Financial Risk Management in Social and Pension Insurance

Chapter I: Introduction

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 13, 2020

Table of Contents

1. Administrative matters

About the lecturer: Peter Blum

Educational background:

- Professional degree in electronics and software engineering.
- Diploma in mathematics from ETH Zurich.
- PhD in financial and insurance mathematics from ETH Zurich [3].
- Chartered Financial Analyst (CFA), CFA Institute.
- Fully qualified actuary (Aktuar SAV) of the Swiss Actuarial Society

Professional experience:

- Industrial software engineering at Landis & Gyr and Siemens.
- Asset / Liability Management, research and financial engineering at Zurich Re, then Converium (now Scor).
- With Suva - The Swiss National Accident Insurance Fund since 2005, initially as an analyst, then heading teams in asset allocation, treasury and research.
- Deputy head of portfolio management during the Great Financial Crisis.
- Chief Risk Officer and Chief of Staff of the Finance Department since 2011, with responsibilities both on the asset and on the liability side.
- Also heavily involved in the strategic management of the Suva Pension Fund.

Institutional background

Suva - The Swiss National Accident Insurance Fund (figures as of 2019)

- Provider of compulsory accident insurance for the secondary and public sectors in Switzerland since 1918.
- 2.067 million insured persons in 130'000 insured companies and institutions.
- 479'746 cases of accident and professional disease processed.
- Benefits of CHF 4.5 billion paid, including medical costs, daily indemnities and 83'709 disability and survivors' pensions totaling CHF 1'648 million.
- Assets under management of CHF 53.8 billion.

Suva Pension Fund (figures as of 2019)

- Pension fund according to Swiss law (2nd Pillar) for the employees of Suva, in operation since 1985.
- 3'958 active and 2'121 retired members.
- Contributions of CHF 115 million, benefits of CHF 111 million.
- Assets under management of CHF 3.3 billion.

Course schedule

Lectures every Wednesday afternoon from 16:15 to 18:00

- Lectures will take place online via Zoom.
- Lectures will be recorded, and recordings will be available to students.
- Changes in course schedule will be announced by e-mail.

Lecturer available for questions after the lectures and during the break.

Otherwise: peter.blum@math.ethz.ch

Course language: English.

Documentation

Slide presentations as made available on the ETH web platform plus whiteboard notes made in the lectures.

Slides are made such that they can be printed black-and-white and 2-up without loss of information.

Not all presentations have been published yet. Publication of further chapters or revised versions of already-published chapters will be announced by e-mail.

If you find mistakes in the presentations, please tell the lecturer.

Excel sheets and R code made available on the web platform are for illustration only.

References given, unless otherwise stated, are for information only.

Exams

Oral exam of 30 minutes.

During the ordinary exam session (next one: January 25 to February 19, 2021).

Examination language: English (German may be available upon demand).

Further information and instructions in the last lecture.

2. Properties of social insurance

Risks covered by social insurance

The most common forms of social insurance include

- Old-age provision, i.e. pension insurance. In its capital-based variant the most important type of social insurance covered in this course.
- Health insurance, at least the basic, compulsory part of it.
- Accident insurance, if it is separate from health insurance. May be further subdivided into insurance against occupational accidents and diseases (Workers' Compensation) and insurance against spare-time accidents.
- Disability insurance, to the extent that it is not yet covered by other types of insurance providers.
- Unemployment insurance.
- Etc. (In Switzerland, for instance, building / fire insurance is essentially also a type of social insurance.)

Organization, benefits and financing of the social insurance system vary greatly from country to country.

Properties of social insurance

Social insurance is usually mutually compulsory, i.e. clients must take out the insurance, and insurers must accept the clients:

- There is a defined market; growth corresponds to the growth of the client base as specified by law.
- No risk selection; in particular, no adverse selection. Insurers can and must insure both "good" and "bad" risks.
- On the other hand, there is an assured client base for the insurer over a long time, which is beneficial for long-term planning.

Defined insurance benefits for defined insured risks:

- Little to no room for product design.
- No growth opportunities through product innovation.

Generally high regulation density; usually regulation through special laws and special regulatory bodies differing from the ones for private insurance. This may include special accounting standards.

Social insurance institutions are often not publicly listed companies. Many social insurance institutions have a special legal status:

- Swiss pension funds are usually foundations, legally and financially separated from their sponsoring institutions. There is also a legal obligation to keep the foundation funded. This is an extremely effective means of risk management.
- Many social insurance institutions are outright governed by special law, e.g. also Suva.
- But, in some cases, also normal publicly listed companies or private mutual societies may be carriers of social insurance.
- However, in general, analytic approaches based on the economic theory of the firm (see e.g. [6]) are not applicable in a social and pension insurance context and will not be discussed any further in this course.

Social insurance is usually not-for-profit, but financial stability and financial sustainability are of utmost importance:

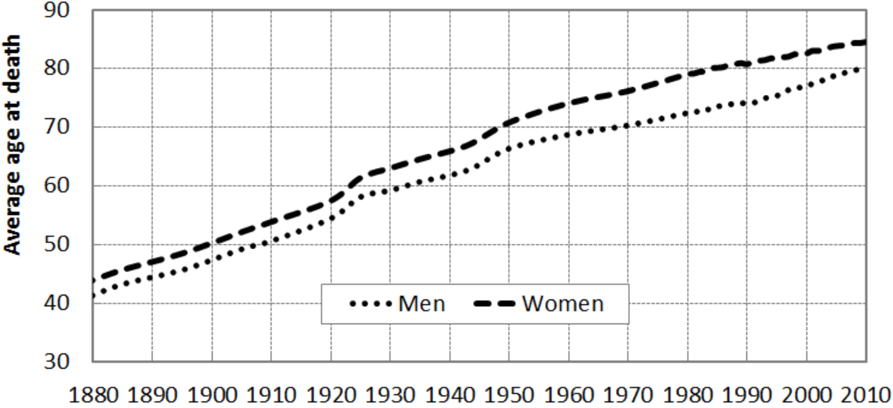
- Therefore, social insurance is still insurance - even under particularly demanding conditions - and thus a valid field for actuarial reasoning.
- We do not consider social benefits that are paid from the general public budget and, hence, are not insurance in our sense.
- Limited profit-taking may be allowed if private companies provide social insurance services, in order to compensate them for their cost of capital.

The most important feature of social and pension insurance, however, is the often very long time horizon:

- Think of pension insurance: The saving process starts e.g. at age 25, whereas life expectancy is above 80 (and increasing), with a substantial chance for people to make it well into their nineties.
- Taking into account disability and long-term care, the time horizons in accident insurance are by no means shorter.
- Or think of asbestos-related diseases: a malignant mesothelioma may break out 20 or 40 or even more years after the actual exposure.

Illustration: life expectancy

Pension insurance and also other forms of social insurance must remain available over the entire life span of their clients.



Basic financing modes

For pension insurance, there are two basic financing modes:

Pay-as-you-go:

- Current benefits are paid from current contributions.
- Current contributors acquire the right to receive benefits in the future.
- No (or no substantial amount of) money is accumulated. Hence, there are no (substantial) investment proceeds that contribute to the financing, and no risks either from investing the money.

Capital-based insurance:

- Contributions of each contributor are accumulated and then used to pay old-age benefits for that person when they are due.
- While being accumulated, the money is invested. This generates investment proceeds that are a source of funding.
- Substantial financial risk may arise from the investment of the accumulated contributions.

Illustration: pay-as-you-go

Pay-as-you-go is basically an inter-generational contract.

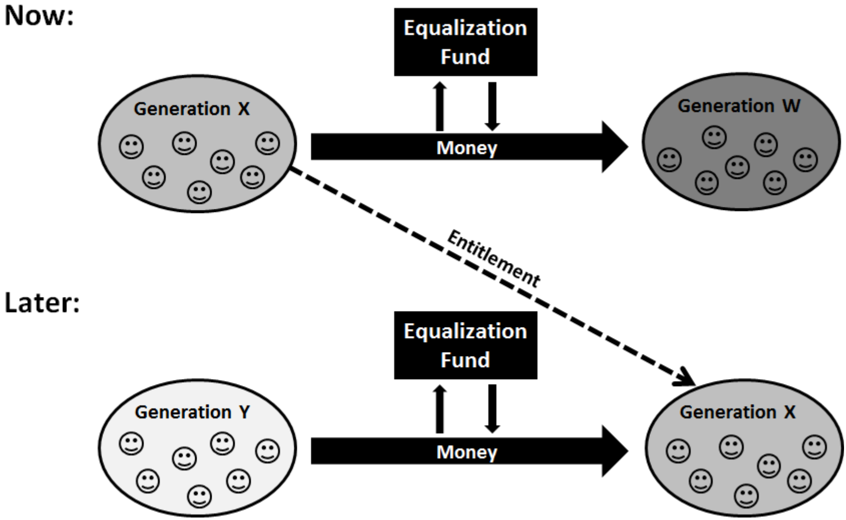


Illustration: capital-based insurance

In capital-based insurance, each generation (in principle) cares for itself.

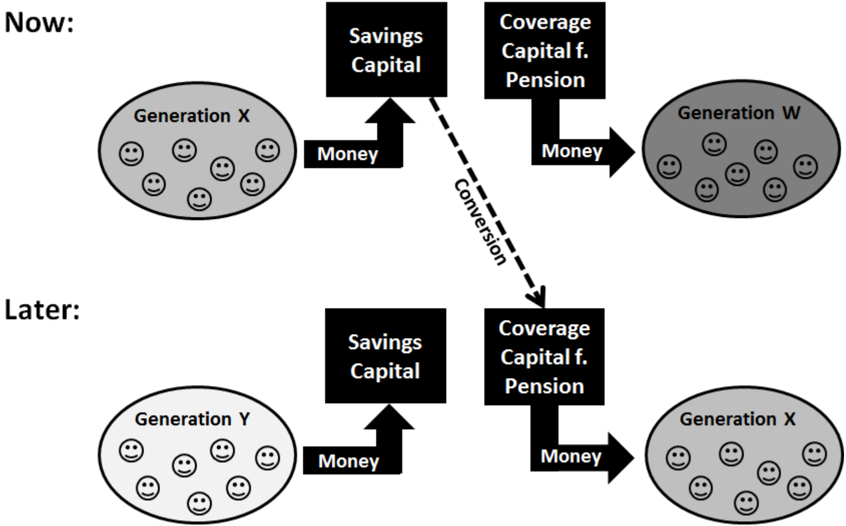
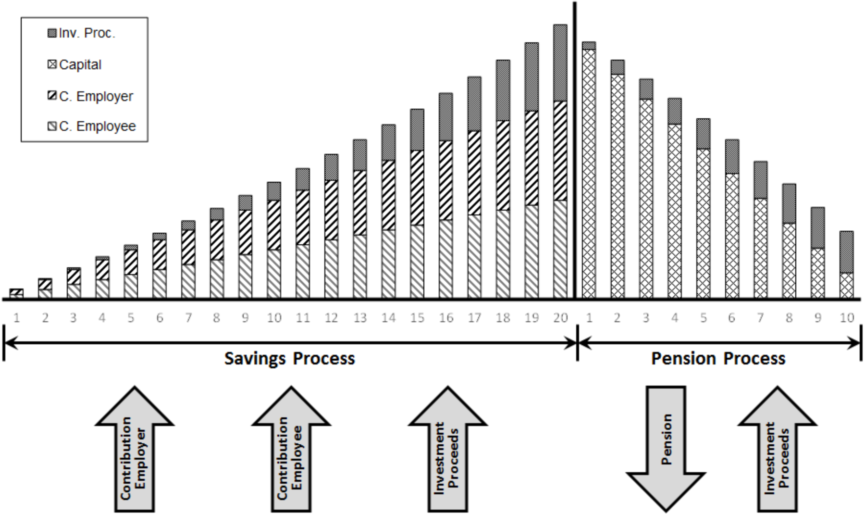


Illustration: capital-based old-age insurance process

Our big question: How much can we expect from the investment proceeds? And what investment risk do we take in turn?



3. About this course

Focus and goal of this course

Since substantial financial risk only arises in the capital-based setup, we will focus on the latter and we will not treat pay-as-you-go any further.

Given the properties of social and pension insurance as outlined before, we will have to work in a long-term, multi-period framework. This represents a big challenge:

- Financial risk in short-term, single-period and asset-only setups is very well understood; see [5].
- Financial risk in a long-term, multi-period setup and in the presence of liabilities is less well-understood, and some original reasoning will be necessary.

