E}\left(\log\left(\mathbf{h}_{t+1}^V\right)\middle|\,\mathcal{F}_t\right) \leq \log \mathbb{E}\left(\mathbf{h}_{t+1}^V\middle|\,\mathcal{F}_t\right) \leq \mathbb{E}\left(\mathbf{h}_{t+1}^V\middle|\,\mathcal{F}_t\right).$$

Therefore, by Dominated Convergence Theorem, we can interchange the order of taking a derivative and expectation as follows. Let's consider a partial derivative for fixed $j \in \{1, ..., N\}$ at the optimal weights

$$\begin{aligned}
&\frac{\partial \mathbb{E}\left(\log\left(\left(1-\sum_{i=1}^N w_{t,i}^{V^*}\right)\mathbf{h}_{t+1,0}+\sum_{i=1}^N w_{t,i}^{V^*}\mathbf{h}_{t+1,i}\right)|\mathcal{F}_t\right)}{\partial w_{t,j}^{V^*}} \\
&= \mathbb{E}\left(\frac{\partial \log\left(\left(1-\sum_{i=1}^N w_{t,i}^{V^*}\right)\mathbf{h}_{t+1,0}+\sum_{i=1}^N w_{t,i}^{V^*}\mathbf{h}_{t+1,i}\right)}{\partial w_{t,i}^{V^*}}|\mathcal{F}_t\right) \\
&= \mathbb{E}\left(\frac{\mathbf{h}_{t+1,j}-\mathbf{h}_{t+1,0}}{\left(1-\sum_{i=1}^N w_{t,i}^{V^*}\right)\mathbf{h}_{t+1,0}+\sum_{i=1}^N w_{t,i}^{V^*}\mathbf{h}_{t+1,i}}|\mathcal{F}_t\right) \\
&= \mathbb{E}\left(\frac{\mathbf{h}_{t+1,j}-\mathbf{h}_{t+1,0}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) = 0.
\end{aligned}$$

Now let's multiply the last expression by $w_{t,j}^{V^*}$ and sum over j

$$\begin{aligned}
&\sum_{j=1}^N w_{t,j}^{V^*}\cdot\mathbb{E}\left(\frac{(\mathbf{h}_{t+1,j}-\mathbf{h}_{t+1,0})}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) \\
&= \mathbb{E}\left(\frac{\sum_{j=1}^N w_{t,j}^{V^*}(\mathbf{h}_{t+1,j}-\mathbf{h}_{t+1,0})}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) + \mathbb{E}\left(\frac{\mathbf{h}_{t+1,0}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) - \mathbb{E}\left(\frac{\mathbf{h}_{t+1,0}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) \\
&= \mathbb{E}\left(\frac{(1-\sum_{j=1}^N w_{t,j}^{V^*})\mathbf{h}_{t+1,0}+\sum_{j=1}^N w_{t,j}^{V^*}\mathbf{h}_{t+1,j})}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) - \mathbb{E}\left(\frac{\mathbf{h}_{t+1,0}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) \\
&= 1-\mathbb{E}\left(\frac{\mathbf{h}_{t+1,0}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) = 0
\end{aligned} \tag{3.4}$$

Therefore we have for the asset 0

$$\mathbb{E}\left(\frac{\mathbf{h}_{t+1,0}}{\mathbf{h}_{t+1}^{V^*}}|\mathcal{F}_t\right) = 1.$$

Summing this with (3.4) we get the same condition for all assets

$$\mathbb{E}\left(\frac{\mathbf{h}_{t+1,0}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t\right) + \mathbb{E}\left(\frac{\mathbf{h}_{t+1,j} - \mathbf{h}_{t+1,0}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t\right) = 1 + 0,$$

so that

$$\text{For all j } \in \{1, ..., N\} : \mathbb{E}\left(\frac{\mathbf{h}_{t+1,j}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t\right) = 1. \tag{3.5}$$

□

Remark 3.6. Note that in this theorem the first asset is used just for the convenience of the proof. We did not assume that this asset has deterministic returns' structure.

Corollary 3.7. *Under Assumption 3.2.1 the benchmarked value process of any admissible portfolio is a martingale in natural filtration* $\mathcal{F}_t$

$$\mathbb{E}\left(\frac{\mathbf{W}_{t+1}}{\mathbf{W}^{V^*}_{t+1}}|\mathcal{F}_t\right) = \frac{\mathbf{W}_t}{\mathbf{W}^{V^*}_t}.$$

Proof. Let's consider some portfolio $V \in \mathbb{V}$ with corresponding weights' vector w^V_t. For all $i \in \{0, .., N\}$ multiply (3.5) by the corresponding portfolio weight $w^{V^*}_{t,i}$ and sum up Conditions (3.5).

$$\sum_{i=0}^{N} w^V_{t,i} \cdot \mathbb{E}\left(\frac{\mathbf{h}_{t+1,i}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t\right) = \sum_{i=0}^{N} w^V_{t,i} = 1$$

$$\mathbb{E}\left(\frac{\mathbf{h}^V_{t+1}}{\mathbf{h}^{V^*}_{t+1}}|\mathcal{F}_t\right) = 1.$$

We can multiply (2.5) by $\frac{\mathbf{W}_0 \cdot \mathbf{h}^V_1 \cdot \mathbf{h}^V_2 \cdots \mathbf{h}^V_t}{\mathbf{W}^{V^*}_0 \cdot \mathbf{h}^{V^*}_1 \cdot \mathbf{h}^{V^*}_2 \cdots \mathbf{h}^{V^*}_t}$ that is $\mathcal{F}_t$ measurable and get

$$\mathbb{E}\Big(\frac{\mathbf{W}_{t+1}}{\mathbf{W}^*_{t+1}}|\mathcal{F}_t\Big) = \frac{\mathbf{W}_t}{\mathbf{W}^*_t}. \tag{3.8}$$

□

The martingale property of the growth optimal portfolio under Condition 3.2.1 leads to the concept that is called *fair pricing*. Namely, similarly to the risk-neutral measure pricing, we have a martingale that allows us to define the price given the distribution of GOP and underlying assets. For further details see [5].

3.3 Example 1: Betting

Let's consider the gambling game where we have a countable set of possible outcomes $\Omega = \cup_{i\in\mathbb{I}}\{\omega_i\}$, where $i \in \mathbb{I}$ - is a countable set. A player chooses an outcome and receives α times his original bet if he guessed right and nothing otherwise.

A smart player who wants to play the game optimally in the sense of growth optimal portfolio tries to find weights that correspond to GOP in this game or, in other words, he allocates his wealth across different outcomes.

Let's define sets $A_i = \{\omega_i\}$, then they partition the outcomes space $\Omega = \cup_{i\in\mathbb{I}} A_i$. From the portfolio theory perspective, we have one period market and a set of assets (bets on outcomes) where each asset generates a certain return. Namely, the growth rate of the outcome i is

$$\mathbf{h}_i = \alpha \cdot \mathbb{1}_{A_i}.$$

The growth rate of portfolio is

$$\mathbf{h}^V = \sum_{i\in\mathbb{I}} w_i \cdot \alpha \cdot \mathbb{1}_{A_i},$$

where $(w_0, w_1, ..., w_N)$ are the portfolio weights.

Now, let's compute GOP weights by applying Theorem 3.2.1

$$1 = \mathbb{E}\left(\frac{\mathbf{h}_i}{\mathbf{h}^{V^*}}\right) = \mathbb{E}\left(\frac{\alpha \cdot \mathbb{1}_{A_i}}{\sum_{i\in\mathbb{I}} w_i^{V^*} \cdot \alpha \cdot \mathbb{1}_{A_i}}\right) = \frac{\mathbb{P}(A_i)}{w_i^{V^*}}.$$

Therefore, the optimal weights are equal to the probabilities of the outcomes

$$w_i^{V^*} = \mathbb{P}(A_i). \qquad (3.9)$$

3.4 Example 2: Complete markets with countable probability space

Now let's consider again a one-period market with a countable probability space $\Omega = \cup_{i\in\mathbb{I}}\{\omega_i\}$. We assume that the market is *complete* meaning that all claims are attainable by some portfolios of original securities. We can interpret this market as a set of indicator claims, that are defined as

$$\tilde{S}_{1,i}(\omega) = \mathbb{1}(\omega = \omega_i),$$

where $i \in \mathbb{I}$.

Since the market is complete, we can replicate these claims by portfolios $\tilde{V}_i$

$$\mathbb{1}(\omega = \omega_i) = \sum_{j=0}^{N} w_j^{\tilde{V}_i} \cdot S_{1,i}.$$

Now we are in the setting of Section 3.3, to wit, the investor has a set of mutually exclusive alternatives $\{\tilde{S}_{1,i}\}_{i\in\mathbb{I}}$ and $\alpha = 1$. Therefore, we already know the weights of GOP in terms of indicators from the previous section result (3.4)

$$\begin{aligned} W_t^{V^*} &= \sum_{i\in\mathbb{I}} \mathbb{P}(\omega = \omega_i)\tilde{S}_{1,i} = \sum_{i\in\mathbb{I}} \mathbb{P}(\omega = \omega_i) \sum_{j=0}^{N} w_j^{\tilde{V}_i} \cdot S_{1,i} \\ &= \sum_{j=0}^{N} \left[\sum_{i\in\mathbb{I}} \mathbb{P}(\omega = \omega_i) w_j^{\tilde{V}_i} \right] S_{1,i}. \end{aligned}$$

Thus, GOP weights can be computed as

$$w_j^{V^*} = \sum_{i\in\mathbb{I}} \mathbb{P}(\omega_i \in \Omega) w_j^{\tilde{V}_i}.$$

3.5 Example 3: A bond and a risky asset with log-normally distributed growth ratio

Let's consider a bit more complicated case where we will see exactly how the restrictions on the portfolio weights affect the martingale properties of GOP. Now we are considering the multiperiod market with two securities where the first security is a bond with a constant price $S_{t,0} = 1$ for all $t \in \mathbb{T}$ and the second one is a risky asset with i.i.d growth ratio such that for all $t \in \mathbb{T}$, $h_{t,1} \in (0, \infty)$. Since we are considering the set of strictly positive portfolios, as described at the beginning of this chapter, the weights should be strictly positive. As far as weights sum up to one for $V \in \mathbb{V}$ we have

$$w_{t,1}^V \in [0, 1],$$

and

$$w_{t,1}^V = 1 - w_{t,0}^V.$$

The growth rate is then

$$\mathbf{g}_t^V = \mathbb{E}\left(\log\left(1 + w_{t,1}^V(\mathbf{h}_{t+1,1} - 1)|\mathcal{F}_t\right)\right).$$

As discussed in the proof of Theorem 3.2.1 we can pass the partial derivative inside the conditional expectation. The first order condition becomes

$$\frac{\partial \mathbf{g}_t^V}{\partial w_{t,1}^V} = \mathbb{E}\left(\frac{\mathbf{h}_{t+1,1}-1}{1+w_{t,1}^V(\mathbf{h}_{t+1,1}-1)}|\mathcal{F}_t\right) = 0.$$

The second partial derivative is then always negative

$$\frac{\partial^2 \mathbf{g}_t^V}{\partial {w_{t,1}^V}^2} = -\mathbb{E}\left(\frac{(\mathbf{h}_{t+1,1}-1)^2}{(1+w_{t,1}^V(\mathbf{h}_{t+1,1}-1))^2}|\mathcal{F}_t\right) \leq 0,$$

which implies that the growth rate is concave. Let's compute the derivative on the boundaries of weights' interval

$$\frac{\partial \mathbf{g}_t^V}{\partial w_{t,1}^V}|_{w_{t,1}^V=0} = \mathbb{E}\left(\mathbf{h}_{t+1,1}|\mathcal{F}_t\right) - 1,$$

$$\frac{\partial \mathbf{g}_t^V}{\partial w_{t,1}^V}\Big|_{w_{t,1}^V=1} = 1 - \mathbb{E}\left(\frac{1}{\mathbf{h}_{t+1,1}}|\mathcal{F}_t\right).$$

Since the second partial derivative is not positive, the first derivative is decreasing. Therefore, for the first order condition to have a solution the following condition should be satisfied

$$\mathbb{E}\left((\mathbf{h}_{t+1,1})^\lambda|\mathcal{F}_t\right) \geq 1, \text{ for } \lambda = \pm 1. \tag{3.10}$$

Otherwise, if (3.10) is violated then the maximum is obtained on the boundaries.