We will take a micro perspective, i.e. the point of view of a single social insurance institution or pension fund. This is the situation that actuaries will most likely be faced with in practice.

Macro considerations aimed at designing entire pension systems are undoubtedly also interesting, but beyond the scope of this course. See [2] for more on this.

About stochastic simulation

In practice, settings are usually too complicated for comprehensive analytical tractability. Therefore, stochastic simulation is usually applied as a tool of analysis.

Hence, a solid understanding of the methods and numerics of stochastic simulation is indispensable for an analyst wanting to solve practical problems related to financial risk in social and pension insurance. A good textbook is e.g. [1].

- The usefulness of many studies is hampered or even negated by technical mistakes such as too few simulation runs, lack of care for variance and dependence, lack of care for the tails, etc.

Numerical quality is necessary, but not sufficient. Most importantly, models must be well-designed, i.e.

- They must be parsimonious and focus on the really important influence factors and distinguish them from the less relevant ones.
- They must treat the relevant influence factors properly.
- In particular, they must give due care to the dependencies and the interplay between the influence factors.

In brief, good models must be simple, focused on the relevant factors and risk measures, and modular. This is where this course attempts to add value.

Call from practice:

If you have to present important results in front of a board of directors or trustees (which usually mainly consists of non-quantitative people) you must:

- perfectly understand what you are talking about, and
- be able to explain everything in simple, understandable and conclusive words.

If you are able to do this, people will likely accept your proposals, and they will likely trust the scientific models underneath the conclusions, even if they cannot understand them in detail.

Two voices on this topic:

- "Make things as simple as possible, but not simpler." (Attributed to Einstein)
- "Whereof one cannot speak, thereof one must remain silent." (Wittgenstein, Tractatus logico-philosophicus)

Concept of this course

Our focus will be on Asset / Liability Management (ALM), i.e. the interface between assets and liabilities, throughout the course. For assets and liabilities on a standalone basis, there are other courses.

We will develop a relatively simple modeling framework that focuses on the top-level influence factors, policy variables and risk measure as well as on their interplay. This modeling framework abstracts, but comprises many subordinated details.

Under some assumptions, we will obtain an analytically tractable model that provides closed-form solutions for a number of important questions. In this analytic model, we will be able to develop a solid understanding of the most relevant ALM factors and their interplay:

- required return / liability growth rate
- expected return
- risk-taking capability
- short-term (investment) risk
- long-term (financial) risk

The most interesting and most intriguing aspect will be the relationship between short-term and long-term risk. We will see that reducing short-term risk will not necessarily reduce long-term risk.

On good risk management

Good risk management is not simply the avoidance or reduction of risk. Good risk management actually means considerate risk-taking with long-term financial stability as the goal.

The tools developed and the insights gained in the analytical framework are also useful in more general settings where stochastic simulation must be used:

- They allow to build well-focused and well-structured models.
- They provide useful analytics to assess the financial condition of an institution.
- They allow to better understand and appraise the outcomes of studies of all kinds and to draw valid conclusions.

Very fundamentally, it is not the goal of this course to provide cookbook-type recipes. This is rather about enabling people to think independently and in a well-structured manner about problems related to financial risk in a social and pension insurance context and to develop valid and sensible tailor-made solutions for these problems.

Mathematical prerequisites

While this course aims at being as self-contained as possible, some mathematical knowledge is nevertheless taken for granted:

- Knowledge of basic concepts of probability theory and statistics on the level of an undergraduate course.
- Knowledge of calculus and linear algebra on the level of undergraduate courses.
- Other concepts will be introduced in due course.

We will see that even fairly simple mathematical reasoning - when applied rigorously and focused on the relevant factors - can provide very deep and useful insights.

This is very useful nowadays when plentiful computing power allows to create simulation models that overwhelm the analyst with a plethora of more or less useful information. Staying focused on the relevant factors and issues is really crucial.

It would be possible to formalize the problems treated in this course in much more sophisticated manners (e.g. in continuous time with diffusion processes and an incomplete-markets setup); this is, however, still work in progress.

4. Assuring future pensions: the danger of taking no risk

Introduction

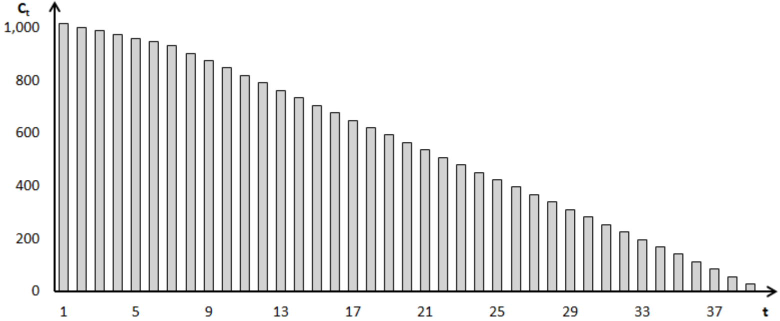
The following presentation is intended to motivate the most important problems, notions and concepts that will be treated throughout this course.

It is from a talk given on Risk Day 2017 at ETH Zurich.

It specifically shows the situation of Swiss pensions funds. But it is representative for many other capital-based social insurance institutions worldwide.

The Task: Financing Future Pensions

Let t denote time in years. We have some stream of future promised cashflows $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$, assumed to be deterministic here, e.g.:



At time $t = 0$ (i.e. now), we must put up an amount A_0 of assets in such a way that the payment of all future cashflows C_t is assured. These assets can be invested, and the returns from doing so must be taken into account.

First Approach: Immunization

Let $P(0, s)$ denote the value at time $t = 0$ of a zero-coupon bond (ZCB) maturing at time $t = s$, with $P(s, s) = 1$.

Now, we replicate the cashflow stream $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ by using ZCBs: At time $t = 0$ and for each time $t > 0$, we buy C_t units of the ZCB maturing at t, at a price of $P(0, t)$ per unit. The total cost of this is

$$A_0 = \mathsf{PV}_0^{\mathsf{ZCB}}(\mathbf{C}) = \sum_{t=1}^{T} P(0, t)\, C_t$$

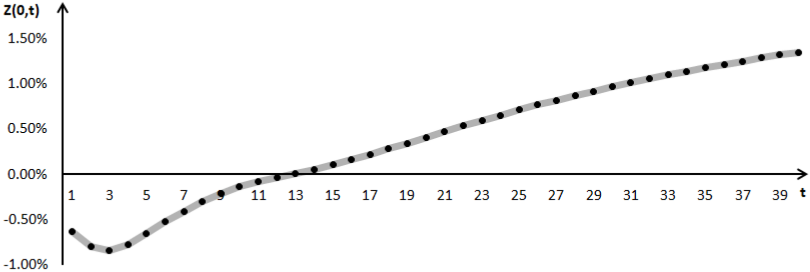
In the absence of credit risk, the payment of all future promised cashflows C_t is thus assured. That is, we have managed to create an immunized position at time $t = 0$, and all subsequent financial risk is eliminated.

Therefore, the immunized value $\mathsf{PV}_0^{\mathsf{ZCB}}(\mathbf{C})$ is a viable valuation of the cashflow stream $\mathbf{C}$.

Note: There is no ZCB market in CHF, but an essentially equivalent portfolio can be constructed from coupon bonds, interest rate swaps and bond futures - at least for time horizons not exceeding 30 years.

The Cost of Immunization

We use the ZCB curve for the CHF as mandated by the Finma for the SST 2017:



Based on this curve, we obtain the following immunized value for our example cashflow stream:

$$\text{PV}_0^{\text{ZCB}}(\mathbf{C}) = \sum_{t=1}^{T} P(0,t)\, C_t = \text{20'726 MCHF}$$

This corresponds to 95% of the sum of undiscounted cashflows of 21'730 MCHF. That is, we must put up almost the entire amount up front; there is only a very small contribution of interest rates to the financing.

The Statutory Setup of Swiss Pension Funds

In reality, Swiss pension funds do not use market-derived ZCB curves for the valuation of their liabilities. Rather, they use a constant discount rate λ:

$$L_0 = \mathsf{PV}_0(\mathbf{C}, \lambda) = \sum_{t=1}^{T} \frac{C_t}{(1+\lambda)^t}$$

This discount rate λ, called technical interest rate, is fixed by the board of trustees of each pension fund.

In order to make this comparable to the immunized approach, we can specify the equivalent immunized discount rate

$$\lambda^{\mathsf{ZCB}} := \mathsf{IRR}\left(\mathsf{PV}_0^{\mathsf{ZCB}}(\mathbf{C}),\, \mathbf{C}\right)$$

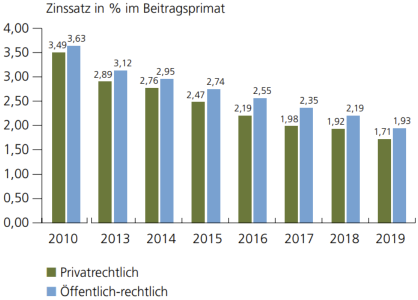
which equates the two valuations, i.e.

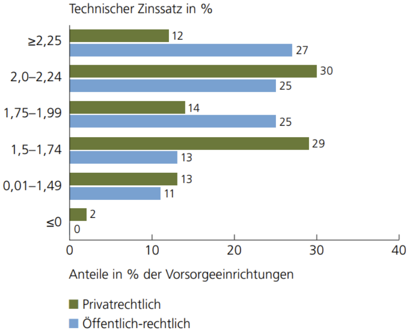
$$\mathsf{PV}_0\left(\mathbf{C}, \lambda^{\mathsf{ZCB}}\right) = \mathsf{PV}_0^{\mathsf{ZCB}}(\mathbf{C})$$

In our specific example, we have $\lambda^{\mathsf{ZCB}} = 0.35\%$, and this would also be the order of magnitude for other Swiss pension funds.

Technical Interest Rate: Development & Current Situation

From Swisscanto, Schweizer Pensionskassenstudie 2020 [7] (520 institutions with total AuM of CHF 772 bln. and 3.8 mln. insured persons).





Technical interest rates λ were reduced considerably over the past ten years. But with values well above 2%, they are far away from values in the immunized case, which would be below 0.5%.

The Effective Situation of Swiss Pension Funds

Swiss pension funds are far away from immunized positions. Achieving this would have a strong impact on the valuation of their liabilities; in our example:

Technical Interest Rate λ	Value of Liablities $PV_0(C,\lambda)$	Cost of Immunization
0.35%	20,726	n/a
1.50%	17,854	16%
2.00%	16,799	23%
2.50%	15,840	31%

Achieving immunized positions would thus require drastic measures:

- Either the injection of large sums of additional financing.
- Or severe reductions of benefits: At $\lambda = 0.35\%$, conversion rates (UWS) would be below 4%, whereas at $\lambda = 2\%$, they are between 4.5% and 5%.

Thus, rather than immunizing their liabilities, Swiss pension funds try to finance their higher discount rates by running mixed investment strategies with considerable shares of risky assets such as equities or real estate.

This then creates financial risk, i.e. the possibility that some of the promised payments may not be made due to investment losses. How can this situation be dealt with?

The Essential Asset / Liability Management Setup

With A_0 and L_0 given, we consider the following generic Asset / Liability model:

$$\left.\begin{aligned} A_t &= A_{t-1}(1+R_t) + C_t \\ L_t &= L_{t-1}(1+\lambda) + C_t \end{aligned}\right\} \qquad \text{for} \quad t \in \{1, \ldots, T\}$$

$A_t =$ value of assets at time t

$L_t =$ value of liabilities at time t

$C_t =$ net cashflow from insurance, i.e. contributions - paid benefits - costs

$\lambda =$ intrinsic growth rate of liabilities; here equal to the technical interest rate

$R_t =$ investment returns: $R_t \sim$ iid with $\mathbf{E}[R_t] \equiv \mu$ and $\mathbf{Var}[R_t] \equiv \sigma^2 < \infty$

That is, the liabilities develop deterministically, whereas the assets follow a random walk. The quantity of interest for further investigations is the funding ratio, i.e.

$$\mathsf{FR}_t = \frac{A_t}{L_t}$$

This is then also a random walk. We would like to have $\mathsf{FR}_t \geq 1$, i.e. the assets at time t cover the liabilities from time t on.

Long-term Financial Risk

Long-term financial risk consists of the possibility that some of the promised cash-flows may not be paid due to adverse investment returns R_t over one or several periods. Quantification of long-term financial risk is, therefore, associated with the state of underfunding, i.e. $\mathsf{FR}_\tau < 1$ for some fixed time horizon $1 \leq \tau \leq T$.

The most widespread risk measure used is the probability of underfunding, i.e.

$$\psi_\tau := \mathbf{P}\left[\mathsf{FR}_\tau \leq 1 | \mathsf{FR}_0\right]$$

As usual in quantitative risk management [5], we would like to have a risk measure that also incorporates the extent of the underfunding. To this end, we define the Expected Funding Shortfall, i.e.

$$\mathsf{EFS}_{\alpha,\tau} = 1 - \mathbf{E}\left[\mathsf{FR}_\tau | \mathsf{FR}_\tau \leq q_\alpha(\mathsf{FR}_\tau)\right]$$

Here, $q_\alpha(\mathsf{FR}_\tau)$ is the left α-quantile of the distribution of FR_τ given FR_0. We consider the difference to 1 because this is the funding shortfall that must be filled up in order to re-attain fully-funded status.

A Simple Analytical Model for Long-term Financial Risk

If the fund is in equilibrium, i.e. $C_t \equiv 0$, and if we assume (somewhat dangerously) that $R_t \sim$ iid $\mathcal{N}(\mu, \sigma^2)$, then FR_τ given FR_0 is log-normally distributed, and we have for the probability of underfunding:

$$\psi_\tau(\mathsf{FR}_0, \lambda, \mu, \sigma^2) = \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)\tau}{\sigma\sqrt{\tau}}\right)$$

and for the Expected Funding Shortfall [4]:

$$\mathsf{EFS}_{\alpha,\tau}(\mathsf{FR}_0, \lambda, \mu, \sigma^2) = 1 - \tfrac{1}{\alpha}\,\mathsf{FR}_0 \exp\left\{\left(\mu - \lambda + \tfrac{1}{2}\sigma^2\right)\tau\right\}\,\Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{\tau}\right)$$

These formula allow to get a good intuition of the interplay between the various variables, in particular:

$\frac{\partial\,\mathsf{EFS}_{\alpha,\tau}}{\partial\,\lambda} > 0 \;\Rightarrow\;$ required return

$\frac{\partial\,\mathsf{EFS}_{\alpha,\tau}}{\partial\,\mu} < 0 \;\Rightarrow\;$ expected return

$\frac{\partial\,\mathsf{EFS}_{\alpha,\tau}}{\partial\,\mathsf{FR}_0} < 0 \;\Rightarrow\;$ risk-taking capability

$\frac{\partial\,\mathsf{EFS}_{\alpha,\tau}}{\partial\,\sigma} > 0$ under realistic conditions $\Rightarrow$ short-term investment risk

The Task of Asset / Liability Management

If we cannot avoid long-term financial risk altogether by immunizing, we should at least set up our fund in such a way that long-term financial risk is minimized.

- The main policy variable is the technical interest rate λ that can be fixed by the board of trustees.
- Moreover, the investment strategy can be designed so as to attain a certain expected return μ. For a sustainable funding, we must have $\mu \geq \lambda$, and we specify here that $\mu = \lambda$.

However, given the choices of λ and $\mu = \lambda$, the other variables cannot be chosen freely anymore:

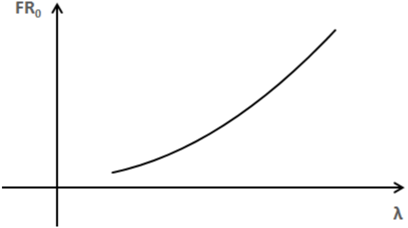
- Since initial assets A_0 are given, the initial funding ratio FR_0 depends on the technical interest rate λ, i.e. $\mathsf{FR}_0 = \mathsf{FR}_0(\lambda)$. This is the liability profile, and its properties are determined by the liabilities of the pension fund.
- A certain level of return $\mu = \lambda$ entails a certain level of short-term investment risk σ, i.e. $\sigma = \sigma(\mu) = \sigma(\lambda)$. This is the risk / return profile, and its properties are determined by the financial markets and by applicable investment constraints.

We must, therefore, optimize long-term financial risk under these contingencies.

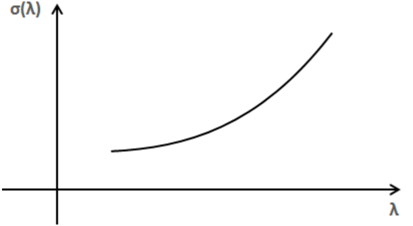
The Liability Profile and the Risk / Return Profile

In realistic settings, the liability profile and the risk-return profile are both convex:

Liability profile:



Risk / return profile:



In principle, these developments are compatible with one another:

- Higher technical interest rate $\lambda \Rightarrow$ higher required return $\mu \Rightarrow$ higher short-term investment risk σ, but also higher risk-taking capability FR_0.
- And vice versa.

The question now is which one of the two profiles dominates, and whether there is an optimum in terms of long-term financial risk somewhere within the range of feasible policies.

The Full Picture in the Analytical Model

Letting $\mu = \lambda$, and plugging the liability profile $\mathsf{FR}_0 = \mathsf{FR}_0(\lambda)$ and the risk / return profile $\sigma = \sigma(\lambda)$ into the formula for the Expected Funding Shortfall, we obtain:

$$\mathsf{EFS}_{\alpha,\tau}(\lambda) = 1 - \frac{1}{\alpha}\,\mathsf{FR}_0(\lambda)\exp\left\{\frac{1}{2}\,\left(\sigma(\lambda)\right)^2\tau\right\}\,\Phi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{\tau}\right)$$

Thus, the Expected Funding Shortfall simply becomes a function of the technical interest rate λ. This can be evaluated easily over a range of realistic values for λ.

The optimal technical interest rate λ^* is then simply the one that minimizes the Expected Funding Shortfall, i.e.

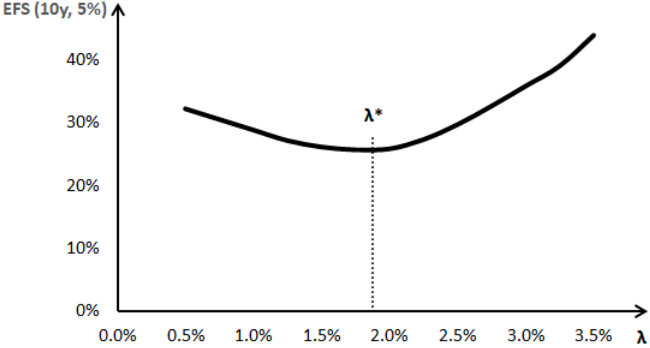
$$\lambda^* = \underset{\lambda\in[\mu_{\min},\,\mu_{\max}]}{\arg\min}\;\mathsf{EFS}_{\alpha,\tau}(\lambda)$$

In more complex settings, evaluations along the exact same logic can be done by using stochastic simulation.

The essential point is not the formula, but the logic with the liability profile (institutional side) and the risk / return profile (financial markets side).

Minimizing Long-term Financial Risk

For our example and plugging in a risk / return profile that reflects current conditions in the financial markets, we obtain the following result:



That is, it makes sense to fix a moderately elevated technical interest rate λ^* and to assume the associated short-term investment risk $\sigma(\lambda^*)$ in order to minimize long-term financial risk.

Conclusions

Since the cost of immunization is currently too high, Swiss pensions fund are heavily exposed to financial risk.

In the interest of their fiduciary duty, they should set themselves up in such a manner that long-term financial risk is minimized.

This is, however, not necessarily achieved by choosing the lowest possible technical interest rates and by assuming the lowest possible amounts of investment risk.

Fixing a higher technical interest rate and assuming more investment risk in the short term may, at least to some extent, reduce financial risk in the long term.

Good strategic risk management is, thus, not necessarily myopic risk avoidance, but rather considerate risk-taking with long-term financial stability in mind.

Bear in mind that the optimum always depends on the specific characteristic of the institutions under investigation. Do only what you thoroughly understand based on serious qualitative and quantitative analysis.

Structure of the course

Chapter I	Introduction
Chapter II	Preliminaries
Chapter III	Financing Liabilities
Chapter IV	The Asset / Liability Framework
Chapter V	The Lognormal Model
Chapter VI	ALM Study 1 - Dealing with the Risk / Return Profile
Chapter VII	ALM Study 2 - Incorporating Required Return
Chapter VIII	ALM Study 3 - Valuation and Risk Management
Chapter IX	Portfolio Construction and the Risk / Return Profile
Chapter X	Synopsis, Wrap-up and Conclusions

Bibliography I

[1] ASSMUSSEN, S., AND GLYNN, P.
Stochastic Simulation: Algorithms and Analysis.
Springer, Berlin, 2007.

[2] BLAKE, D.
Pension Economics.
John Wiley & Sons, Chichester, 2006.

[3] BLUM, P.
On some mathematical aspects of dynamic financial analysis.
Doctoral thesis, ETH, Zurich, 2005.

[4] BLUM, P.
Financial Risk Management in Social and Pension Insurance.
Lecture notes, ETH Zürich, Zürich, 2016.

[5] MCNEIL, A., FREY, R., AND EMBRECHTS, P.
Quantitative Risk Management, revised ed.
Princeton University Press, Princeton, 2015.

Bibliography II

[6] Scherer, B.
Liability Hedging and Portfolio Choice.
Risk Books, London, 2005.

[7] Swisscanto.
Schweizer Pensionskassenstudie 2020.
Survey, Swisscanto, Zürich, 2020.

Financial Risk Management in Social and Pension Insurance

Chapter II: Preliminaries

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 16, 2020

Table of contents

1. Cashflows and their properties

Time scale conventions

A discrete time scale will be used throughout the course.
Let t denote time in years: $t \in \{0,\, 1,\, \dots\,,\, T\} \subseteq \mathbb{N}_0$ with final time $T < \infty$.

Time $t = 0$ denotes the present, at which valuations usually take place; values $t > 0$ denote the future.

If t denotes a time interval, then this is to be interpreted as $(t-1, t]$. Usually, it suffices to identify t with the end of the respective year.

Note: All considerations in this course could easily be made also on quarterly, monthly or even continuous time scales. For the ease of presentation, however, only the yearly time scale will be used hereinafter.

Cashflows

Let $C_t \in \mathbb{R}$ denote some promised cashflow that occurs at time t, e.g.

- one installment of an old-age or disability pension,
- a reimbursement of a medical treatment,
- a reimbursement for long-term care,
- an unemployment benefit.

By convention, the cashflow C_t always takes place at the end of year t.

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ denote a stream of cashflows taking place over times $t \in \{1, \ldots, T\}$.

Special case: If $C_t = c$ for all t, then $\mathbf{C}$ is called an annuity.

Attention: There is no universal convention regarding signs. When only liabilities are considered, a positive sign may mean a cash outflow from the institution. In other instances, a positive sign may denote an inflow, whereas a negative sign may denote an outflow.

Mean Time to Payment

Definition 1 (Mean Time to Payment (MTP))

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ *denote a stream of cashflows with* $C_t \geq 0$ *for all* t*. The Mean Time to Payment is defined as:*

$$\mathit{MTP}(\mathbf{C}) = \sum_{t=1}^{T} \frac{t\,C_t}{\sum_{s=1}^{T} C_s}$$

MTP is a weighted time to payment where each time t is weighted by the relative contribution of the cashflow C_t to the total payment $\sum_{s=1}^{T} C_s$.

MTP is a useful summary statistic which is independent of any discount rates. It measures the average time horizon of the cashflow stream.

If $\mathbf{C}$ is an annuity, then we have $\text{MTP}(\mathbf{C}) = \frac{T+1}{2}$. This is due to the fact that $\sum_{t=1}^{T} t = \frac{T(T+1)}{2}$.