Now let's consider a case when growth ratios are independently identically distributed log-normal variables

$$\log(\mathbf{h}_{t,1}) \sim \mathcal{N}(\mu, \sigma^2).$$

Given this information we can compute the expectation in (3.10)

$$\begin{aligned}
\mathbb{E}\left((\mathbf{h}_{t+1,1})^\lambda|\mathcal{F}_t\right) &= \mathbb{E}\left((\mathbf{h}_{t+1,1})^\lambda\right) \\
&= \int_{-\infty}^{\infty} e^{\lambda x}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x^2-2x\mu+\mu^2-2\sigma^2\lambda x)}{2\sigma^2}}dx \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x^2-2x(\mu-\sigma^2\lambda)+(\mu-\sigma^2\lambda)^2-(\mu-\sigma^2\lambda)^2+\mu^2)}{2\sigma^2}}dx
\end{aligned}$$

$$= e^{\frac{\lambda^2\sigma^2}{2}+\lambda\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-(\mu-\sigma^2\lambda))^2}{2\sigma^2}} dx = e^{\frac{\lambda^2\sigma^2}{2}+\lambda\mu}.$$

Now we can solve (3.10). We require that

$$e^{\frac{\lambda^2\sigma^2}{2}+\lambda\mu} \geq 1, \text{ for } \lambda = \pm 1.$$

Thus

$$e^{\frac{\sigma^2}{2}+\mu} \geq 1 \wedge e^{\frac{\sigma^2}{2}-\mu} \geq 1,$$
$$\frac{\sigma^2}{2} + \mu \geq 0 \wedge \frac{\sigma^2}{2} - \mu \geq 0$$

are satisfied. To summarize,

$$\frac{|\mu|}{\sigma^2} \leq \frac{1}{2}. \tag{3.11}$$

Inequality (3.11) shows us under which condition solution to Optimisation Problem 3.1 lies inside the domain. Whenever this inequality holds we are in the setting of Theorem 3.2.1 and Corollary 3.7 holds implying that benchmarked price of each asset is a martingale.

If Inequality 3.11 is violated we have two possible cases. The first one $\frac{\mu}{\sigma^2} > \frac{1}{2}$ corresponds to the situation when the risky asset performs extremely well and, therefore, in GOP all capital is allocated to this asset. The benchmarked bond's price becomes a strict supermartingale, while benchmarked stock's price always equals one. On the contrary, if $\frac{\mu}{\sigma^2} < -\frac{1}{2}$, then the stock significantly underperforms and all wealth in GOP is allocated to the bond. In this case, the stock's benchmarked price is a strict supermartingale, while bond's benchmarked price is always one.

This simple example with two assets shows us that we cannot always apply Theorem 3.2.1 to compute GOP's weights and demonstrates the source of complexity in GOP discrete time construction. Moreover, it provides an economic intuition regarding possible violations to the *fair pricing* mentioned above. Namely, the fair pricing breaks down if the growth of risky asset is quite low comparing to the risk-neutral asset or when the market price of risk is too high.

4. Construction of GOP in the case of Log-Normally distributed growth ratios

In this chapter, we will consider a model that can be applied to construct Growth Optimal portfolio on the real market. The usual way of building GOP involves continues time models and further approximations. For example, Platen shows that GOP can be approximated by the well-diversified portfolio, see Chapter 10 at [8] or [11]. The other approach is to assume a certain distribution of the returns, then use Monte Carlo approximation for the target function and finally solve this problem directly by the suitable general optimization method. The example with stationary log-normally distributed returns can be found in Maier's paper, see [13]. However, Maier constructs the optimal portfolio that has an unconditional expectation as a target function. This setting corresponds to the one-period market and, obviously, on a multi-period market the solution will not have attractive properties that were described in Chapter 2 even if the distribution assumption is valid.

4.1 Log-normality assumption

In this section, we will consider a model-based approach of constructing GOP. This means that we need to make some assumptions on the growth ratios' process. Our final goal is to construct GOP for a set of securities that can be modelled by the common time-series models such as ARIMA and VAR, which will be specified in the following chapter. The main assumption for using those models is a stationarity assumption

Assumption 4.1.1. *The logarithm of growth ratio of each risky security on the market forms a stationary process with normally distributed components.*

One can argue that this assumption is too artificial and we cannot really model the securities on the market with such simple models. Nevertheless, the most famous option pricing Black-Scholes-Merton model implies that the growth ratio of a stock is log-normally distributed

$$\frac{S_{t+\Delta t}}{S_t} \sim \exp\Big(N((\mu - \frac{1}{2}\sigma^2)\Delta t, \sigma^2 \Delta t) \Big).$$

Moreover, there is a variety of papers where scientists successfully model different stocks by above-mentioned time-series models, for examples see [14], [15], [16]. The fact that these models actually can explain certain stocks does not automatically imply that one can abuse them in order to obtain an arbitrage. On the contrary, in most of the cases, confidence intervals for predictions are huge and eliminate the possibility of arbitrage operations.

Finally, we want to emphasize that by utilizing these time-series models we do not aim to obtain just a best possible prediction for the growth ratio, rather than we try to use the current information on the market to assess the conditional distribution that would allow us to solve Optimization Problem 1.2.6.

4.2 GOP optimization problem for time-series models

In this section, we will specify the model the general setting while in the following chapter we will provide specific models that we used for stocks modelling. For simplicity we are considering the case when the market only consists of N risky securities, the risk-less asset will be added in the following section as an extension to this model. For convenience, let's firstly denote the vector of growth ratios as

$$\mathbf{h}_t = (\mathbf{h}_{t,1}, \mathbf{h}_{t,2}, ..., \mathbf{h}_{t,N}).$$

Assumption 4.2.1. *We assume that for all $t \in \mathbb{T}$ error vectors ε_t are i.i.d. and follow Multivariate Normal distribution with zero mean*

$$\varepsilon_t = (\boldsymbol{\varepsilon}_{t,1}, \boldsymbol{\varepsilon}_{t,2}, ..., \boldsymbol{\varepsilon}_{t,N}) \sim N(0, \Sigma),$$

where

$$\varepsilon_t \in \mathcal{F}_t,$$

and Σ - covariance matrix, i.e. parameter of Multivariate Normal distribution.

Given this assumption, the general time series model setting is the following

$$\text{For all } i \in \overline{1,N}, t \in \mathbb{T} : \log(\mathbf{h}_{i,t}) = f_i(\mathbf{h}_{t-1}, \mathbf{h}_{t-2}, ..., \mathbf{h}_{t-k}, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, ..., \boldsymbol{\varepsilon}_{t-l}) + \boldsymbol{\varepsilon}_{t,i},$$

where k, l are predefined number of lags in the model and f_i don't depend on t.

Now let's plug in this model into Optimization Problem 1.2.6 at time t

$$\mathbf{g}_t^* = \sup_{V\in\mathbb{V}} \mathbf{g}_t^V = \sup_{V\in\mathbb{V}} \mathbb{E}\left(\log\left(\mathbf{h}_{t+1}^V\right)|\mathcal{F}_t\right) = \sup_{w_t^V\in\mathbb{W}_t} \mathbb{E}\left(\log\left(\sum_{i=1}^N w_{t,i}^V \mathbf{h}_{t+1,i}\right)|\mathcal{F}_t\right)$$

(4.1)

$$= \sup_{w_t^V\in\mathbb{W}_t} \mathbb{E}\left(\log\left(\sum_{i=1}^N w_{t,i}^V \exp(f_i\left(\mathbf{h}_t, \mathbf{h}_{t-1}, ..., \mathbf{h}_{t-k+1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, ..., \boldsymbol{\varepsilon}_{t-l+1}\right) + \boldsymbol{\varepsilon}_{t+1,i})\right)|\mathcal{F}_t\right).$$

Now note that $f_i\left(\mathbf{h}_t, \mathbf{h}_{t-1}, ..., \mathbf{h}_{t-k+1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, ..., \boldsymbol{\varepsilon}_{t-l+1}\right) \in \mathcal{F}_t$ by Assumption 4.2.1 and Definition 1.2.1. It means that at time t, we already know realized values of $\mathbf{h}_t = \hat{h}_t, \mathbf{h}_{t-1} = \hat{h}_{t-1}, ..., \mathbf{h}_{t-k+1} = \hat{h}_{t-k+1}$ and $\boldsymbol{\varepsilon}_t = \hat{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1} = \hat{\varepsilon}_{t-1}, ..., \boldsymbol{\varepsilon}_{t-l+1} = \hat{\varepsilon}_{t-k+1}$. On the other hand we have that $\boldsymbol{\varepsilon}_{t+1,i} \perp\!\!\!\perp \mathcal{F}_t$, therefore, we actually have an optimization problem with unconditional expectation and (4.1) can be restated as

(4.2)

$$\mathbf{g}_t^* = \sup_{w_t^V\in\mathbb{W}_t} \mathbb{E}\left(\log\left(\sum_{i=1}^N w_{t,i}^V \exp\left(f_i\left(\hat{h}_t, \hat{h}_{t-1}, ..., \hat{h}_{t-k+1}, \hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, ..., \hat{\varepsilon}_{t-l+1}\right) + \boldsymbol{\varepsilon}_{t+1,i}\right)\right)\right).$$

If $X \sim N(\mu, \sigma)$ then $X + a \sim N(\mu + a, \sigma)$. So we can define a new random vector $\mathbf{Y}_t = (\mathbf{Y}_{t,1}, \mathbf{Y}_{t,2}, ..., \mathbf{Y}_{t,N}) \sim N(\mu, \Sigma)$ as follows

$$\text{For all } i \in \overline{1, N}: \quad \mathbf{Y}_{t,i} = f_i\left(\hat{h}_t, \hat{h}_{t-1}, ..., \hat{h}_{t-k+1}, \hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, ..., \hat{\varepsilon}_{t-l+1}\right) + \boldsymbol{\varepsilon}_{t+1,i}. \tag{4.3}$$

Following this, we can rewrite Problem 4.2 as

$$\mathbf{g}_t^* = \sup_{w_t^V\in\mathbb{W}_t} \mathbb{E}\left(\log\left(\sum_{i=1}^N w_{t,i}^V \exp(\mathbf{Y}_{t,i})\right)\right). \tag{4.4}$$

Domain $\mathbb{W}_t$ of the optimization problem, defined at 3.2, in this case correspond to non-negative or, following financial jargon, long-only portfolio weights, as it is explained in Section 3.1. Thus

$$\mathbb{W}_t = \{w_t^V \in \mathcal{F}_{t-1} | \forall i \;\; w_{t,i}^V \geq 0 \text{ and } \sum_{i=1}^N w_{t,i}^V = 1\}. \tag{4.5}$$

The exact density of the sum of Log-normally distributed random variables is hard to derive even in the simplest independent case. Another common approach is to use

Fenton-Wilkinson approximation [21] for the sum of log-normally distributed variables. However, according to my tests on randomly sampled parameters μ and Σ, Fenton-Wilkinson approximation does not provide accurate enough distribution to compute an expectation of form 4.4. Therefore, we need to build a Monte-Carlo type of algorithm to calculate optimal weights.