Time value of money

We stand at $t = 0$ and we must finance some cashflow C_t taking place at some future time $t > 0$. Hence, we must incorporate the time value of money:

Let there be an account in which we can deposit money and which grants an interest of δ, i.e. if we invest an amount M at $t = 0$, we have at $t = 1$:

$$M + \delta M = M(1 + \delta)$$

And over several time periods, due to the compounding of interest, we have

$$M(1 + \delta)(1 + \delta) \cdots (1 + \delta) = M(1 + \delta)^t$$

If there is some set amount C_t at time t, we can equate

$$M(1 + \delta)^t = C_t$$

Solving for M, we obtain

$$M = \frac{C_t}{(1 + \delta)^t}$$

M is the present value as of time $t = 0$ and with discount rate δ of the future cashflow C_t. I.e. if we can set aside M at time $t = 0$ and if the interest rate δ is guaranteed, then we end up with C_t at time t.

Note: This whole course is basically about

1. selecting a sensible value for the discount rate δ, and
2. making sure that δ will actually be earned.

Note: In life insurance mathematics, one usually considers the so-called discount factor $v = (1 + \delta)^{-1}$, and one considers this factor as deterministic and given; see e.g. [4].

Note: In the past, one used to assume that $\delta > 0$, and one used to make considerable efforts to build interest rate models that avoid negative rates, see e.g. [1]. Nowadays, we must relax this assumption and also allow for negative interest rates and discount rates.

Present Value of a stream of cashflows

Definition 2 (Present Value)

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ *be a stream of cashflows, and let* δ *be a discount rate. The Present Value of* $\mathbf{C}$ *as of time* 0 *is defined as*

$$PV_0(\mathbf{C}, \delta) = \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t}$$

That is, each single cashflow is discounted back to time $t = 0$, and these discounted values are summed up. Hence, if the discount rate δ is guaranteed, we can set aside the amount $\mathrm{PV}_0(\mathbf{C}, \delta)$ at time $t = 0$, and this suffices to finance all committed cashflows C_t of $\mathbf{C}$ with certainty.

This formula can easily be generalized to some general valuation date $s > 0$:

$$\mathrm{PV}_s(\mathbf{C}, \delta) = \sum_{t=s+1}^{T} \frac{C_t}{(1+\delta)^{t-s}}$$

For general cashflows streams, the present value must be computed numerically; see example spreadsheets. For regular cashflows, in particular for annuities, we can obtain explicit formulae:

Proposition 1 (Present value of an annuity)

Let $\mathbf{C}^T$ be a T-year annuity, i.e. $\mathbf{C}^T = (c, \ldots, c)' \in \mathbb{R}^T$, and let δ be the discount rate with $\delta \in (-1, \infty) \setminus \{0\}$. Then we have:

$$\mathit{PV}_0(\mathbf{C}^T, \delta) = \frac{c}{\delta}\left(1 - \frac{1}{(1+\delta)^T}\right)$$

Proof: Using the definition of the present value, we have:

$$\mathrm{PV}_0(\mathbf{C}, \delta) = c \sum_{t=1}^{T} \frac{1}{(1+\delta)^t} =: c \sum_{t=1}^{T} v^t \quad \text{for} \quad v = \frac{1}{1+\delta}$$

Here, the v^t form a geometric sequence $(a_t)_t$ with $a_1 = v$ and $a_{t+1}/a_t = v$.

Proof: (cont'd) Therefore, we have for the T-th partial sum, see [2]:

$$\sum_{t=1}^{T} v^t = v \frac{1 - v^T}{1 - v}$$

Re-inserting $v = \frac{1}{1+\delta}$ and rearranging, we obtain the claim. $\square$

This formula and proof also work for negative discount rates, provided that they are greater than -100% (which amounts to a full confiscation).

There are other such standardized cashflow streams in life insurance mathematics; see [4] for more details.

Internal Rate of Return

Definition 3 (Internal Rate of Return (IRR))

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ be a stream of cashflows, and assume that its present value is known to be PV_0^*. Assume that there exists some discount rate δ^* such that $PV_0(\mathbf{C}, \delta^*) = PV_0^*$. Then, δ^* is called the Internal Rate of Return, formally $IRR(\mathbf{C}, PV_0^*)$.

The IRR is simply the discount rate that equates the present value of some given cashflow stream to some given value:

$$\mathrm{PV}_0(\mathbf{C}, \mathrm{IRR}(\mathbf{C}, \mathrm{PV}_0^*)) = \mathrm{PV}_0^*$$

When it comes to bonds, the IRR is also called yield to maturity, see next section.

When some promised cashflow stream $\mathbf{C}$ is specified, and when a certain amount M of money is given to finance it, then $\mathrm{IRR}(\mathbf{C}, M)$ is the discount rate that must be guaranteed so that M suffices to pay $\mathbf{C}$.

In general, the IRR must be determined numerically; when cashflows have opposing signs, this can cause problems.

Dependence of present value on discount rate

For a stream of cashflows $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$, the present value is given by

$$\mathsf{PV}_0(\mathbf{C}, \delta) = \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t}$$

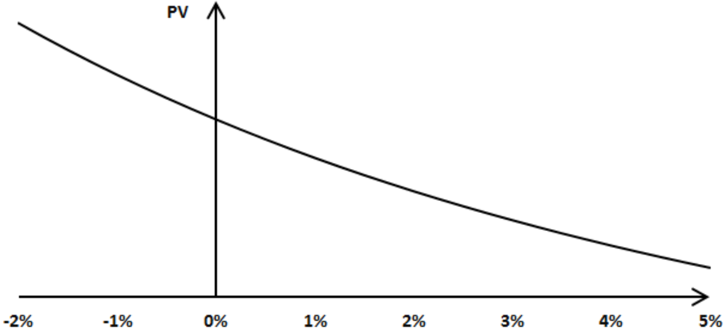
If $C_t > 0$ for all t, then PV_0 decreases as δ increases, and vice versa.

Example: Consider an annuity with $T = 20$ and $C_t = c = 1'000$:

Discount Rate	Present Value	Discount Rate	Present Value
5%	12'462	1%	18'046
4%	13'590	0%	20'000
3%	14'877	-1%	22'263
2%	16'351	-2%	24'894

If we are able to sustain a discount rate of 5%, we have to put up 12'462 at time zero to finance the annuity. If we are only able to sustain 2%, we have to put up 16'351, i.e. 31% more. These differences are dramatic. Therefore, it is crucially important to choose the discount rate δ carefully.

For a given cashflow stream $\mathbf{C}$, the relationship $\delta \mapsto \mathsf{PV}_0(\mathbf{C}, \delta)$ can be computed explicitly. For the annuity example above, this looks as follows:



As one would expect from the formula, the relationship is non-linear.

In order to express the relationship between δ and $\mathsf{PV}_0(\mathbf{C}, \delta)$ more concisely, there exist the notions of duration and convexity.

Duration of a stream of cashflows

Definition 4 (Duration)

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ *be a stream of cashflows with* $C_t > 0$ *for all* t*.*
The Duration of $\mathbf{C}$ *is then defined as*

$$D(\mathbf{C}, \delta) = -\frac{\partial}{\partial \delta} PV_0(\mathbf{C}, \delta) \Big/ PV_0(\mathbf{C}, \delta)$$

The duration is the relative change in value of $\mathbf{C}$ in response to an infinitesimal change in the discount rate δ.

In the finance literature, this form of duration is called Modified Duration. If the C_t do not depend on δ, it also equals the Effective Duration. It does, however, differ slightly from the often-cited Macaulay Duration which equals $(1 + \delta) D(\mathbf{C}, \delta)$.

If useful, we can use the relationship $\mathsf{D}(\mathbf{C}, \delta) = -\frac{\partial}{\partial \delta} \log \mathsf{PV}_0(\mathbf{C}, \delta)$.

The term $-\frac{\partial}{\partial \delta} \mathsf{PV}_0(\mathbf{C}, \delta)$ is the absolute change in value and often referred to as the Dollar (or whatever currency you like) Duration.

Proposition 2 (Duration of a stream of cashflows)

Under the assumptions of Definition 4, we have

$$D(\mathbf{C}, \delta) = \frac{1}{1+\delta} \sum_{t=1}^{T} \frac{t\, C_t}{(1+\delta)^t} \Bigg/ \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t}$$

Proof: Take first derivatives and rearrange terms. $\square$

$\mathsf{D}(\mathbf{C}, \delta)$ can thus be interpreted as a specially weighted mean time to payment, hence the name duration. One should, however, be careful with this interpretation.

If $C_t > 0$ tor all t, then $\mathsf{D}(\mathbf{C}, \delta) > 0$ and $\frac{\partial}{\partial \delta} \mathsf{PV}_0(\mathbf{C}, \delta) < 0$. I.e. PV_0 decreases whenever δ increases, and vice versa.

Attention: Care must be taken with the sign whenever using the duration.

Convexity of a stream of cashflows

Definition 5 (Convexity)

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ be a stream of cashflows with $C_t > 0$ for all t. The Convexity of $\mathbf{C}$ is then defined as

$$K(\mathbf{C}, \delta) = \frac{\partial^2}{\partial \delta^2} PV_0(\mathbf{C}, \delta) \Big/ PV_0(\mathbf{C}, \delta)$$

This is just the second order term, the change in the change of value.

Proposition 3 (Convexity of a stream of cashflows)

Under the assumptions of Definition 5, we have

$$K(\mathbf{C}, \delta) = \frac{1}{(1+\delta)^2} \sum_{t=1}^{T} \frac{t\,(t+1)\,C_t}{(1+\delta)^t} \Bigg/ \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t}$$

Proof: Take second derivatives and rearrange terms. □

If the cashflows C_t are positive and and fixed, the convexity is always positive.

Approximation formula

We can now use duration and convexity in the classical manner to make first or second order Taylor approximations for the change of $\mathsf{PV}_0(\mathbf{C}, \delta)$. Let δ_0 denote the starting point, and let $\Delta\delta$ denote the change in discount rate. Then:

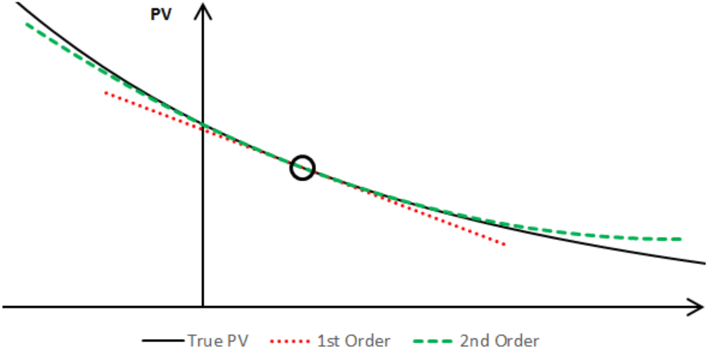
First order:

$$\frac{\mathsf{PV}_0(\mathbf{C}, \delta_0 + \Delta\delta) - \mathsf{PV}_0(\mathbf{C}, \delta_0)}{\mathsf{PV}_0(\mathbf{C}, \delta_0)} = -\mathsf{D}(\mathbf{C}, \delta_0)\Delta\delta + o((\Delta\delta)^2)$$

Second order:

$$\frac{\mathsf{PV}_0(\mathbf{C}, \delta_0 + \Delta\delta) - \mathsf{PV}_0(\mathbf{C}, \delta_0)}{\mathsf{PV}_0(\mathbf{C}, \delta_0)} = -\mathsf{D}(\mathbf{C}, \delta_0)\Delta\delta + \tfrac{1}{2}\,\mathsf{K}(\mathbf{C}, \delta_0)(\Delta\delta)^2 + o((\Delta\delta)^3)$$

How good is this approximation in practice?



In the presence of significant non-linearity, the first-order approximation does rather poorly for larger changes of the discount rate.

Call from practice

Actual cashflow streams from social and pension insurance often exhibit significant convexity. Hence, using duration alone as a measure for sensitivity against changes of interest rate / discount rate may be misleading and dangerous. Therefore, in practice:

1. Use explicit calculation of true present value if possible.
2. Or use at least the second order approximation.

Particularly, if larger changes of discount rates have to be dealt with.

2. Bonds and yield curves

Bond basics

Bonds are the most basic and most frequent type of security in which capital-based social and pension insurance institutions invest the money that they hold in order to cover their liabilities. For a comprehensive reference, see [3].