4.2.1 Information theory interpretation of GOP optimization problem

In this subsection, we will provide an interpretation of GOP optimisation problem from the information theory perspective. It gives a good intuition behind the goal of the optimisation problem and since we haven't found this interpretation in the literature it might be a good starting point for further research of GOP.

Let's assume that components of $\mathbf{Y}_t$ are independent and have the following covariance structure

$$\Sigma = \begin{bmatrix} \sigma_{t,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{t,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & & \dots & \sigma_{t,N}^2 \end{bmatrix}.$$

Let $\mathbf{X}_{t,i} \sim N(0,1)$ i.e. $(\mathbf{X}_{t,1}, ..., \mathbf{X}_{t,N}) \sim (N((0,...,0), I_N)$, where I_N is $N \times N$ identity matrix. Then, we can rewrite the target function from Optimization Problem 4.4 as

$$\begin{aligned} &\mathbb{E}\left(\log\left(\sum_{i=1}^{N} w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right) \\ &= \mathbb{E}\left(\log\left(\sum_{i=1}^{N} \tilde{w}_{t,i}^V \exp\left(\sigma_{t,i}\mathbf{X}_{t,i} - \frac{1}{2}\sigma_{t,i}^2\right)\right) - \log\left(\sum_{k=1}^{N} \tilde{w}_{t,k}^V \exp\left(-\mu_{t,k} - \frac{1}{2}\sigma_{t,k}^2\right)\right)\right), \end{aligned} \tag{4.6}$$

where $\tilde{w}_{t,k}^V = \frac{w_{t,k}^V \exp\left(\mu_{t,k} + \frac{1}{2}\sigma_{t,k}^2\right)}{\sum_{i=1}^N w_{t,i}^V \exp\left(\mu_{t,i} + \frac{1}{2}\sigma_{t,i}^2\right)}$.

Remark 4.7. Note that if $\mathbf{Y} = \exp(\mu + \sigma\mathbf{X})$ where $X \sim N(0,1)$, i.e. $\mathbf{Y}$ is log-normally distributed, then $\mathbb{E}(\mathbf{Y}) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$. Thus, we have expectations of corresponding log-normally distributed r.v. in the second term.

Let's denote the density of Normal distribution with parameters μ, σ as $dN(\mu, \sigma)$ and rewrite (4.6) as

$$
\begin{aligned}
&= \mathbb{E}\left(\log\left(\sum_{i=1}^{N} \tilde{w}_{t,i}^{V} \frac{dN(\sigma_{t,i},1)(\mathbf{X}_{t,i})}{dN(0,1)(\mathbf{X}_{t,i})}\right)\right) - \log\left(\sum_{k=1}^{N} \tilde{w}_{t,k}^{V} \exp\left(-\mu_{t,k} - \frac{1}{2}\sigma_{t,k}^{2}\right)\right) \\
&= \mathbb{E}\left(\log\left(\sum_{i=1}^{N} \tilde{w}_{t,i}^{V} \frac{dN((0,0,...,\sigma_{t,i},0,...,0),I_N)(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N}))}{dN((0,0,...,0),I_N)(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N})))}\right)\right) \\
&- \log\left(\sum_{k=1}^{N} \tilde{w}_{t,k}^{V} \exp\left(-\mu_{t,k} - \frac{1}{2}\sigma_{t,k}^{2}\right)\right) \\
&= \mathbb{E}\left(\log\left(\frac{\sum_{i=1}^{N} \tilde{w}_{t,i}^{V} dN((0,0,...,\sigma_{t,i},0,...,0),I_N)(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N}))}{dN((0,0,...,0),I_N)(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N})))}\right)\right) \\
&- \log\left(\sum_{k=1}^{N} \tilde{w}_{t,k}^{V} \exp\left(-\mu_{t,k} - \frac{1}{2}\sigma_{t,k}^{2}\right)\right). \qquad (4.8)
\end{aligned}
$$

Denote

$$Q(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N}) = \sum_{i=1}^{N} \tilde{w}_{i}^{V} dN((0,0,...,\sigma_{t,i},0,...,0),I_N) \tag{4.9}$$

and

$$P(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N}) = dN((0,0,...,0,0,...,0),I_N). \tag{4.10}$$

Note that because of $\tilde{w}_{t}^{V} \in \mathbb{W}_{t}$, $Q(\mathbf{X}_{t,1},...,\mathbf{X}_{t,N})$ is a multivariate normal mixture density and the first term is Kullback-Leibler divergence between P and Q. Thus, using (4.8), Optimization Problem 4.4 can be restated as

$$\mathbf{g}_{t}^{*} = \sup_{\tilde{w}_{t}^{V} \in \mathbb{W}_{t}} \left(D_{KL}(P||Q) - \log\left(\sum_{k=1}^{N} \tilde{w}_{t,k}^{V} \exp\left(-\mu_{t,k} - \frac{1}{2}\sigma_{t,k}^{2}\right)\right)\right). \tag{4.11}$$

The first term can be interpreted as a distance between Multivariate Normal Distribution mixture and a Standard Normal Distribution. The second term is a regularization term that tights weights to the corresponding index of the largest component $\mu_{t,k} + \frac{1}{2}\sigma_{t,k}^{2}$.

4.3 Iterative algorithm for finding GOP weights

In this section, we will present the iterative algorithm for construction of GOP. The main idea of the algorithm is that there exists a fixed point equation for optimal weights. I want to thank Dario Gasbarra for proposing the idea and suggesting several improvements and extensions for the algorithm that we will present in the following sections. So let's start with a fixed point equation.

Theorem 4.3.1 (Fixed point equation)**.** *Let's denote logarithm of growth ratio* $\log(\mathbf{h}_t)$ *vector as* $\mathbf{Y}_t$*. If* $\mathbf{Y}_t \sim N(\mu, \Sigma)$ *and the components of* $\mathbf{Y}_t$ *are independent i.e. it has the following covariance structure*

$$\Sigma = \begin{bmatrix} \sigma_{t,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{t,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & & \dots & \sigma_{t,N}^2 \end{bmatrix},$$

then for all $t \in \mathbb{T}$ *and* $j \in \{1, ..., N\}$ *there exists a fixed point equation*

$$w_{t,j}^{V^*} = \sigma_{j,t}^{-1} \mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=1}^{N} w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i}) \right) \right). \tag{4.12}$$

Proof. Let $\mathbf{X}_{t,i} \sim N(0,1)$ then, Optimization Problem 4.4 can be rewritten as

$$\mathbf{g}_t^* = \sup_{\forall i \ w_{t,i}^V \geq 0 \text{ and } \sum_{i=1}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(\sum_{i=1}^{N} w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i}) \right) \right). \tag{4.13}$$

Now we have the optimization problem with N inequalities and one equality constraint. Let's denote corresponding Kuhn-Tucker multipliers as γ_i and λ, then Lagrangian for the problem is

$$\mathcal{L}(w_t^V, \gamma_1, ..., \gamma_N, \lambda) = \mathbb{E}\left(\log\left(\sum_{i=1}^{N} w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i}) \right) \right) - \sum_{k=1}^{N} \gamma_i w_{t,i}^V - \lambda \cdot (\sum_{i=1}^{N} w_{t,i}^V - 1).$$

As discussed in Section 3.2, we can pass the partial derivative inside the expectation in this problem. Kuhn-Tucker's conditions for optimality are then the following
For all $j \in \overline{1, N}$

$$\frac{\partial \mathcal{L}(w_t^{V^*}, \gamma_1, ..., \gamma_N, \lambda)}{\partial w_{t,j}^{V^*}}$$

$$= \frac{\partial \mathbb{E}\left(\log\left(\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right) - \sum_{k=1}^N \gamma_i w_{t,i}^{V^*} - \lambda \cdot (\sum_{i=1}^N w_{t,i}^{V} - 1)}{\partial w_{t,j}^{V^*}}$$

$$= \mathbb{E}\left(\frac{\exp(\mu_{t,j} + \sigma_{t,j}\mathbf{X}_{t,j})}{\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})}\right) - \gamma_j - \lambda = 0, \tag{4.14}$$

and

$$\gamma_j \cdot w_{t,j}^{V^*} = 0, \tag{4.15}$$

where $\gamma_j \geq 0$.
By the same procedure as in Section 3.2 we can derive that $\lambda = 1$. To see this, we can multiply (4.14) by corresponding weight and take the sum over weights

$$\sum_{j=1}^N w_{t,j}^{V^*} \mathbb{E}\left(\frac{\exp(\mu_{t,j} + \sigma_{t,j}\mathbf{X}_{t,j})}{\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})}\right) = \sum_{j=1}^N w_{t,j}^{V^*} (\gamma_j + \lambda), \text{ that is}$$

$$\mathbb{E}\left(\frac{\sum_{j=1}^N w_{t,j}^{V^*} \exp(\mu_{t,j} + \sigma_{t,j}\mathbf{X}_{t,j})}{\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})}\right) = \sum_{j=1}^N w_{t,j}^{V^*} \gamma_j + \sum_{j=1}^N w_{t,j}^{V^*} \lambda, \text{ so that}$$

$$\lambda = 1 - \sum_{j=1}^N w_{t,j}^{V^*} \gamma_j.$$

By Complementary Slackness Condition (4.15) we indeed get

$$\lambda = 1. \tag{4.16}$$

For the next step, we need Gaussian integration by parts (Stein's lemma) formula. The proof can be found in [12], Lemma 1.1.1.

Lemma 4.17 (Stein's lemma for d-dimensional normal vector)**.** *For d dimensional random vector* $\mathbf{Z} \sim N(0, I_d)$ *and for any function* f*, such that* $\mathbb{E}\left(\left|\left|\frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}}\right|\right|\right) < \infty$ *the following equation holds*

$$\mathbb{E}\left(\nabla f(\mathbf{Z})\right) = \mathbb{E}\left(\mathbf{Z} \cdot f(\mathbf{Z})\right). \tag{4.18}$$

Let's apply this lemma to the target function inside the expectation in Problem 4.13

$$\mathbb{E}\left(\frac{\partial \log\left(\sum_{i=0}^N w_{t,i}^{V} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)}{\partial \mathbf{X}_{t,j}}\right) = \mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=1}^N w_{t,i}^{V} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right),$$

so,

$$\mathbb{E}\left(\frac{w_{t,j}^V \sigma_{t,j} \exp(\mu_{t,j} + \sigma_{t,j}\mathbf{X}_{t,j})}{\sum_{i=1}^N w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})}\right) = \mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=0}^N w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right).$$

Therefore we get

(4.19)

$$w_{t,j}^V =$$
$$\sigma_{j,t}^{-1}\left(\mathbb{E}\left(\frac{\exp(\mu_{t,j} + \sigma_{t,j}\mathbf{X}_{t,j})}{\sum_{i=0}^N w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})}\right)\right)^{-1} \mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=1}^N w_{t,i}^V \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right).$$

For the optimal weights by Condition (4.14) and (4.16)

$$w_{t,j}^{V^*} = \sigma_{j,t}^{-1}\left(\gamma_j + 1\right)^{-1} \mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right). \tag{4.20}$$

Let's denote the set of indices with positive optimal weights as $\mathbb{I}_{>0} = \{i \in \overline{1,N} | w_{t,i}^{V^*} > 0\}$ and with zero weights as $\mathbb{I}_{=0} = \{i \in \overline{1,N} | w_{t,i}^{V^*} = 0\}$.

Now by Complementary Slackness Condition (4.15) for non-zero weights we have

$$\text{For all } j \in \mathbb{I}_{>0} : w_{t,j}^{V^*} = \sigma_{j,t}^{-1}\mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right) \tag{4.21}$$

While for the zero weights since $\gamma_j \geq 0$ and $\{\mathbf{X}_{t,i}\}_{i=1}^N$ are independent

$$\text{For all } j \in \mathbb{I}_{=0} : 0 = w_{t,j}^{V^*} = \mathbb{E}\left(\mathbf{X}_{t,j} \log\left(\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)\right) = 0 \tag{4.22}$$

□

Remark 4.23. Note that random variables $\mathbf{X}_{t,j}$ and $\log\left(\sum_{i=1}^N w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right)$ are non-negatively correlated since logarithm is an increasing function. Thus, all updates for the weights should be non-negative: $\forall t, j : w_{t,j}^{V^*} \geq 0$. However, Monte-Carlo estimations can be negative, so we should truncate the updated weights at zero.