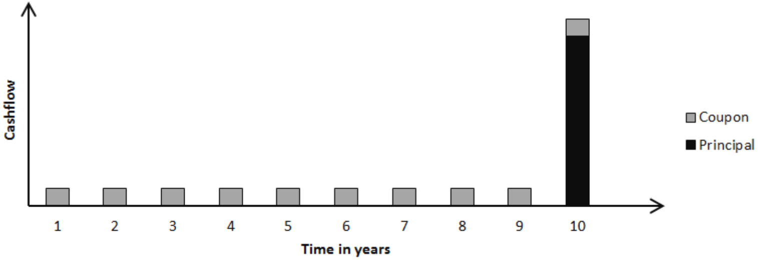
A (bullet) bond is a security by which an issuer (e.g. the Swiss Confederation) takes on some amount of money P, called the Principal and promises to pay a fixed Coupon c each year for a specified number of years N as well as to pay back the principal at the time of Maturity, i.e. when the N years have expired.

Basically, the owner of the bond, e.g. a pension fund, extends a loan of P to the issuer, e.g. the Swiss Confederation, and is indemnified for doing so by receiving the regular coupon payments c. The loan is to be paid back in full at maturity.

The lifetime N of a bond is fixed at issuance. We are, however, more interested in its residual lifetime at the time when we do the valuation. i.e. usually at $t = 0$. We call this the Time to Maturity M.

- A 10-year bond ($N = 10$) issued 5 years ago (at $t = -5$) is now (at $t = 0$) a bond with 5 years time to maturity ($M = 5$).
- The same holds for a 30-year bond issued 25 years ago.

Hence, a bond is basically a standardized stream of cashflows:



In the terminology introduced before, we can, therefore, express the bond as follows:

$$\mathbf{C} = (C_1 = c,\ \ldots\ ,\ C_{M-1} = c,\ C_M = c + P)' \in \mathbb{R}^M \tag{1}$$

Thus, we can apply all the cashflow concepts introduced above also to bonds.

Outlook: One intuitive idea would be to match the cashflows of some bonds with the promised cashflows from the insurance side in order to obtain an immunized position. This will be looked at in the following chapter.

For the time being, we make the following simplifying assumptions:

- There is no credit risk, i.e. we can be reasonably sure that all the payments will actually be made.
- There is only one coupon payment per year.
- All payments take place at the end of the respective year.

Convention: We will always quote the principal P as 100. Thus, everything else corresponds to a percentage of the principal. This corresponds to standard practice in the bond markets. If we need a position of more than 100, we simply use several units of the bond.

Market price and Yield to maturity

A bond is a standardized security that can, in principle, be bought and sold in an organized market at any time during its lifespan.

Let B_0 denote the market price of a bond with cashflows $\mathbf{C}$ as in Formula 1 prevailing at the valuation time $t = 0$. We can now determine the discount rate δ^* that equates the present value of the bond to its market price:

$$\mathsf{PV}_0(\mathbf{C}, \delta^*) \stackrel{!}{=} B_0 \quad \Leftrightarrow \quad \delta^* = \mathsf{IRR}(\mathbf{C}, B_0)$$

In the bond world, this internal rate of return δ^* is called Yield to Maturity and abbreviated by $R(0, M)$.

The yield to maturity is, in principle, the annual rate of return that can be achieved if one buys the bond at price B_0 at $t = 0$ and holds it up until maturity at time $t = M$.

This is based on the assumption that the coupon payments c can be reinvested at the interest rate $R(0, M)$ between time t when they are made and maturity M. This is not always realistic, as interest rates change over time.

The traded price B_0 of a bond is not necessarily equal to the par value P, and the yield to maturity $R(0, M)$ is not necessarily equal to the coupon rate c/P. Both B_0 and $R(0, M)$ depend on supply and demand as they prevail in the market, e.g.:

Country	Ticker	Maturity	Coupon	Market Price	Yield to maturity
Switzerland	SWISS 4 04/08/28	10 years	4.00	139.27	-0.03%
Italy	BTPS 2 02/01/28	10 years	2.00	95.28	2.58%
Australia	ACGB 2 1/4 05/21/28	10 years	2.25	96.65	2.64%
Portugal	PGB 2 1/8 10/17/28	10 years	2.125	103.19	1.78%
Italy	BTPS 0.95 03/01/23	5 years	0.95	96.78	1.68%
Switzerland	SWISS 4 02/11/23	5 years	2.00	114.75	-0.87%

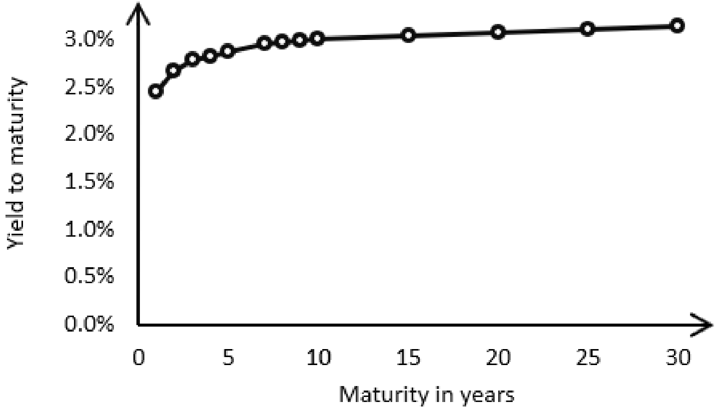
For a bond with $P = 100$, the following nomenclature is common:

- A bond with a market price $B_0 < 100$ is said to trade at a discount.
- A bond with a market price $B_0 > 100$ is said to trade at a premium.
- A bond with a market price $B_0 \approx 100$ is said to trade at par.

The yield curve

In the market, there are usually comparable bonds for a range of different maturities $M_1 < M_2 < \ldots < M_n$, each one with its traded price $B_0(M_i)$ and its yield to maturity $R(0, M_i)$. The graph $\{(M_i, R(0, M_i)) : i = 1, \ldots, n\}$ is called the Yield Curve.

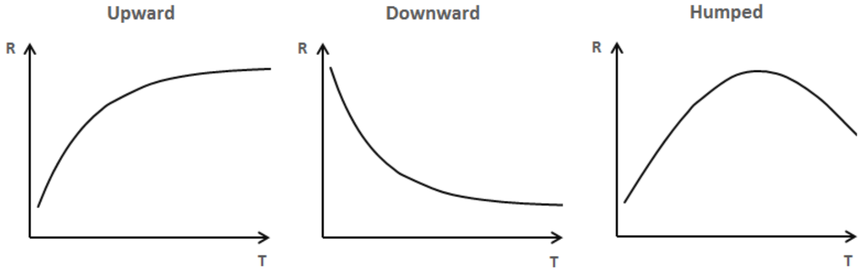
Example: Yield curve for bonds of the US Treasury as of August 2018:



Yield curve shapes

Normally, the yield curve is upward-sloping: The longer you tie up your money, the higher the yield to maturity that you get as a reward.

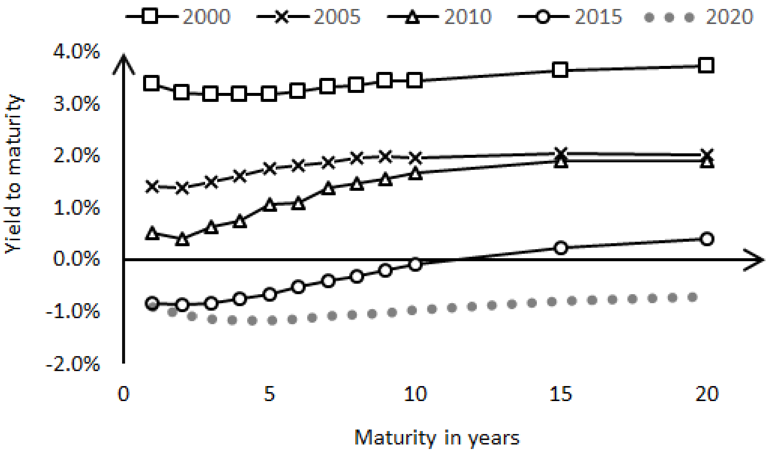
However, the yield curve may, at times, also be downward-sloping ("inverted"), or it may be humped, or it may have even fancier shapes:



For sloped yield curves, the steepness of the slope is of considerable interest.

Yield curve history

Over the past decades, yields have decreased dramatically over the entire range of maturities; here e.g. the Swiss case:



This is a major problem for capital-based social and pension insurance institutions, as we will see in the next chapters.

Zero-coupon bonds

Coupon bonds as introduced are rather cumbersome due to the intermediate coupon cashflows. A simpler alternative:

A Zero-Coupon Bond (ZCB) with maturity M pays the amount of 1 at maturity M, and no coupons in between $t = 0$ and $t = M$. Let $P(0, M)$ be the market price of a ZCB at time $t = 0$. Applying the usual valuation method, we then must have for some discount rate δ^*:

$$\mathsf{PV}_0\left(\mathbf{Z}^M, \delta^*\right) = P(0, M) = \frac{1}{(1+\delta^*)^M} \quad \text{or} \quad \delta^* = \frac{1}{P(0, M)^{1/M}} - 1$$

The resulting discount rate $\delta^* =: Z(0, M)$ is called Zero-Coupon Yield or also Spot Rate. It has no underlying issues with reinvestment.

Convention: Zero coupon bonds are always quoted with a final value of 1.

Therefore, the price of a zero coupon bond can be directly used as an alternative discount factor for a cashflow C_T occurring at time T:

$$\widetilde{\mathsf{PV}}_0(C_T, Z(0, T)) := P(0, T)\, C_T$$

Based on this, we can express a coupon bond as a portfolio of zero-coupon bonds:

- For each time $t \in \{1, \ldots, M\}$, we hold c units of a ZCB with maturity t.
- For $t = M$, we hold an additional P units of a ZCB with maturity M.

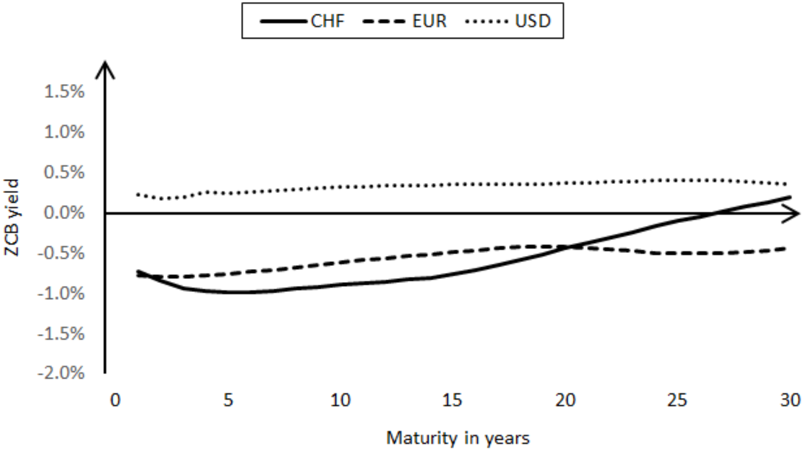
This ZCB portfolio produces exactly the same pattern of cashflows as a coupon bond. Therefore, if the market is arbitrage-free, the traded price B_0 of the coupon bond must be equal to the price of the ZCB portfolio:

$$B_0 \stackrel{!}{=} \sum_{t=1}^{M} c\, P(0,t) + P(0,M)P \tag{2}$$

In most markets, zero coupon bonds do not actually exist. Formula 2, applied to a number of different maturities can, however, be used to compute artificial equivalent ZCB rates from the prices of actually traded coupon bonds by means of a procedure called bootstrapping.

We will not treat this further here. ZCB curves are readily available from different sources, e.g. from Bloomberg or also from supervisory authorities for statutory purposes.

Example: Zero-coupon rate curves computed by FINMA for use in the Swiss Solvency Test 2020:



Valuation based on ZCB curves

The valuation logic based on ZCB prices and rates as shown above can also be applied to general cashflow streams, i.e. instead of computing

$$\mathsf{PV}_0(\mathbf{C}, \delta) = \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t}$$

for a single discount rate δ, one can also compute

$$\widetilde{\mathsf{PV}}_0(\mathbf{C}) = \sum_{t=1}^{T} C_t \, P(0,t) = \sum_{t=1}^{T} \frac{C_t}{(1+Z(0,t))^t}$$

This amounts to applying a specific discount rate for each time t. If we start with the latter approach, we can link this to the former one by letting

$$\delta^* = \mathsf{IRR}(\mathbf{C}, \widetilde{\mathsf{PV}}_0) \quad \text{such that} \quad \mathsf{PV}_0(\mathbf{C}, \delta^*) = \widetilde{\mathsf{PV}}_0(\mathbf{C})$$

δ^* is then the equivalent single discount rate.