Update rule Fixed Point Equation (4.12) holds only for optimal weights. However, we can make a guess that starting with some initial weights w_t^0 we will converge to the optimal solution by iterating weights with this equation. At step $n \in \mathbb{N}$, we update the weight $j \in \{1, ..., N\}$ as

$$w_{t,j}^{n+1} = \sigma_{j,t}^{-1} \mathbb{E} \left(\mathbf{X}_{t,j} \log \left(\sum_{i=1}^{N} w_{t,i}^{n} \exp(\mu_{t,i} + \sigma_{t,i} \mathbf{X}_{t,i}) \right) \right). \tag{4.24}$$

Given this update rule, we can estimate expectations with Monte Carlo sampling and propose the following algorithm for computing GOP weights. Starting from initial weights, we use Update rule (4.24) and then re-normalize weights such that they sum up to one. We iterate until the difference between w_t^{n+1} and w_t^n becomes negligible. The pseudo code for the algorithm is the following.

Algorithm 1 GOP weights for log-normally distributed growth ratio

function GOPWEIGHTS($weights_0, N, \mu, \sigma$)
  $\varepsilon = 10^{-6}$ ▷ Set precision.
  $weights_{cur} = weights_0$
  $weights_{new} = weights_0 + \varepsilon$
  while $||weights_{cur} - weights_{new}||_{L1} > \varepsilon$ **do**
    $weights_{cur} = weights_{new}$
    $samples$ = generateSamples(N) ▷ Generate samples from N dimensional standard normal distribution.
    for $j \in \overline{1, N}$ **do**
      $weights_{new}[j] = \sigma^{-1}[j] \cdot$ computeExpectation($samples, weights_{cur}, \mu, \sigma$)
    $weights_{new} = \frac{weights_{new}}{||weights_{new}||_{L1}}$
  return $weights_{new}$
function COMPUTEEXPECTATION($\mathbf{X}, weights, \mu, \sigma$)
  $L = length(\mathbf{X}_j)$
  return $\max(\frac{1}{L} \sum_{k=1}^{N} \mathbf{X}_j[k] \cdot \log\left(\sum_{i=1}^{N} weights[i] \exp(\mu[i] + \sigma[i] \mathbf{X}_i[k])\right), 0)$

Remark 4.25. Note that Equations (4.16) or (4.22) can be also used to check optimality of the computed solution.

This algorithm is quite simple and can be a good alternative to the usual general gradient descent methods. Firstly, it does not require parameters tuning. For instance, we do not need to provide it with a learning rate. Secondly, it might provide a significant speed improvement when we solve the problem with a big number of assets, especially if the gradient descent method uses also second derivatives. Finally, we can use fixed Monte Carlo samples for it or can generate new ones each step. Which means that compared to the gradient descent methods, Algorithm 1 will not overfit to the provided samples.

4.4 Variance reduction techniques

Update rule (4.24) involves expectation operators that are too complicated to be computed analytically, thus we compute them by Monte Carlo simulation. In order to converge to the optimal weights, we need accurate estimations, for these reasons we use different variance reduction tricks.

4.4.1 Antithetic Variates

The first technique that we use to decrease variance is *Antithetic variates* trick described in Kroese book [19], section 9.2. For the Standard Normal Distribution, it works as follows.

Suppose that we want to compute the statistics $\mathbb{E}(h(\mathbf{X}))$ for normally distributed random variable $\mathbf{X} \sim N(0,1)$. Then, the *antithetic variates* trick is to compute expectation of $g(\mathbf{X})) = \frac{1}{2}(h(\mathbf{X})) + h(-\mathbf{X}))$ instead of $h(\mathbf{X})$. Random variable $g(\mathbf{X})$ has the same expectation as $h(\mathbf{X})$ but reduced variance if $h(\mathbf{X})$ is monotone function, since correlation between $h(\mathbf{X})$ and $h(-\mathbf{X})$ is negative.

4.4.2 Static Control Variate

In addition to general *antithetic variates* trick, we can use another variance reduction trick called *Static Control Variate* that is described in Pagès (2018) [22], Section 3.1.

Suppose that we want to compute $\mathbb{E}(\mathbf{X})$, the idea of this trick is to find another random variable $\mathbf{Y}$ such that $\mathbb{E}(\mathbf{Y})$ can be computed analytically and the variance of random variable $\mathbf{Z} = \mathbf{X} - \mathbf{Y} \geq 0$ became smaller. Let's find a control variate for the expectation in Updating Rule 4.24

$$
\begin{aligned}
&\mathbb{E}\left(\mathbf{X}_{t,j}\log\left(\sum_{i=1}^{N} w_{t,i}^{V^*}\exp(\mu_{t,i}+\sigma_{t,i}\mathbf{X}_{t,i})\right)\right) = \mathbb{E}\left(\mathbf{X}_{t,j}\log\left(\sum_{i=1}^{N} w_{t,i}^{V^*}\exp(\mu_{t,i}+\sigma_{t,i}\mathbf{X}_{t,i})\right)\right)\\
&\quad + \mathbb{E}\left(\mathbf{X}_{t,j}\sum_{i=1}^{N} w_{t,i}^{V^*}(\mu_{t,i}+\sigma_{t,i}\mathbf{X}_{t,i})\right) - \mathbb{E}\left(\mathbf{X}_{t,j}\sum_{i=1}^{N} w_{t,i}^{V^*}(\mu_{t,i}+\sigma_{t,i}\mathbf{X}_{t,i})\right)\\
&=\mathbb{E}\left(\mathbf{X}_{t,j}\left\{\log\left(\sum_{i=1}^{N} w_{t,i}^{V^*}\exp(\mu_{t,i}+\sigma_{t,i}\mathbf{X}_{t,i})\right) - \sum_{i=1}^{N} w_{t,i}^{V^*}(\mu_{t,i}+\sigma_{t,i}\mathbf{X}_{t,i})\right\}\right) + w_{t,j}^{V^*}\cdot\sigma_{t,j}.
\end{aligned}
$$

By Jensen's inequality for all $\omega \in \Omega$ we have

$$\log\left(\sum_{i=1}^{N} w_{t,i}^{V^*} \exp(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i})\right) \geq \sum_{i=1}^{N} w_{t,i}^{V^*}(\mu_{t,i} + \sigma_{t,i}\mathbf{X}_{t,i}).$$

Combining these two reduction methods we can rewrite function *computeExpectation* In Algorithm 1.

---

function COMPUTEEXPECTATION($\mathbf{X}, w, \mu, \sigma$)

$L = length(\mathbf{X}_j)$

$$g_X = \frac{1}{L}\sum_{k=1}^{N} \mathbf{X}_j[k] \cdot \left\{ \log\left(\sum_{i=1}^{N} w[i]\exp(\mu[i] + \sigma[i]\mathbf{X}_i[k])\right) - \log\left(\sum_{i=1}^{N} w[i]\exp(\mu[i] - \sigma[i]\mathbf{X}_i[k])\right) - \sum_{i=1}^{N}(w[i]\mu[i] - \sigma[i]\mathbf{X}_i[k]) \right\}$$

return $g_X + w[j] \cdot \sigma[j]$

In order to test these two variance reduction methods, we experimented with sampled parameters from standard Normal distribution: $\mu \sim N(0, I_N)$, $\sigma \sim |N(0, I_N)|$ and normalized $w \propto |N(0, I_N)|$. We have computed the Monte Carlo variance of expectation from Update Rule 4.21 on samples of size 10^4 for 100 times and on average the variance was reduced by factor 11.

4.5 Convergence of the algorithm

In this section, we will provide theoretical and practical arguments for the convergence of Algorithm 1. From the theoretical point of view, we will consider the convergence under the second order approximation of Logarithm function. In practice, we will compare the algorithm to the standard *scipy* library's optimizer that uses Sequential quadratic programming implemented by Dieter Kraft [18].

4.5.1 Theoretical argument

We will consider the algorithm at a certain time t, so I will omit the time index for simplicity. Let w^* - target actual weights of GOP, $w = w^* + \Delta w$ - current weights, where Δw is a current error. Our goal is to show that after one iteration the updated weights become closer to the actual weights w^*.

Lemma 4.26. *If Δw is small enough such that we can use second order approximation of the logarithm*

$$\log(1+x) = x - \frac{x^2}{2} + O(x^3), \text{ as } x \to \infty. \tag{4.27}$$

and if, in addition,

1. *we can neglect a second order difference term i.e.* $(\Delta w)^2 \ll \Delta w$,
2. *for all i we have* $\exp(\mu_i + \sigma_i^2/2) < 2$,
3. *underlying assets are not dramatically different from each other* $\sum_{i=1}^N \Delta w_i \exp(\mu_i + \sigma_i^2/2) \approx 0$ [1] ,

then the following condition should be satisfied so that a new weight $w_j^u, j \in \{1, ..., N\}$ after update (4.24) is closer to the optimal w_j^* than before. Namely, if

(4.28)

$$\left|\exp\big(\mu_j + \sigma_j^2/2\big) \cdot \left(2 - \max_{i \in \overline{1,N}} \left(\exp\big(\mu_i + \sigma_i^2/2\big)\right)\right) - 2\left(\exp\big(2\mu_j + 2\sigma_j^2\big) - \exp\big(2\mu_j + \sigma_j^2\big)\right)\right| < 1$$

then $|w_j^u - w_j^*| < |w_j - w_j^*|$.

Proof. Let's consider non-negative weights from the set $\mathbb{I}_{>0}$, by (4.21) we have

$$\begin{aligned}
w_j^* &= \sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j \log\left(\sum_{i=1}^N w_i^* \exp(\mu_i + \sigma_i \mathbf{X}_i)\right)\right) \\
&= \mathbb{E}\left(\mathbf{X}_j \log\left(1 + \sum_{i=1}^N w_i^*(\exp(\mu_i + \sigma_i \mathbf{X}_i) - 1)\right)\right) \\
&\approx \sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j \left(\left(\sum_{i=1}^N w_i^*(\exp(\mu_i + \sigma_i \mathbf{X}_i) - 1)\right) - \frac{\left(\sum_{i=1}^N w_i^*(\exp(\mu_i + \sigma_i \mathbf{X}_i) - 1)\right)^2}{2}\right)\right).
\end{aligned}$$

[1] Note that $\sum_{i=1}^N \Delta w_i = 0$.

The updated weight w_j^u after one iteration of (4.21) is then approximately equals

$$\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\left(\sum_{i=1}^{N}(w_i^*+\Delta w_i)(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)\right.\right.$$
$$\left.\left.-\frac{\left(\sum_{i=1}^{N}(w_i^*+\Delta w_i)(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)^2}{2}\right)\right).$$

We can take the difference of these weights and use linearity of expectation

(4.29)

$$w_j^u-w_j^*\approx$$
$$\approx\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\sum_{i=1}^{N}(w_i^*+\Delta w_i)(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right.\right.$$
$$\left.\left.-\frac{\left(\sum_{i=1}^{N}(w_i^*+\Delta w_i)(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)^2}{2}\right)\right)$$
$$-\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\sum_{i=1}^{N}w_i^*(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)-\frac{\left(\sum_{i=1}^{N}w_i^*(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)^2}{2}\right)\right)$$

(4.30)

The first terms are linear and have common multiplier and for the second terms we can apply simple difference of squares formula, so (4.30) is then

(4.31)

$$=\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\sum_{i=1}^{N}\Delta w_i(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right.\right.$$

$$\left.\left.-\frac{\left(\sum_{i=1}^{N}\Delta w_i(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)\left(\sum_{i=1}^{N}(\Delta w_i+2w_i^*)(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)}{2}\right)\right).$$

Now we can use Assumption 1 meaning that we are quite close to the optimal portfolio, so we can ignore second order term $\frac{1}{2}\left(\sum_{i=1}^{N}\Delta w_i(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)^2$, then (4.31) can be rewritten as

$$
\begin{aligned}
&\approx \sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\sum_{i=1}^{N}\Delta w_i\exp(\mu_i+\sigma_i\mathbf{X}_i)\right.\right.\\
&\left.\left.-\left(\sum_{i=1}^{N}\Delta w_i\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)\left(\sum_{i=1}^{N}w_i^*(\exp(\mu_i+\sigma_i\mathbf{X}_i)-1)\right)\right)\right)\\
&=\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\left(\sum_{i=1}^{N}\Delta w_i\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)\left(2-\sum_{i=1}^{N}w_i^*\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)\right)\right). \qquad (4.32)
\end{aligned}
$$

Since we assume that $\mathbf{X}_i$ and $\mathbf{X}_j$ are independent for $i\neq j$ most of the elements inside the expectation are zero because for $i\neq j$: $\mathbb{E}\left(\mathbf{X}_j\cdot g(\mathbf{X}_i)\right)=\mathbb{E}(\mathbf{X}_j)\cdot\mathbb{E}(g(\mathbf{X}_i))=0$. Thus, we can rearrange (4.32) and again use linearity of expectation

$$
\begin{aligned}
&=2\sigma_j^{-1}\Delta w_j\mathbb{E}\left(\mathbf{X}_j\exp(\mu_j+\sigma_j\mathbf{X}_j)\right)\\
&-\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\Delta w_j\exp(\mu_j+\sigma_j\mathbf{X}_j)\left(\sum_{i=1,i\neq j}^{N}w_i^*\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)\right)\\
&-\sigma_j^{-1}\mathbb{E}\left(\mathbf{X}_j\left(\left(\sum_{i=1,i\neq j}^{N}\Delta w_i\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)w_j^*\exp(\mu_j+\sigma_j\mathbf{X}_j)\right)\right)\\
&-\sigma_j^{-1}\Delta w_j w_j^*\mathbb{E}\left(\mathbf{X}_j\exp(\mu_j+\sigma_j\mathbf{X}_j)^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&=2\sigma_j^{-1}\Delta w_j\mathbb{E}\left(\mathbf{X}_j\exp(\mu_j+\sigma_j\mathbf{X}_j)\right)\\
&-\sigma_j^{-1}\Delta w_j^*\mathbb{E}\left(\mathbf{X}_j\exp(\mu_j+\sigma_j\mathbf{X}_j)\right)\sum_{i=1,i\neq j}^{N}w_i^*\mathbb{E}\left(\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)\\
&-\sigma_j^{-1}w_j^*\mathbb{E}\left(\mathbf{X}_j\exp(\mu_j+\sigma_j\mathbf{X}_j)\right)\sum_{i=1,i\neq j}^{N}\Delta w_i\mathbb{E}\left(\exp(\mu_i+\sigma_i\mathbf{X}_i)\right)\\
&-\sigma_j^{-1}\Delta w_j^* w_j^*\mathbb{E}\left(\mathbf{X}_j\exp(\mu_j+\sigma_j\mathbf{X}_j)^2\right) \qquad (4.33)
\end{aligned}
$$

By the same derivation of expectation as in Section 3.5 we can compute

$$\mathbb{E}\left(\exp(\mu_j+\sigma_j\mathbf{X}_j)\right)=\exp\left(\mu_j+\sigma_j^2/2\right), \qquad (4.34)$$

similarly

$$\mathbb{E}\left(\mathbf{X}_j \exp(\mu_j + \sigma_j \mathbf{X}_j)\right) = \sigma_j \exp\bigl(\mu_j + \sigma_j^2/2\bigr), \tag{4.35}$$

and

$$\mathbb{E}\left(\mathbf{X}_j \exp(\mu_j + \sigma_j \mathbf{X}_j)^2\right) = 2\sigma_j \exp\bigl(2\mu_j + 2\sigma_j^2\bigr). \tag{4.36}$$

Now plug them in (4.33)

$$\begin{aligned}
&= 2\Delta w_j \exp\bigl(\mu_j + \sigma_j^2/2\bigr) - \Delta w_j \exp\bigl(\mu_j + \sigma_j^2/2\bigr) \sum_{i=1, i\neq j}^{N} w_i^* \exp\bigl(\mu_i + \sigma_i^2/2\bigr) \\
&- w_j^* \exp\bigl(\mu_j + \sigma_j^2/2\bigr) \sum_{i=1, i\neq j}^{N} \Delta w_i \exp\bigl(\mu_i + \sigma_i^2/2\bigr) - 2\Delta w_j w_j^* \exp\bigl(2(\mu_j + \sigma_j^2)\bigr) \\
&= 2\Delta w_j \exp\bigl(\mu_j + \sigma_j^2/2\bigr) + 2\Delta w_j w_j^* \exp\bigl(2\mu_j + \sigma_j^2\bigr) \\
&- \Delta w_j \exp\bigl(\mu_j + \sigma_j^2/2\bigr) \sum_{i=1}^{N} w_i^* \exp\bigl(\mu_i + \sigma_i^2/2\bigr) \\
&- w_j^* \exp\bigl(\mu_j + \sigma_j^2/2\bigr) \sum_{i=1}^{N} \Delta w_i \exp\bigl(\mu_i + \sigma_i^2/2\bigr) - 2\Delta w_j w_j^* \exp\bigl(2(\mu_j + \sigma_j^2)\bigr).
\end{aligned}$$

We assume that we are close to GOP and we don't have dramatically different assets $\sum_{i=1}^{N} \Delta w_i \exp(\mu_i + \sigma_i^2/2) \approx 0$ i.e. Assumption 2. Moreover, by Assumption 3 for all i we have $\exp(\mu_i + \sigma_i^2/2) < 2$. Given these assumptions we get

$$\begin{aligned}
&|w_j^u - w_j^*| \lessapprox \\
&\lessapprox |\Delta w_j| \, \Bigl|2\exp\bigl(\mu_j + \sigma_j^2/2\bigr) + 2w_j^* \left(\exp\bigl(2\mu_j + \sigma_j^2\bigr) - \exp\bigl(2(\mu_j + \sigma_j^2)\bigr)\right) \\
&- \exp\bigl(\mu_j + \sigma_j^2/2\bigr) \sum_{i=1}^{N} w_i^* \exp\bigl(\mu_i + \sigma_i^2/2\bigr)\Bigr| \\
&\leq |\Delta w_j| \left|\exp\bigl(\mu_j + \sigma_j^2/2\bigr) \cdot \left(2 - \max_{i \in \overline{1,N}} \left(\exp\bigl(\mu_i + \sigma_i^2/2\bigr)\right)\right)\right. \\
&\left. -2\left(\exp\bigl(2\mu_j + 2\sigma_j^2\bigr) - \exp\bigl(2\mu_j + \sigma_j^2\bigr)\right)\right|
\end{aligned}$$

Finally, we have the rough criterion that allow us to check if one step of the algorithm drives the non-negative weight towards the optimal one. Namely, if

$$\left|\exp\left(\mu_j + \sigma_j^2/2\right) \cdot \left(2 - \max_{i \in \overline{1,N}} \left(\exp\left(\mu_i + \sigma_i^2/2\right)\right)\right) - 2\left(\exp\left(2\mu_j + 2\sigma_j^2\right) - \exp\left(2\mu_j + \sigma_j^2\right)\right)\right| < 1$$

then $|w_j^u - w_j^*| \lessapprox |\Delta w_j|$ and we are closer to the optimal solution than before. If Property (4.28) is satisfied for all non-negative weights, then we make a step to the optimal direction and re-normalization will not make an allocation worse. In practice, the values of μ and σ vectors are quite small, so Property (4.28) is usually satisfied.

□

4.5.2 Simulations

The calculations in previous Subsection 4.5.1 show us that a convergence criterion for Algorithm 1 is hard to derive even if we allow a number of assumptions and approximations. Therefore, let's see how Algorithm 1 works in practice.

We approximate expectations from the algorithm by Monte-Carlo method, drawing $n = 10^5$ from the N-dimensional Normal Distribution. In order to have more accurate results, we used variance reduction tricks described in Section 4.4.

Let's firstly set $N = 10$ and use $\mu \sim N(0, I_N)$ and $\sigma \sim |N(0, I_N)|$. As a benchmark for the results, we use standard optimization method from *scipy* python library.

#	Initial	Scipy Optimizer	Iterative Alg.	Difference %	Criterion (4.28)
1	1.22	2.27	2.25	0.80	True
2	1.04	2.49	2.49	0.10	True
3	1.15	2.07	2.05	1.30	True
4	1.05	1.72	1.63	4.80	True
5	1.99	2.76	2.75	0.30	True
6	0.94	1.60	1.58	1.20	True
7	0.69	1.02	1.01	0.80	True
8	0.34	1.27	1.25	1.70	True
9	0.26	0.75	0.74	0.70	True
10	1.46	2.28	2.25	1.00	True
Mean	1.02	1.82	1.80	1.27	-

Table 4.1: Optimization results for randomly sampled parameters

We see that in all cases Algorithm 1 performs almost as good as scipy optimizer. The maximum difference between optimizers was 4.8% and the average difference is 1.27%.

Note that we compute the target function by Monte-Carlo and, therefore, it also contains some error. Overall we can say that this simulation test is passed and Algorithm 1 converges to the optimum.

Now let's take 10 assets from the real market and construct 10 portfolios where each time $N = 5$ randomly chosen assets are selected. For estimate μ and σ of each asset, we use mean values calculated from monthly close prices from March 2009 to 1 March 2019, because now the purpose is to understand whether we can apply Algorithm 1 for real assets at all, rather than to construct a proper GOP portfolio. We took the returns data from *Investing.com* and then transformed it to the log growth ratios. The assets and their parameters are

Ticker	μ	σ
AAPL	0.021	0.073
AMD	0.018	0.170
BAC	0.011	0.098
BA	0.019	0.071
INTC	0.011	0.061
FB	0.021	0.100
FDX	0.011	0.076
MS	0.005	0.089
WMT	0.005	0.048
XOM	0.001	0.048

Table 4.2: Stocks' monthly estimated parameters 09-19

#	Initial	Scipy Optimizer	Iterative Alg.	Difference %	Criterion (4.28)
1	0.015	0.023	0.020	13.1	True
2	0.016	0.022	0.021	4.4	True
3	0.018	0.024	0.024	0.1	True
4	0.016	0.024	0.024	0.0	True
5	0.011	0.019	0.018	7.4	True
6	0.015	0.023	0.021	8.4	True
7	0.014	0.023	0.022	6.5	True
8	0.013	0.022	0.022	1.0	True
9	0.016	0.024	0.023	4.1	True
10	0.016	0.024	0.023	3.3	True
Mean	0.015	0.023	0.022	4.8	-

Table 4.3: Optimization results with randomly picked 5 assets from the real market

We see that on average the results of algorithm are close to the optimal ones. However, there are cases like 1 or 6 when it looks like our method did not converge to global optima while Criterion (4.28) is satisfied. However, in all cases, we got an improvement to the initial value, so we conclude that we can use this algorithm as a fast approximation for GOP that does not require computing gradients.

4.6 Extensions of the algorithm

In this section, we will provide two important extensions of Algorithm 1. Namely, we will consider the case of correlated securities and the case where we have a riskless asset. The case of correlated risky assets combined with a risk-free instrument involves complicated Inverse problem 4.48. We provide a possible structure of the solution, however, it is not guaranteed that in general setting this form of solution is positive definite. Thus, further research is needed to identify the structure that works for any form of the covariance matrix.