Duration and convexity of bonds

Since bonds are simply streams of cashflows, we can apply the concepts of duration and convexity directly to them.

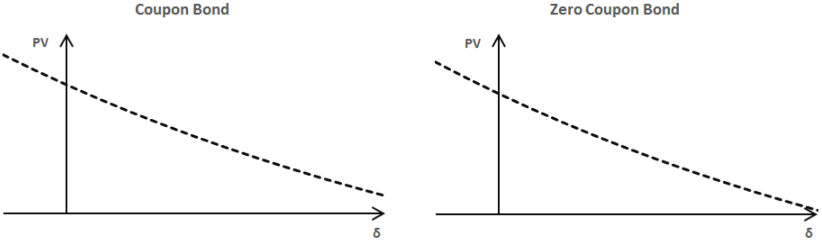
For a coupon bond, we have $\mathbf{C} = (c, \ldots, c, c+P)' \in \mathbb{R}^M$; letting $\delta = R(0, M)$ and using Propositions 2 and 3, we obtain:

$$\mathsf{D}(\mathbf{C}, \delta) = \frac{1}{1+\delta} \left[\sum_{t=1}^{M} \frac{c\,t}{(1+\delta)^t} + \frac{MP}{(1+\delta)^M} \right] \Bigg/ \left[\sum_{t=1}^{M} \frac{c}{(1+\delta)^t} + \frac{P}{(1+\delta)^M} \right]$$

$$\mathsf{K}(\mathbf{C}, \delta) = \frac{1}{(1+\delta)^2} \left[\sum_{t=1}^{M} \frac{c(t+1)t}{(1+\delta)^t} + \frac{M(M+1)P}{(1+\delta)^M} \right] \Bigg/ \left[\sum_{t=1}^{M} \frac{c}{(1+\delta)^t} + \frac{P}{(1+\delta)^M} \right]$$

For a zero coupon bond, the cashflow stream is $\mathbf{C} = (0, \ldots, 0, 1)' \in \mathbb{R}^M$, and, letting $\delta = Z(0, T)$, the formulae simplify to:

$$\mathsf{D}(\mathbf{C}, \delta) = \frac{M}{1+\delta} \approx M \qquad \text{and} \qquad \mathsf{K}(\mathbf{C}, \delta) = \frac{M\,(M+1)}{(1+\delta)^2}$$



Also with bonds, convexity is an issue, and second-order approximation should be preferred over first-order approximation.

Duration and convexity of a portfolio of bonds

Proposition 4 (Portfolio duration and convexity)

Assume that we hold n_1 units of a bond $\mathbf{C}_1$ valued at PV_0^1 per unit, and n_2 units of a bond $\mathbf{C}_2$ priced at PV_0^2 per unit. Then we have for the combined position:

$$D(n_1\mathbf{C}_1 + n_2\mathbf{C}_2, \delta) = w_1 D(\mathbf{C}_1, \delta) + w_2 D(\mathbf{C}_2, \delta)$$
$$K(n_1\mathbf{C}_1 + n_2\mathbf{C}_2, \delta) = w_1 K(\mathbf{C}_1, \delta) + w_2 K(\mathbf{C}_2, \delta)$$
$$\textit{where} \quad w_i = \frac{n_i PV_0^i}{n_1 PV_0^1 + n_2 PV_0^2} \quad \textit{for} \quad i \in \{1,\, 2\}.$$

Proof: Use $\frac{\partial}{\partial\delta}\left[n_1 \mathrm{PV}_0^1 + n_2 \mathrm{PV}_0^2\right] = n_1 \frac{\partial}{\partial\delta}\mathrm{PV}_0^1 + n_2 \frac{\partial}{\partial\delta}\mathrm{PV}_0^2$, then divide by the portfolio value $n_1 \mathrm{PV}_0^1 + n_2 \mathrm{PV}_0^2$ and rearrange.

Same procedure with second derivatives for convexity. □

This naturally generalizes to portfolios with n bonds. The relation only holds exactly if all bonds have the same yield δ. Otherwise it is a (usually good) approximation.

Bibliography I

[1] Blum, P.
On some mathematical aspects of dynamic financial analysis.
Doctoral thesis, ETH, Zurich, 2005.

[2] Bronstein, I., Semendjajew, K., Musiol, G., and Mühlig, H.
Taschenbuch der Mathematik.
Verlag Harri Deutsch, Thun, 1995.

[3] Fabozzi, F.
Fixed Income Analysis, second ed.
John Wiley & Sons, Chichester, 2007.

[4] Gerber, H.
Life Insurance Mathematics, third ed.
Springer, Berlin, 1997.

Financial Risk Management in Social and Pension Insurance

Chapter III: Financing Liabilities

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 16, 2020

Table of Contents

1. Introduction and overwiew

The task

Let $t \in \{0, 1, \ldots, T\}$ for $T < \infty$ denote time in years.

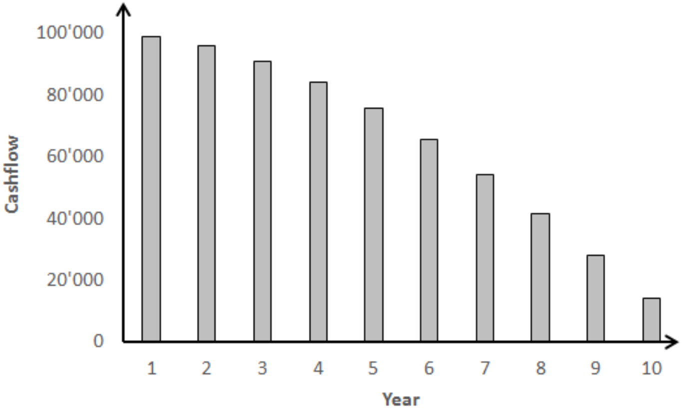
Assume that we have a stream $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ of promised cashflows with $C_t \geq 0$ for all t. This can be e.g. a portfolio of old-age or disability pensions. Assume, for the time being, that these cashflows are deterministic and known.

We stand at time $t = 0$ and must put up a portfolio of assets A_0 such that the payment of all cashflows C_t for all times $t \in \{1, \ldots, T\}$ is assured (at least with high probability).

How can this be achieved?

Example: cashflow stream

A typical cashflow stream in social and pensions insurance (albeit with an unusually short time horizon, for simplicity's sake):



Naïve approach: bank account

Assume that there is a bank account which pays an annual interest rate δ. Then we can put up the amount

$$A_0 = \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t} \quad (= \mathsf{PV}_0(\mathbf{C}, \delta))$$

and all future payments are assured; c.f. Section 1 of Chapter II.

Does this work? If the earning of the interest rate δ is guaranteed for the entire time span $t = 0,\, 1,\, \ldots,\, T$: YES.

Is the interest rate guaranteed? NO! It can actually vary quite considerably over time. If this is the case, payment of all liabilities is no longer assured.

- in the 1980-ies: $\delta \approx 4\%$
- around 2000: $\delta \approx 2\%$
- now: $\delta \approx 0\%$

This does not work. FORGET IT!

Alternative: bonds

Recall from Section 2 of Chapter II that a bond is basically a standardized stream of cashflows:

$$\mathbf{B} = (c, \ldots, c, c + P)' \in \mathbb{R}^M$$

where c is the (annual) coupon, P is the principal and M is the number of years to maturity, i.e. the redemption of the principal. We shall follow the convention of Chapter II and always let $P = 100$.

By paying the market price B_0 prevailing at $t = 0$, we can acquire the bond, i.e. the right to receive exactly this stream of cashflows.

- Important: After the purchase, there is no more uncertainty about the cash-flows ...
- ... under the tacit assumption that the issuer of the bond does not go bankrupt until the maturity of the bond.

Ansatz: At time $t = 0$, we buy a portfolio of different bonds, composed in such a way that their cashflows are equal to the liability cashflows to be financed.

2. Cashflow matching and immunized value

Bond portfolio

Assume that we have a bond market with N different bonds $\mathbf{B}^i$, $i \in \{1, \ldots, N\}$. Each bond has its specific coupon c^i and its specific maturity M^i. We assume that $M^i \leq T$ for all i. Following the usual convention, we also assume that the principal P^i of each bond equals 100. That is, we have a number of standardized streams of cashflows

$$\mathbf{B}^i = \left(c^i, \ldots c^i, 100 + c^i\right)' \in \mathbb{R}^{M^i}$$

We assume that all these bonds are (reasonably) free of credit risk, which is the case e.g. for bonds of the Swiss confederation.

We stand at $t = 0$ as usual. All bonds are assumed to be traded in the market, with current market prices B_0^i for $i \in \{1, \ldots, N\}$. This also means that each bond has its yield to maturity

$$R(0, M^i) = \mathsf{IRR}\left(\mathbf{B}^i, B_0^i\right)$$

i.e. the discount rate that equates the present value of the cashflow pattern to the price paid in the market.

Cashflow matrix

Assume w.l.o.g. that $1 = M^1 \leq M^2 \leq \cdots \leq M^N = T$. Then, we can express the cashflows in matrix form: $B \in \mathbb{R}^{T \times N}$ such that $B_{t,i}$ expresses the cashflow of bond i at time t:

$$B = \begin{pmatrix} 100 + c^1 & c^2 & \cdots & c^i & \cdots & c^N \\ 0 & 100 + c^2 & & \vdots & & \vdots \\ \vdots & 0 & & \vdots & & \\ \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & c^i & & \vdots \\ \vdots & \vdots & & 100 + c^i & & \vdots \\ \vdots & \vdots & & 0 & & \vdots \\ \vdots & \vdots & & \vdots & & c^N \\ 0 & 0 & \cdots & 0 & \cdots & 100 + c^N \end{pmatrix}$$

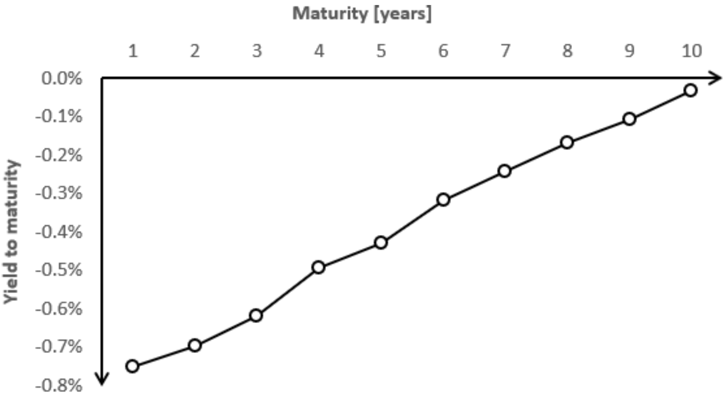
Example: bond market data

Market data from the Swiss bond market as of Summer 2018:

Maturity [years]	Coupon [CHF]	Principal [CHF]	Market Price [CHF]	Yield to maturity [CHF]
1	3.00	100	103.78	-0.75%
2	2.25	100	105.95	-0.70%
3	2.00	100	107.95	-0.62%
4	2.00	100	110.10	-0.49%
5	4.00	100	122.44	-0.43%
6	1.25	100	109.51	-0.32%
7	1.50	100	112.31	-0.24%
8	1.25	100	111.44	-0.17%
9	3.25	100	130.37	-0.11%
10	4.00	100	140.41	-0.03%

Example: yield curve

Yield curve resulting from the bond market data:



This does not look like in the textbooks. But it is the reality that Swiss institutional investors have to cope with nowadays.

Example: cashflow matrix

Cashflow matrix resulting from the bonds indicated above:

Calendar year t	Maturity of Bond i: 1	2	3	4	5	6	7	8	9	10
1	103.00	2.25	2.00	2.00	4.00	1.25	1.50	1.25	3.25	4.00
2	0.00	102.25	2.00	2.00	4.00	1.25	1.50	1.25	3.25	4.00
3	0.00	0.00	102.00	2.00	4.00	1.25	1.50	1.25	3.25	4.00
4	0.00	0.00	0.00	102.00	4.00	1.25	1.50	1.25	3.25	4.00
5	0.00	0.00	0.00	0.00	104.00	1.25	1.50	1.25	3.25	4.00
6	0.00	0.00	0.00	0.00	0.00	101.25	1.50	1.25	3.25	4.00
7	0.00	0.00	0.00	0.00	0.00	0.00	101.50	1.25	3.25	4.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	101.25	3.25	4.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	103.25	4.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	104.00

Immunization: idea

At time $t = 0$, from each bond $\mathbf{B}^i$, we can buy n_i units at the market price B_0^i. This results in a portfolio

$$\mathbf{n} := (n_1, \ldots, n_N)' \in \mathbb{R}^N$$

Given the cashflow matrix B, this portfolio will produce a cashflow stream $B\mathbf{n} \in \mathbb{R}^T$ over time. If we manage to fix $\mathbf{n}$ in such a way that

$$B\mathbf{n} = \mathbf{C} \qquad \text{or} \qquad (B\mathbf{n})_t = C_t \quad \text{for all} \quad t \in \{1, \ldots, T\}$$

where C_t is the given liability cashflow at time t, then we have achieved the state of immunization, i.e.

- We buy the portfolio $\mathbf{n}$ at time $t = 0$ for the price $\sum_{i=1}^{N} n_i B_0^i$.
- At each time $t \in \{1, \ldots, T\}$, this portfolio gives us an income of $(B\mathbf{n})_t$.
- And this income is then used (and sufficient) to pay the cashflow C_t owed.

That is, by purchasing the portfolio $\mathbf{n}$ at time $t = 0$, we have covered all our obligations for $t \in \{1, \ldots, T\}$.

Simple immunization

Let $\mathbf{C} = (C_1, \dots, C_T)' \in \mathbb{R}^T$ denote the liability cashflow stream.

Assume that we have exactly $N = T$ different bonds $\mathbf{B}^i$, $i \in \{1, \dots, N = T\}$, and each of these bonds matures in a different year. That is, for each single year $t \in \{1, \dots, T\}$, we we have one bond with $i = t$ and cashflow pattern

$$\mathbf{B}^t = \left(c^t, \dots, c^t, 100 + c^t\right)' \in \mathbb{R}^t$$

This yields a cashflow matrix that is quadratic, i.e. $B \in \mathbb{R}^{T \times T}$ and that has upper triangular form. If the coupon rates are non-degenerate, then B also has full rank. With the portfolio vector $\mathbf{n} = (n_1, \dots, n_T)' \in \mathbb{R}^T$, we obtain the following condition for immunization:

$$B\mathbf{n} = \mathbf{C}$$

and this has the (under these circumstances unique) solution

$$\mathbf{n} = B^{-1}\mathbf{C}$$

The problem with this simple solution is that certain elements of this solution may be negative, i.e. $n_i < 0$ for some $i \in \{1, \dots, T\}$.

This amounts to a short position in bond $\mathbf{B}^i$. Such a short position can, in principle, be implemented. But in the context of social and pension insurance, short positions are often forbidden by law or regulation. And even if they are not forbidden, they are not desirable for reasons related to operations and risk management.

Anyway, in realistic settings with longer time horizons than 10 years, it is usually difficult or impossible to set up a cashflow matrix that is exactly upper triangular.

Therefore, we need a more general formulation of the immunization problem. It must accommodate the constraint that $\mathbf{n} \geq 0$ (to be understood as $n_i \geq 0$ for all components i), and it should also be able to deal with more general cashflow matrices where $N \neq T$.

Note: A setting where each (reasonably measurable) cashflow can be perfectly replicated by securities available in the market is called a complete market.

General immunization problem

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ with $C_t \geq 0$ for all $t \in \{1, \ldots, T\}$ be a given stream of liability cashflows.

Assume that we have N bonds $\mathbf{B}^i$ with coupons c^i and with maturities such that $1 = M^1 \leq \cdots \leq M^N = T$, each one with its market price B_0^i and its yield to maturity $R^i(0, M^i) = \mathsf{IRR}(\mathbf{B}^i, B_0^i)$ at time $t = 0$.

The resulting cashflow matrix is $B = (B_{t,i})_{t \in \{1, \ldots, T\}, i \in \{1, \ldots, N\}}$, where $B_{t,i}$ denotes the cashflow from bond i at time t.

Let $\mathbf{n} = (n_1, \ldots, n_N)' \in \mathbb{R}^N$ denote the portfolio, i.e. n_i is the number of units of bond i that we purchase at time $t = 0$. We always impose $\mathbf{n} \geq 0$.

If we hold some bond portfolio $\mathbf{n}$, then for each time $t \in \{1, \ldots, T\}$, we have a total cash inflow $(B\mathbf{n})_t$ and a cash outflow of C_t from the liabilities, resulting in a discrepancy $(\mathbf{C} - B\mathbf{n})_t$. In the simplest case, we would simply minimize the sum of these discrepancies, i.e.

$$\min_{\mathbf{n}} \mathbf{1}' (\mathbf{C} - B\mathbf{n}) \quad \text{s.t.} \quad \mathbf{n} \geq 0$$

This is a simple linear programming problem as proposed e.g. in [1].

The solution may, however, not be satisfactory. We might have the situation where a large positive discrepancy at some time t_1 may be compensated by a large negative discrepancy at some other time t_2. This is contrary to our intention of matching cashflows from the bond holdings with cashflows for the liabilities as best as we can. To the latter end, we should rather optimize absolute deviations, i.e.

$$\min_{\mathbf{n}} \sum_{t=1}^{T} \left((B\mathbf{n})_t - C_t\right)^2 \quad \text{s.t.} \quad \mathbf{n} \geq 0 \tag{1}$$

The square makes sure that larger discrepancies (in whatever direction) are more penalized than smaller ones. Using matrix notation, this problem is equivalent to

$$\min_{\mathbf{n}} \tfrac{1}{2}\, \mathbf{n}'(B'B)\mathbf{n} - (\mathbf{C}'B)\mathbf{n} \quad \text{s.t.} \quad \mathbf{n} \geq 0 \tag{2}$$

This is a standard quadratic programming problem that can be solved by any standard software package. If necessary, we can also impose further equality or inequality constraints.

Proof: (Equivalence of equations 1 and 2)

We have $(B\mathbf{n})_t = \sum_{i=1}^{N} B_{t,i}\, n_i$. Therefore:

$$((B\mathbf{n})_t - C_t)^2 = \left(\sum_{i=1}^{N} B_{t,i}\, n_i - C_t\right)^2 = \left(\sum_{i=1}^{N} B_{t,i}\, n_i\right)^2 - 2\, C_t \sum_{i=1}^{N} B_{t,i}\, n_i + C_t^2$$

The term C_t^2 does not depend on $\mathbf{n}$ and needs not be considered any further. For the second term, we can sum over t to obtain:

$$\begin{aligned}\sum_{t=1}^{T} C_t \sum_{i=1}^{N} B_{t,i}\, n_i &= \sum_{t=1}^{T} \sum_{i=1}^{N} C_t\, B_{t,i}\, n_i \\ &= \sum_{i=1}^{N} \left(\sum_{t=1}^{T} C_t\, B_{t,i}\right) n_i = (\mathbf{C}' B)\, \mathbf{n}\end{aligned}$$

For the first term, we have

$$\left(\sum_{i=1}^{N} B_{t,i}\, n_i\right)^2 = ((B\mathbf{n})_t)^2 = \left(\mathbf{n}' B'\right)_t (B\mathbf{n})_t$$

and therefore

$$\sum_{t=1}^{T} \left(\mathbf{n}' B'\right)_t (B\mathbf{n})_t = \mathbf{n}'(B'B)\mathbf{n}$$

Putting everything together and multiplying by $\frac{1}{2}$, we obtain

$$\tfrac{1}{2}\mathbf{n}'(B'B)\mathbf{n} - (\mathbf{C}' B)\, \mathbf{n} \qquad \square$$

Immunizing portfolio

Let $\mathbf{n}^*$ denote the solution of Optimization Problem 2, i.e. the bond portfolio that best matches the given liability cashflows $\mathbf{C}$. What does this mean?

- At time $t = 0$, we buy n_i units of Bond $\mathbf{B}^i$ at market price B_0^i for each $i \in \{1, \ldots, N\}$. This results in a total purchase price of $\sum_{i=1}^N n_i^* B_0^i$.
- We then hold this portfolio unaltered over time, without buying or selling. Thus, at each time $t \in \{1, \ldots, T\}$ we receive a cashflow of $(B\mathbf{n}^*)_t$.
- We then use this cashflow in order to pay the liability cashflow C_t.

Generally, there will be no exact match between $(B\mathbf{n}^*)_t$ and C_t. But if the discrepancies are small relative to the value of the bond portfolio, i.e.

$$\sqrt{\sum_{t=1}^T \left((B\mathbf{n}^*)_t - C_t\right)^2} \ll \sum_{i=1}^N n_i^* B_0^i$$

this is a viable approximation. We simply need to hold a relatively small amount of additional cash D^* to make sure that the discrepancies will be covered.

Example: results

Composition of portfolio, purchase price and residual:

Bond / Maturity i	Number of units n_i	Price per unit B_0^i	Yield to maturity $R(0,M)$
1	851	103.78	-0.75%
2	847	105.95	-0.70%
3	816	107.95	-0.62%
4	764	110.10	-0.49%
5	693	122.44	-0.43%
6	620	109.51	-0.32%
7	514	112.31	-0.24%
8	396	111.44	-0.17%
9	267	130.37	-0.11%
10	136	140.41	-0.03%

Total purchase price of immunizing portfolio:	**658'718**
Residual discrepancy:	**0.02**
Equivalent immunizing yield:	**-0.35%**

Immunized value

By following the procedure above, for some given stream of liability cashflows, we have the following situation:

- By buying the bond portfolio $\mathbf{n}^*$ at time $t = 0$ for the price $\sum_{i=1}^{N} n_i^* B_0^i$, and by putting up additional cash D^*,
- we can make sure that all subsequent liability cashflows at all future times $t \in \{1, \ldots, T\}$ can be paid with certainty.

This situation is completely independent of the subsequent development of the bond prices. By paying B_0^i at $t = 0$, we have assured the receipt of the cashflows $(c^i, \ldots, c^i, 100 + c^i)$, irrespective of what the price for doing this would be in the future.

That is, we have effectively created an immunized position, and the price for doing so is the immunized value of the cashflow stream $\mathbf{C}$:

$$\mathsf{IV}_0(\mathbf{C}) := \sum_{i=1}^{N} n_i^* B_0^i + D^* \approx \sum_{i=1}^{N} n_i^* B_0^i$$

If we want to remain in the present value framework according to Chapter II, we can also define the equivalent immunized discount rate δ_{IM} as

$$\delta_{\mathsf{IM}} := \mathsf{IRR}\left(\mathbf{C}, \mathsf{IV}_0(\mathbf{C})\right)$$

such that the usual present value becomes equal to the immunized value, i.e.

$$\mathsf{PV}_0(\mathbf{C}, \delta_{\mathsf{IM}}) = \sum_{t=1}^{T} \frac{C_t}{(1+\delta_{\mathsf{IM}})^t} = \mathsf{IV}_0(\mathbf{C})$$

The value $\mathsf{IV}_0(\mathbf{C})$ is a viable valuation of the cashflow stream $\mathbf{C}$ because it is based on a discount rate that we can actually earn. It has a special role with respect to other valuations based on other discount rates, because no financial risk is involved.

Example: immunized value

For our example, we obtain the following results:

- Immunized value: $\mathsf{IV}_0(\mathbf{C}) = 658'718 + 0.02 = 658'718$
- Equivalent immunized discount rate: $\delta_{\mathsf{IM}} = -0.35\%$
- Undiscounted sum of liability cashflows: $648'841$

Since the equivalent immunizing discount rate is negative, the purchase price of the immunizing portfolio is higher than the undiscounted sum of cashflows.

That is, in order to achieve an immunized position under the current market conditions, we must put up more money at the beginning than we will pay out over the course of time. We have managed to eliminate financial risk, but the price for this is a negative contribution from financial returns.

Immunization with zero-coupon bonds

Another drawback of the immunization approach as presented is that the handling of the coupon bonds is relatively cumbersome.

The situation would be less complicated if we had zero-coupon bonds. Recall from section 2 of Chapter II that a zero-coupon bond is a security that pays the amount of 1 at maturity $M > 0$ and nothing in-between times 0 and M. One unit of the zero coupon bond can be purchased at the price $P(0, M)$ at time $t = 0$.

Assume that for each time $t \in \{1, \ldots, T\}$, we have a zero-coupon bond available with purchase price $P(0, t)$ at time 0.