4.6.1 Extension 1: Correlated log ratios

Let's use notation from Section 4.3

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{X}_t + \mu_t, \\ \mathbf{X}_t &\sim N(0, \Sigma), \end{aligned} \tag{4.37}$$

where Σ is positive definite covariance matrix.

Given this notation, Optimization Problem 4.4 can be rewritten as

$$\mathbf{g}_t^* = \sup_{\forall i\ w_{t,i}^V \geq 0 \text{ and } \sum_{i=1}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(\sum_{i=1}^N w_{t,i}^V \exp(\mu_{t,i} + \mathbf{X}_{t,i})\right)\right). \tag{4.38}$$

To obtain stationary point equation 4.21 we will again use Stein's lemma. Since a covariance matrix of a normally distributed random vector is symmetric and positive definite we can use LU decomposition, that in a symmetric case gives

$$\Sigma = AA^T. \tag{4.39}$$

On the other hand if $\mathbf{Z} \sim N(0, I_d)$ and $\mathbf{X} = A\mathbf{Z}$, then the covariance matrix of $\mathbf{X}$ is

$$Cov(\mathbf{X}) = \mathbb{E}((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^T) = \mathbb{E}((A\mathbf{Z} - \mathbb{E}(A\mathbf{Z}))(A\mathbf{Z} - \mathbb{E}(A\mathbf{Z}))^T)$$

$$= A\mathbb{E}((\mathbf{Z} - \mathbb{E}(\mathbf{Z}))(\mathbf{Z} - \mathbb{E}(\mathbf{Z}))^T)A^T = ACov(\mathbf{Z})A^T = AI_dA^T = AA^T.$$

Therefore, since Multivariate Normal Distribution is characterized by its mean and covariance matrix, $\mathbf{X}$ is just a linear transformation of some standard normally distributed vector $\mathbf{X} = A\mathbf{Z}$.

We can rewrite Lemma 4.17 for $\mathbf{X}_t = A\mathbf{Z}_t$ using transformation of gradient rule $\nabla_x f(Ax) = A^T \nabla f(Ax)$ [20]. Let $g(\mathbf{Z}_t) = f(A\mathbf{Z}_t)$, then

$$\mathbb{E}\left(\nabla_Z g(\mathbf{Z}_t)\right) = \mathbb{E}\left(\mathbf{Z}_t \cdot g(\mathbf{Z}_t)\right),$$

by definition of $g(\mathbf{Z}_t)$ and gradient rule

$$\mathbb{E}\left(A^T \nabla f(A\mathbf{Z}_t)\right) = \mathbb{E}\left(\mathbf{Z}_t \cdot f(A\mathbf{Z}_t)\right),$$

returning to the $\mathbf{X}_t$ we get

$$\mathbb{E}\left(A^T \nabla f(\mathbf{X}_t)\right) = \mathbb{E}\left(A^{-1}\mathbf{X}_t \cdot f(\mathbf{X}_t)\right).$$

Therefore the expectation of gradient can be written as

$$\begin{aligned}\mathbb{E}\left(\nabla f(\mathbf{X}_t)\right) &= (A^T)^{-1}A^{-1}\mathbb{E}\left(\mathbf{X}_t \cdot f(\mathbf{X}_t)\right)\\ &= (AA^T)^{-1}\mathbb{E}\left(\mathbf{X}_t \cdot f(\mathbf{X}_t)\right) = \Sigma^{-1}\mathbb{E}\left(\mathbf{X}_t \cdot f(\mathbf{X}_t)\right).\end{aligned} \tag{4.40}$$

Note that $\mathbb{E}\left(\nabla f(\mathbf{X}_t)\right) = \left(\mathbb{E}\left(\frac{\partial f(\mathbf{X}_{t,1})}{\partial \mathbf{X}_{t,1}}\right), \mathbb{E}\left(\frac{\partial f(\mathbf{X}_{t,2})}{\partial \mathbf{X}_{t,2}}\right), ..., \mathbb{E}\left(\frac{\partial f(\mathbf{X}_{t,N})}{\partial \mathbf{X}_{t,N}}\right)\right)^T$, and these components we have already computed in Section 4.3. For non-negative weights by (4.16) and (4.15) we have

$$\mathbb{E}\left(\frac{\partial \log\left(\sum_{i=0}^N w_{t,i}^{V*} \exp(\mu_{t,i} + \mathbf{X}_{t,i})\right)}{\partial \mathbf{X}_{t,j}}\right) = w_{t,j}^{V*}\mathbb{E}\left(\frac{\exp(\mu_{t,j} + \sigma_{t,j}\mathbf{X}_{t,j})}{\sum_{i=1}^N w_{t,i}^{V*} \exp(\mu_{t,i} + \mathbf{X}_{t,i})}\right) = w_{t,i}^{V*}. \tag{4.41}$$

Therefore, by (4.40) and (4.41) we can again propose an update rule for weights $w_{t,i}^K$ in vector notation

$$w_t^{K+1} = \Sigma^{-1}\mathbb{E}\left(\mathbf{X}_t \log\left(\sum_{i=0}^N w_{t,i}^K \exp(\mu_{t,i} + \mathbf{X}_{t,i})\right)\right). \tag{4.42}$$

And the updated version of Algorithm 1 is

Algorithm 2 GOP weights for correlated log-normally distributed growth ratios

function GOPWEIGHTS($weights_0, N, \mu, \Sigma$)
  $\varepsilon = 10^{-6}$ ▷ Set precision.
  $weights_{cur} = weights_0$
  $weights_{new} = weights_0 + \varepsilon$
  while $||weights_{cur} - weights_{new}||_{L1} > \varepsilon$ **do**
    $weights_{cur} = weights_{new}$
    $samples = \text{generateSamples}(N, \Sigma)$ ▷ Generate samples from N dimensional normal distribution with covariance matrix Σ.
    for $j \in \overline{1, N}$ **do**
      $expVector[j] = \text{computeExpectation}(samples[:, j], weights_{cur}, \mu, \Sigma)$
    $weights_{new} = \Sigma^{-1} \cdot expVector$
    $weights_{new} = \frac{weights_{new}}{||weights_{new}||_{L1}}$
  return $weights_{new}$

4.6.2 Extension 2: Risk-free instrument

Now we can get back to the initial setting where we have a risk-free on the market. With additional risk-free asset Optimization problem 3.1 can be written as

$$\mathbf{g}_t^* = \sup_{\forall i \ w_{t,i}^V \geq 0 \text{ and } \sum_{i=0}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(w_{t,0}^V h_{t,0} + \sum_{i=1}^N w_{t,i}^V \mathbf{h}_{t,i}\right)\right), \tag{4.43}$$

where $h_{t,0}$ is deterministic. Therefore, (4.43) is equivalent to

$$\begin{aligned} \mathbf{g}_t^* &= \sup_{\forall i \ w_{t,i}^V \geq 0 \text{ and } \sum_{i=0}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(w_{t,0}^V + \sum_{i=1}^N w_{t,i}^V \frac{\mathbf{h}_{t,i}}{h_{t,0}}\right)\right) \\ &= \sup_{\forall i \ w_{t,i}^V \geq 0 \text{ and } \sum_{i=0}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(w_{t,0}^V + \sum_{i=1}^N w_{t,i}^V \hat{\mathbf{h}}_{t,i}\right)\right). \end{aligned} \tag{4.44}$$

Let's forget for a moment about this optimization problem and consider the case where we have $N+1$ risky securities. We can use the first risky security as numeraire and rewrite target function inside the expectation as

$$\log\left(\sum_{i=0}^N w_{t,i}^V \mathbf{h}_{t,i}\right) = \log(\mathbf{h}_{t,0}) + \log\left(w_{t,0}^V + \sum_{i=1}^N w_{t,i}^V \tilde{\mathbf{h}}_{t,i}\right). \tag{4.45}$$

And again, as in (4.44)

$$\mathbf{g}_t^* = \sup_{\forall i \ w_{t,i}^V \geq 0 \text{ and } \sum_{i=0}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(\sum_{i=0}^N w_{t,i}^V \mathbf{h}_{t,i}\right)\right) \tag{4.46}$$

$$= \sup_{\forall i \ w_{t,i}^V \geq 0 \text{ and } \sum_{i=0}^N w_{t,i}^V = 1} \mathbb{E}\left(\log\left(w_{t,0}^V + \sum_{i=1}^N w_{t,i}^V \tilde{\mathbf{h}}_{t,i}\right)\right).$$

Let's have Assumptions 4.37 regarding $\mathbf{X}_t$ and $\mathbf{Y}_t$ as in Section 4.6.1, therefore

$$\tilde{\mathbf{h}}_{t,i} = \frac{\mathbf{h}_{t,i}}{\mathbf{h}_{t,0}} = \exp\left(\tilde{\mu}_{t,i} + \tilde{\mathbf{X}}_{t,i}\right), \tag{4.47}$$

where $\tilde{\mu}_{t,i} = \mu_{t,i} - \mu_{t,0}$ and

$$\mathbb{E}(\tilde{X}_{t,i}\tilde{X}_{t,j}) = \tilde{\Sigma}_{i,j} = \Sigma_{i,j} + \Sigma_{0,0} - \Sigma_{0,j} - \Sigma_{0,i}. \tag{4.48}$$

Thus, new dummy variables $\tilde{\mathbf{h}}_{t,i}$ are again log-normally distributed and this optimization problem corresponds to (4.44). Combining (4.44) and (4.46) we see that instead of solving Optimization problem (4.43) we can find corresponding artificial $N+1$ risky assets and solve the problem by Algorithm 1. For simplicity, we can fix $\mu_{t,0} = 0$ in (4.46) and consider already discounted $\mathbf{h}_{t,0}$ as in (4.44). Then, the task is to solve the inverse problem, namely to find a positive definite covariance matrix $\hat{\Sigma}$ for $N+1$ optimization problem that satisfies equation (4.48) given initial $N \times N$ covariance matrix Σ for risky assets.

For the case of the initial diagonal covariance matrix Σ as in Section 4.3 it is not hard to guess a correct solution.

For $\hat{\sigma_{t,0}} \in \mathbb{R}^+$ satisfying

$$\min_{i \in \overline{1,N}} (\sigma_{t,i}^2) > \hat{\sigma}_{t,0} \cdot N \tag{4.49}$$

the matrix

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma_{t,0}}^2 & 0 & \dots & \dots & 0 \\ 0 & \sigma_{t,1}^2 - \hat{\sigma_{t,0}}^2 & -\hat{\sigma_{t,0}}^2 & \dots & -\hat{\sigma_{t,0}}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\hat{\sigma_{t,0}}^2 & \dots & -\hat{\sigma_{t,0}}^2 & \sigma_{t,N}^2 - \hat{\sigma_{t,0}}^2 \end{bmatrix}$$

is positive definite and satisfies Equation (4.48). In the general setting, we can try a matrix with the same structure as $\hat{\Sigma}$. However, the criterion for positive determinant becomes more complicated than (4.49).

5. Growth Optimal Portfolio for ARMA and VAR growth ratio models

In this chapter, we will show how to construct Growth Optimal Portfolio by Algorithms 1, 2 on the real market using time series models. We will consider two different approaches, namely modelling assets' growth ratios independently with ARMA models and joint modelling by VAR model. We will provide only specifications of these time series models, for detailed analysis of these models [24]. For estimations of time-series models we use *statsmodels* package for python [23].

5.1 ARMA model

5.1.1 Specification

Standard Auto Regressive Moving Average with lags p and q i.e. ARMA(p, q) model for a process $\mathbf{X}_t$ states that

$$\mathbf{X}_t = c + \sum_{i=1}^{p} \alpha_i \mathbf{X}_{t-i} + \sum_{j=1}^{q} \beta_j \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{5.1}$$

where $\boldsymbol{\varepsilon}_t$ are i.i.d. normal random variables or strong white noise process and c, α_i, β_i - scalar model parameters.

Let's assume that logarithm of growth ratios $\mathbf{h}_{t,k}$ of each asset k follow some ARMA(p_k, q_k) process.

Assumption 5.1.1. *For each security k*

$$\log(\mathbf{h}_{t,k}) = c_k + \sum_{i=1}^{p_k} \alpha_i \log(\mathbf{h}_{t-i,k}) + \sum_{j=1}^{q_k} \beta_j \boldsymbol{\varepsilon}_{t-j,k} + \boldsymbol{\varepsilon}_{t,k}.$$

Following the notation from Section 4.2 we see that function $f_k(\mathbf{h}_{t-1}, \mathbf{h}_{t-2}, ..., \mathbf{h}_{t-k}, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, ..., \boldsymbol{\varepsilon}_{t-l})$ is a usual forecast of ARMA model

$$\begin{aligned} & f_k(\mathbf{h}_{t-1}, \mathbf{h}_{t-2}, ..., \mathbf{h}_{t-k}, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, ..., \boldsymbol{\varepsilon}_{t-l}) = \mathbb{E}(\log(\mathbf{h}_{t,k})|\mathcal{F}_{t-1}) \\ & = c_k + \sum_{i=1}^{p_k} \alpha_i \log(\mathbf{h}_{t-i,k}) + \sum_{j=1}^{q_k} \beta_j \boldsymbol{\varepsilon}_{t-j,k}. \end{aligned} \tag{5.2}$$

Assuming 5.1.1 we should model each security on the market separately, but these models do not provide us with information on the dependence structure of $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_{t,1}, ..., \boldsymbol{\varepsilon}_{t,N})$. Therefore, we have to make an additional assumption. For the normal distribution it means that we need to specify the covariance matrix. Given some particular assets, professionals can give good enough *a priori* estimations of the dependencies. We will use the simpliest independence assumption and try to find assets for which this assumption is precise enough.

Assumption 5.1.2. *The covariance structure of error vector is*

$$\operatorname{cov}(\boldsymbol{\varepsilon}_t) = I_N,$$

where I_N is an identity matrix of size $N \times N$.

Given Assumptions 5.1.1, 5.1.2 and fitted ARMA models, we can construct GOP by Algorithm 1 for each time step t. The procedure for constructing GOP weights is then the following

Algorithm 3 GOP construction procedure. ARMA case.

```
1: procedure GOPconstructionARMA
2:     for t ∈ 𝕋 do
3:         for k ∈ 1,N (overlined) do
4:             Estimate ARMA model for security k.
5:         μ = forecast from ARMA models.
6:         σ = extract from ARMA models.
7:         GOP weights = Call function GOPweights(μ, σ) from Algorithm 1.
```

5.1.2 Data and Diagnostics Checks

It is quite hard to find securities on the markets for which we can assume that they are independent. Nevertheless, our goal is not to construct the exact growth optimal portfolio with real securities rather than demonstrate how one can do it. Therefore, for analysis, we took aggregate stocks index S&P500, Vanguard VGLT (Vanguard Long Term) ETF that tracks long-term US government bonds and Gold Futures. Stocks and government bonds are usually considered as orthogonal components of investing portfolio while the gold price is often called by investors as *safe heaven* i.e. perfect hedge for stocks and bonds on a bear market. For the broad analysis of the dependence structure of these three assets see [26].

We took monthly data from January 2010 to March 2019 for indices prices that is available online on *Investing.com* website. Note, that indices' prices incorporate dividend yields and, therefore, prices $\mathbf{P}_{t,i}$ match cumulative indices $\mathbf{S}_{t,i}$. For security k logarithms of growth rates are computed as in Equation 1.2.1

$$\log(\mathbf{h}_{t,k}) = \log\left(\frac{\mathbf{P}_{t,k}}{\mathbf{P}_{t-1,k}}\right). \tag{5.3}$$

	Gold Futures	VGLT	S&P 500
Count	110	110	110
Mean	0.0019	0.0023	0.0086
Std	0.047	0.032	0.036
Min	-0.13	-0.082	-0.096
25%	-0.025	-0.021	-0.009
50%	0.0035	0.00024	0.011
75%	0.032	0.02	0.029
Max	0.12	0.092	0.1

Table 5.1: Summary statistics of securities

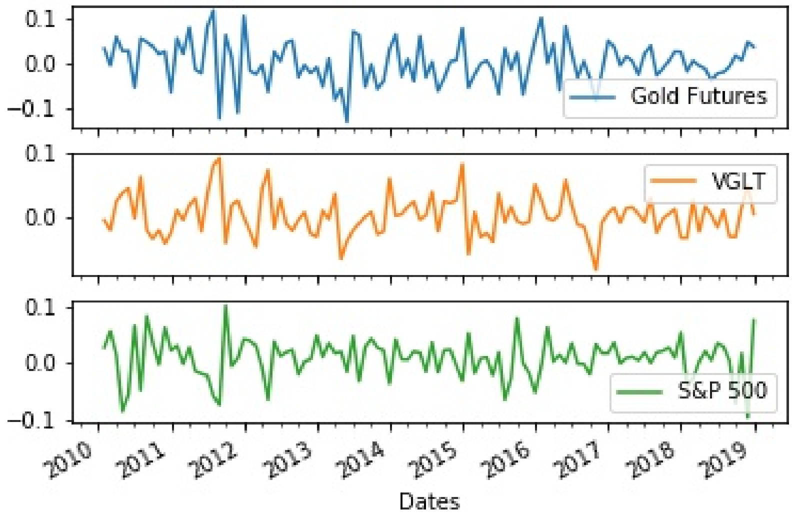

Figure 5.1: Logarithm growth ratios of securities

Plots 5.1 of log-growth-rates do not reflect any non-stationary patterns like deterministic and stochastic trends or volatility clusters. Unit root Dickey-Fuller (DF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests also support visual analysis and do not indicate violations from stationary on significance level 5% for all assets.

	DF test	KPSS test
Gold Futures	7.37-21	>0.1
VGLT	2.88e-15	>0.1
S&P 500	1.58e-21	>0.1

Table 5.2: Unit root tests' P-Values

Remark 5.4. Unit root tests are done by corresponding functions from the library *statmodels* [23]. KPSS function compares test statistics to the table values that are limited up to 10% P-value.

For order selection (p, q) I used *Bayesian Information Criteria* (BIC) criteria. Diagnostic checks were done for residuals from the ARMA models fitted to the whole data set. On significance level 5% Jarque-Bera test does not show violations of normality assumption and Ljung-Box Portmanteau test does not indicate autocorrelation in residuals.

	Jarque-Bera	Ljung-Box Portmanteau
Gold Futures	0.74	0.22
VGLT	0.19	0.23
S&P 500	0.073	0.13

Table 5.3: Diagnostic tests' P-Values

5.1.3 GOP simulation

Using Procedure 5.1.1 and data described the previous section we simulated the construction of Growth Optimal Portfolio with historical data. The simulation does not incorporate any transaction costs, so in practice returns of GOP would have been smaller. On the other hand, we did not do any parameters tuning and lookback tweaking like it is usually done when traders construct strategies, therefore the performance of GOP as a strategy could be better.

The simulation does not start from the first available date in the data set, because the start date should be chosen such that

- ARMA and VAR (Section 5.2) models are provided with enough data for accurate model parameters estimation.
- There is enough test data to see how GOP behaves in the long term.

Let's compare different investing strategies where capital is reinvested in each period. We can invest either to one of the assets or according to allocation suggested by GOP. In addition, let's compare the performance of GOP portfolio on the long run to the *Markowitz portfolio* [28] with the highest *Sharpe ratio* (Max Sharpe) based on historical estimations. For convenience, we constructed this portfolio using the python library *PyPortfolioOpt* [27]. In this simulation, models were retrained and portfolios were rebalanced each month.

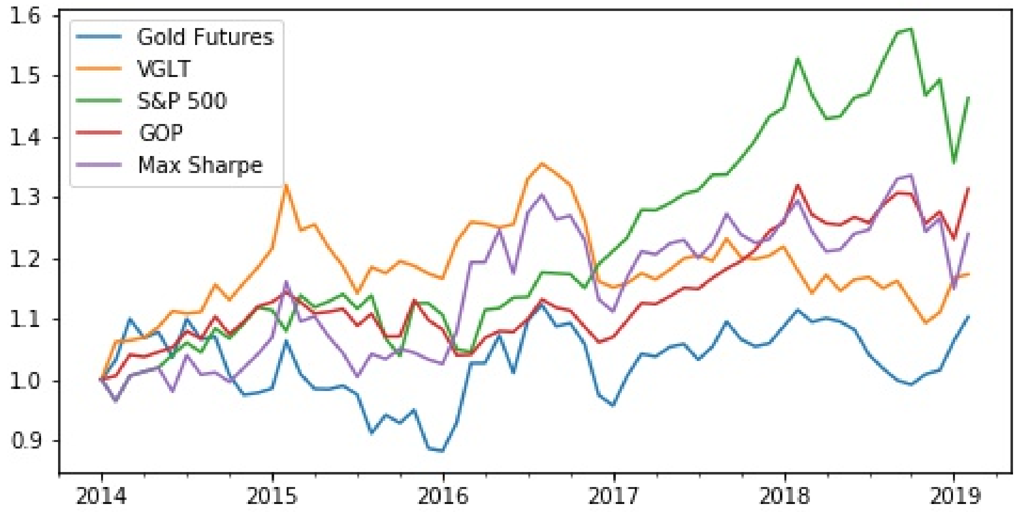

Figure 5.2: GOP simulation with ARMA models

We see that over 5 years of testing period the best performing asset is S&P500 index and GOP is the first runner-up. The testing period is not very large, so we cannot guarantee that even exact GOP portfolio outperforms other assets with high probability. Nevertheless, the overall return of GOP is higher than return Max Sharpe Portfolio and GOP does not have dramatic drawdown periods, unlike all underlying assets. It may seem that performance of GOP is high due to the fact that it tends to allocate all capital to stocks index that occasionally had the best return over the test period, but on average GOP allocated only 63% of weight to S&P500. Moreover, in our simulations, GOP allocations were always diversified despite criticism that GOP tends to produce sparse portfolios, especially in a one-period market setting.

	Avg. Weights
Gold Futures	16%
VGLT	21%
S&P 500	63%

Table 5.4: Average Growth Optimal Portfolio Weights

Another surprising fact is that GOP has a higher Sharpe Ratio then all underlying securities and Max Sharpe portfolio. Meanwhile Max Sharpe portfolio is constructed to maximize this performance measure using historical data.

	Sharpe Ratio
Gold Futures	0.023
VGLT	0.076
S&P 500	0.17
GOP	0.19
Max Sharpe	0.069

Table 5.5: Sharpe Ratios

5.2 VAR model

5.2.1 Specification

Vector Auto Regression model allows us to capture linear interdependencies between components of some stochastic vector process. Standard VAR model with p lags for a stochastic vector process $\mathbf{X}_t = (\mathbf{X}_{t,1}, ..., \mathbf{X}_{t,N})$ states that

$$\mathbf{X}_t = c + \sum_{i=1}^{p} A_i \mathbf{X}_{t-i} + \boldsymbol{\varepsilon}_t, \tag{5.5}$$

where $\boldsymbol{\varepsilon}_t$ strong white noise vector process with multivariate normal distribution i.e.

$$\boldsymbol{\varepsilon}_t \sim N(0, \Sigma), \tag{5.6}$$

and

$$\text{For } t \neq s : \boldsymbol{\varepsilon}_t \perp\!\!\!\perp \boldsymbol{\varepsilon}_s. \tag{5.7}$$

Vector $c \in \mathbb{R}^N$ and matrices $A_i \in \mathbb{R}^{N \times N}$ are model parameters.

Remark 5.8. One can consider a vector analogue of ARMA(p,q) process - VARMA(p,q) model. However, there exists a fundamental identification problem that the coefficient matrices are not generally unique, see Chapter 12 in [25].

As for the ARMA model, we will assume that vector of the growth ratios' logarithms $\log(\mathbf{h}_t)$ is a VAR (p) process.

Assumption 5.2.1. *For any $t \in \mathbb{T}$ the log growth ratio can be represented as*

$$\log(\mathbf{h}_t) = c + \sum_{i=1}^{p} A_i \log(\mathbf{h}_{t-i}) + \boldsymbol{\varepsilon}_t, \tag{5.9}$$

where $\boldsymbol{\varepsilon}_t$ satisfy usual VAR properties 5.6 and 5.7.

To identify the function $f_k(\mathbf{h}_{t-1}, \mathbf{h}_{t-2}, ..., \mathbf{h}_{t-k}, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, ..., \boldsymbol{\varepsilon}_{t-l})$ from Section 4.2 let's rewrite Equation 5.9 for each component k.

$$\log(\mathbf{h}_{t,k}) = c_k + \sum_{i=1}^{p} \sum_{j=1}^{N} a^i_{k,j} \log(\mathbf{h}_{t-i,j}) + \boldsymbol{\varepsilon}_{t,k}, \tag{5.10}$$

where $a^i_{k,j}$ is a $[k, j]$ component of the matrix A_i.
Thus, similarly to ARMA model we get that

$$\begin{aligned} & f_k(\mathbf{h}_{t-1}, \mathbf{h}_{t-2}, ..., \mathbf{h}_{t-k}, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, ..., \boldsymbol{\varepsilon}_{t-l}) = \mathbb{E}(\log(\mathbf{h}_{t,k})|\mathcal{F}_{t-1}) \\ & = c_k + \sum_{i=1}^{p} \sum_{j=1}^{N} a^i_{k,j} \log(\mathbf{h}_{t-i,j}). \end{aligned} \tag{5.11}$$

In contrast to the ARMA model, we do not need to make any additional assumptions on the covariance matrix structure since it is estimated directly during fitting process of a VAR model. Despite this benefit, from representation 5.10 wee see that VAR model involves much more parameters for estimation given the same number of lags. Therefore, we need enough data to avoid overfitting of the model.

The procedure of GOP construction with VAR model is similar to ARMA case.

Algorithm 4 GOP construction procedure. VAR case.

procedure GOPconstructionARMA
  for $t \in \mathbb{T}$ **do**
    Estimate VAR model for all securities.
    μ = forecast from VAR model.
    Σ = extract residuals covariance matrix from VAR model.
    GOP weights = Call function $GOPweights(\mu, \Sigma)$ from Algorithm 2.

5.2.2 GOP simulation

For the simulation, we will use the same dataset as in Section 5.1.2. In the ARMA case, we used Assumption 5.1.2 that log-growth-ratios of underlying securities are independent. However, this assumption is quite unrealistic even for the selected assets. Therefore, we can use the VAR model and estimate the covariance structure of residuals. For the construction, we again use the implementation of VAR model from python *statsmodels* package. The fit function automatically selects the required order of VAR by *BIC* and time series are automatically checked for cointegration. In our case we do not face cointegrated time series, thus, Vector Error Correction Model (VECM) is not needed and we can perform simulations with VAR model.

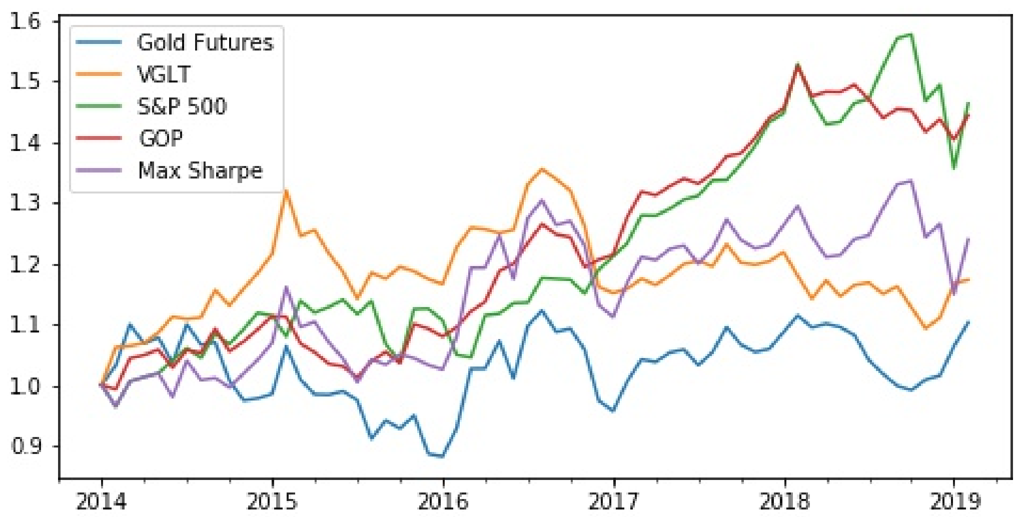

Figure 5.3: GOP simulation with ARMA models

	Avg. Weights
Gold Futures	28%
VGLT	28%
S&P 500	44%

Table 5.6: Average Growth Optimal Portfolio Weights

Plot 5.2.2 shows that cumulative return of GOP with underlying VAR model over the test period is approximately the same as the return of S&P500 index. On average, the portfolio is more diversified than in the case of ARMA models, so again we cannot say that the portfolio performed well just because it favours risky stock index. GOP's Sharpe Ratio (0.25) over this period is higher then Sharpe Ratio of GOP in the ARMA case(0.19). To sum up, we see that the inclusion of linear interdependencies between underlying assets has a positive impact on GOP performance as an investment strategy.

6. Conclusions

In this thesis, we tried to cover most of the aspects of Growth Optimal Portfolio on a discrete market. In the theoretical part of the thesis, we gave proofs for important optimal properties, derived necessary and sufficient conditions for optimality and provided examples of GOP in simple cases when they have closed-form solutions.

The main result of this thesis is Iterative Algorithm 1 (presented in Section 4.3) for constricting GOP in the case when growth ratios of underlying assets are log-normally distributed. We derived iterative equations for updating weights, utilized variance reduction techniques and extended the algorithm to the cases with risk-free asset and non-diagonal covariance structure of growth ratios. Algorithms 1 and 2 have shown good convergence properties on the toy examples. However, there were simulations where general optimization algorithm shows better results, meaning that our method did not converge to the optimal value. For example, Algorithm 2, in general, performed worse in the cases where the covariance matrix has a small determinant due to numerical issues. Finally, we have shown how we can apply Algorithm 1 when assets are modelled by linear time series models (ARMA and VAR).

The simulation of Growth Optimal Portfolio within time period 2014-2019 indicate some interesting aspects of GOP. Namely, GOP demonstrated quite good cumulative return and the best risk-adjusted return measured by Sharpe Ratio. Moreover, the allocations of GOP were quite diversified opposed to the common criticism of GOP for providing sparse solutions in the discrete market setting. Even though the simulation shows favorable performance of GOP, it is clear that for proper VAR modelling we need to use bigger samples sizes, since information criteria always preferred simple models with not more than 2 lags. For example, with monthly data, this makes yearly seasonality difficult to model.

Summing up, I think that Growth Optimal Portfolio is still a promising concept both as a tool for pricing securities and as a trading strategy. I believe that simple construction procedures that are derived in this thesis can shed some light on the utilization of Growth Optimal Portfolio in the investing process.

Bibliography

[1] Kelly, J. (1956). A new interpretation of information rate. IRE Transactions on Information Theory, 2 (3), p. 185-189.

[2] Williams, J. B. (1936). Speculation and the carryover, The Quarterly Journal of Economics 50, 3, pp. 436-455.

[3] Györfi, L. (2012). Machine Learning for Financial Engineering. Chaper 1. Christensen, M. M. "On the History of the Growth-Optimal Portfolio". pp. 1-79.

[4] Breiman, L. (1960). Investment policies for expanding businesses optimal in a long-run sense, Naval Research Logistics Quarterly 7, 4, pp. 647-651.

[5] Buhlmann H. ; Platen E. (2002). A Discrete Time Benchmark Approach for Finance and Insurance, Research Paper Series 74, Quantitative Finance Research Centre, University of Technology, Sydney.

[6] Merton, R. (1969). Lifetime portfolio selection under uncertainty: The continuous time case, Review of Economics and Statistics 51, 3, pp. 247-257.

[7] Björk, T. (1998). Arbitrage theory in continuous time. Oxford: Oxford University Press.

[8] Platen, E. (2006). A Benchmark Approach to Quantitative Finance.

[9] Vander Weide, J. (1977). The Strategy Which Maximizes the Geometric Mean Return is Portfolio Investments. Management Science, 23 (10), p. 1117-1123.

[10] Kuhn, H. W.; Tucker, A. W. (1951). Nonlinear programming. Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. pp. 481-492.

[11] Truc Le, Eckhard Platen, (2006). Approximating the growth optimal portfolio with a diversified world stock index, The Journal of Risk Finance, Vol. 7 Issue: 5, pp.559-574,

[12] Nourdin, I. (2012). Normal approximations with Malliavin calculus: From Stein's method to universality. Cambridge [England] ; New York: Cambridge University Press.

[13] Maier, S. F. (1977). A Monte Carlo Investigation of Characteristics of Optimal Geometric Mean Portfolios. Journal of Financial and Quantitative Analysis, 12(2), pp. 215-233.

[14] Ariyo, A. A. (2014). Stock Price Prediction Using the ARIMA Model.International Conference on Computer Modelling and Simulation , pp. 106-112. IEEE.

[15] Mondal, P. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications, 4(2), pp. 13-29.

[16] Maghyereh, A. (2006). Oil price shocks and emerging stock markets: A generalized VAR approach. In Global stock markets and portfolio management (pp. 55-68). Palgrave Macmillan, London.

[17] Kumar Kattumannil, S. (2009). On Stein's identity and its applications. Statistics and Probability Letters, 79(12), pp. 1444-1449.

[18] Kraft, D. (1994). Algorithm 733: TOMP-Fortran modules for optimal control calculations. ACM Transactions on Mathematical Software (TOMS), 20(3), pp. 262-281.

[19] Kroese, D. P.; Taimre, T.; Botev, Z. I. (2011). Handbook of Monte Carlo methods. John Wiley & Sons.(Chapter 9.3).

[20] J. Duchi, Properties of the Trace and Matrix Derivatives. Available online: https://web.stanford.edu/ jduchi/projects/matrix_prop.pdf

[21] D. Dufresne. (2009). Sums of Lognormals. Available online: https://pdfs.semanticscholar.org/38a3/9907d5af01a6340c3b6f38dce39be3be5e8a.pdf

[22] Pagès, G. (2018). Numerical probability: An introduction with applications to finance. Cham: Springer.

[23] Skipper S., Perktold J. (2010). Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference.

[24] Verbeek, M. (2017). A guide to modern econometrics (Fifth edition.). Hoboken (NJ): Wiley.

[25] Lütkepohl, H. (2007). New Introduction to Multiple Time Series Analysis. Berlin: Springer.

[26] Baur, DG (2010). Is Gold the Hedge or Safe Haven? An Analysis of Stocks, Bonds and Gold. Financial Review, 45 (2), p. 217-229.

[27] Martin R.A. (2018). PyPortfolioOpt v. 0.2.0. Revision d33666e3.

[28] Markowitz, HM (1987). Mean-variance analysis in portfolio choice and capital markets . Oxford: Basil Blackwell.