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ be a given stream of liability cashflows with $C_t \geq 0$ for all t. Then, at time $t = 0$ and for each future time $t \in \{1, \ldots, T\}$, we can simply buy C_t units of the zero-coupon bond maturing at time t. In this case, we are completely immunized, even without an approximation error. The cost of achieving this immunization is

$$\widetilde{\mathsf{PV}}_0(\mathbf{C}) = \sum_{t=1}^{T} P(0, t) C_t$$

where $\widetilde{\mathsf{PV}}_0(\mathbf{C})$ is the alternative present value as in Section 2 of Chapter II.

Here again, we can compute the equivalent immunized discount rate:

$$\delta_{\mathsf{IMZ}} := \mathsf{IRR}\left(\mathbf{C}, \widetilde{\mathsf{PV}}_0(\mathbf{C})\right)$$

such that the immunized value matches the usual present value, i.e.

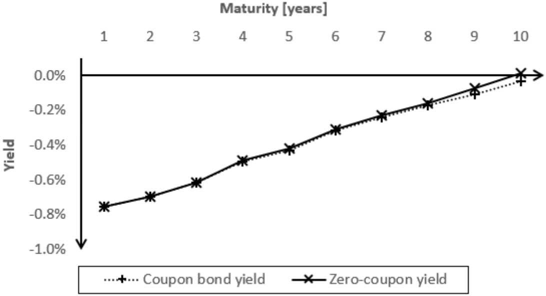
$$\mathsf{PV}_0(\mathbf{C}, \delta_{\mathsf{IMZ}}) = \sum_{t=1}^{T} \frac{C_t}{(1+\delta_{\mathsf{IMZ}})^t} = \widetilde{\mathsf{PV}}_0(\mathbf{C})$$

In most markets except the US, zero-coupon bond do not exist as traded securities. Hence, $\widetilde{\mathsf{PV}}_0(\mathbf{C})$ is, in principle, not a viable valuation.

However, as described in Section 2 of Chapter II, one can compute notional zero-coupon bond prices and rates. And the value $\widetilde{\mathsf{PV}}_0(\mathbf{C})$ obtained from these notional prices is a - typically fairly good - approximation of the actual immunized value.

Example: valuation with zero-coupon bonds

For our example, zero-coupon yields and yields-to-maturity of coupon bonds are not dramatically different:



Therefore, also the valuation resulting from the application of the zero-coupon curve is very similar from the one obtained with coupon bonds:

- Immunized value: $\widetilde{\mathsf{PV}}_0(\mathbf{C}) = 658'718$
- Equivalent immunized discount rate: $\delta_{\mathsf{IMZ}} = -0.35\%$

Immunization: conclusions

The concept of immunization is very appealing: By purchasing a portfolio of securities at time $t = 0$, we can - in principle - ensure the payment of all future promised cashflows without any financial risk remaining.

Therefore, the immunized value of a stream of liability cashflows is an important reference value: It denotes the amount of money that must be put up if one wants to ensure the payment of the liabilities without financial risk interfering. This is related to the economic concept of the certainty equivalent.

The immunized value has, however, also a number of drawbacks, notably:

- For longer-dated liabilities (e.g. $t \gg 20$ years), there may not exist enough bonds with corresponding maturities.
- One could still compute the immunized value by using extrapolated bond prices and yields (in particular for zero-coupon bonds).
- But these immunized values are only meaningful to a limited extent, because the immunizing position cannot actually be implemented. That is, in this case, there is financial risk remaining.

In the current situation, however, the main drawback is that the immunized value is very expensive. With current bond prices and yields (c.f. the example), the equivalent discount rates are around or even below zero, and one must put up very high amounts of money to assure payment of the liabilities without financial risk.

In our example, $\delta_{\mathsf{IM}} = -0.35\%$ for $\mathsf{IV}_0(\mathbf{C}, \delta_{\mathsf{IM}}) = 658'718$. Compare this to the values obtained with other discount rates.

δ	$\mathsf{PV}_0(\mathbf{C}, \delta)$
0%	$648'841$
1%	$621'990$
2%	$596'855$
3%	$573'298$
4%	$551'194$

The rest of this course will be basically about judging whether it makes sense to incur some financial risk in order to sustain a somewhat higher discount rate and achieve a lower valuation of the liabilities.

3. Simplification: duration matching

Introduction

The task here is the same as in the case of cashflow matching: We are given a stream of cashflows $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ with $C_t \geq 0$ for all t, and we want to set up a bond portfolio at time $t = 0$ in such a manner that the payment of $\mathbf{C}$ is assured without financial risk at $t > 0$.

If the assets held are only bonds, then financial risk arises from changes in interest rates. If interest rates change, then the value of the bonds held changes. But - at least in principle - one should also adapt the discount rate of liabilities, which causes a change in the value of liabilities.

If we set up the bond portfolio in such a way that its change in value is equal to the change in value of the liabilities, then we are - at least in principle - immunized:

- If interest rates rise, the value of the bonds diminishes. But so does the value of the liabilities if the discount rate is adapted accordingly.
- If interest rates fall, the discount rate of liabilities should be reduced. This causes the value of liabilities to rise. But so does the value of the bonds.

The approach is now to make sure that the movement on the asset side and on the liability side are (approximately) equal.

Interest rate sensitivity

Recall Chapter II: The measures of interest rate sensitivity for any stream of discounted cashflows - liabilities and bonds alike - are the duration (1st order) and the convexity (2nd order):

$$\text{Duration: } D(\mathbf{C}, \delta) = -\frac{\partial}{\partial \delta} \mathsf{PV}_0(\mathbf{C}, \delta) \Big/ \mathsf{PV}_0(\mathbf{C}, \delta)$$

$$\text{Convexity: } K(\mathbf{C}, \delta) = \frac{\partial^2}{\partial \delta^2} \mathsf{PV}_0(\mathbf{C}, \delta) \Big/ \mathsf{PV}_0(\mathbf{C}, \delta)$$

For a bond, the discount rate δ equals its yield to maturity $R(0, M)$. For a portfolio of bonds, the overall duration and convexity can be calculated or at least approximated by using Proposition 4 of Chapter II.

Duration matching

Let $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ with $C_t \geq 0$ for all t be a stream of liability cashflows, and let δ denote its discount rate.

Let $\mathbf{B}^1 \ldots, \mathbf{B}^N$ denote a set of bonds, each one with its yield to maturity $R^i(0, M^i)$ and market price B_0^i as of $t = 0$. Let $\mathbf{n} = (n_1 \ldots, n_N)' \in \mathbf{R}^N$ with $n_i \geq 0$ for all i denote the number of units of bond i that we hold in our portfolio.

To achieve duration (and convexity) matching, we must select $\mathbf{n}$ at $t = 0$ such that

$$\begin{aligned} D\left(\sum_{i=1}^{N} n_i \, \mathbf{B}^i, \delta_{\mathbf{n}}\right) &= D(\mathbf{C}, \delta) \\ K\left(\sum_{i=1}^{N} n_i \, \mathbf{B}^i, \delta_{\mathbf{n}}\right) &= K(\mathbf{C}, \delta) \\ \delta_{\mathbf{n}} &\geq \delta \\ \mathbf{n} &\geq 0 \end{aligned}$$

The additional convexity condition should normally be included in social and pension insurance settings. It may, however, be omitted in certain situations.

$\delta_{\mathbf{n}}$ denotes the overall yield to maturity of the bond portfolio , i.e.

$$\delta_{\mathbf{n}} = \mathsf{IRR}\left(\sum_{i=1}^{N} n_i \, \mathbf{B}^i, \; \sum_{i=1}^{N} n_i \, B_0^i\right)$$

We will see in the next chapter that the condition $\delta_{\mathbf{n}} \geq \delta$ is important.

If the set of available bonds is sufficiently rich, this problem will not have a unique solution, and we may be able to impose additional constraints or targets, e.g.

- maximize yield $\delta_{\mathbf{n}}$ or minimize purchase price $\sum_{i=1}^{N} n_i \, B_0^i$
- impose cashflow matching for the first few years

Note that, in general, there is absolutely no guarantee that the cashflows will match.

Immunization by duration matching

Assume that the yield in the bond market changes by $\Delta\delta$ for all maturities. This also means that $\delta_{\mathbf{n}}$ changes by $\Delta\delta$. Because of the condition $\delta_{\mathbf{n}} \geq \delta$, this also means that we should change the discount rate δ by $\Delta\delta$.

Relative effect on the asset side:

$$\frac{\Delta\,\mathsf{PV}\left(\sum_{i=1}^{N} n_i\,\mathbf{B}^i, \delta_{\mathbf{n}} + \Delta\delta\right)}{\mathsf{PV}\left(\sum_{i=1}^{N} n_i\,\mathbf{B}^i, \delta_{\mathbf{n}}\right)} = -D\left(\sum_{i=1}^{N} n_i\,\mathbf{B}^i, \delta_{\mathbf{n}}\right)(\Delta\delta) + \frac{1}{2}\,K\left(\sum_{i=1}^{N} n_i\,\mathbf{B}^i, \delta_{\mathbf{n}}\right)(\Delta\delta)^2 + o\left((\Delta\delta)^3\right)$$

Relative effect on the liability side:

$$\frac{\Delta\,\mathsf{PV}\left(\mathbf{C}, \delta + \Delta\delta\right)}{\mathsf{PV}\left(\mathbf{C}, \delta\right)} = -D\left(\mathbf{C}, \delta\right)(\Delta\delta) + \frac{1}{2}\,K\left(\mathbf{C}, \delta\right)(\Delta\delta)^2 + o\left((\Delta\delta)^3\right)$$

Due to the matching conditions, these relative changes will be equal. That is, if liabilities increase, then so will the assets used to cover them. And vice versa. To the extent that duration and convexity are a good measure for interest rate sensitivity, we have at least in principle an immunized position.

Example: duration-matched portfolios

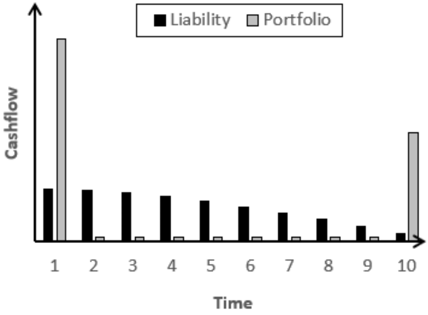
Two possible duration-matched portfolios and the cashflow-matched portfolio for comparison:

Bond i	Number of units in portfolio: Cheapest	CF match t=1,2	Full CF match
1	3'634	838	851
2	0	832	847
3	0	2'673	816
4	0	0	764
5	0	0	693
6	0	0	620
7	0	0	514
8	0	0	396
9	0	0	267
10	1'961	1'373	136
Portfolio price:	**652'486**	**656'489**	**658'718**

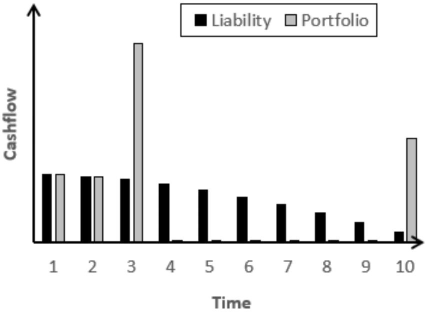
Example: cashflows under duration matching

Unless explicitly imposed, a duration-matched portfolio leads to cashflow patterns that are not necessarily related to the cashflow stream that must be financed.

Cheapest portfolio, no conditions imposed on cashflows:



Cashflow matching condition imposed for times $t = 1, 2$:



Duration matching: conclusions

By applying duration (and convexity) matching, one can create an (approximately) immunized position of assets and liabilities with less restrictive conditions than for cashflow matching.

One can also combine duration matching (for longer time horizons) and cashflow matching (for shorter time horizons).

Due to the restriction $\delta \leq \delta_{\mathbf{n}}$, also duration matching will only allow for very low discount rates under current market conditions, leading to very high liability values similar to the ones attained under cashflow matching.

Basic duration matching as shown here only works for parallel shifts of the yield curve where rates for all maturities move by the same amount $\Delta\delta$, which is not realistic in many situations. To circumvent this, there exist more sophisticated methods based e.g. on so-called key rate durations; see [1].

Duration and convexity can only be determined accurately for fixed income securities. If an institution also holds other asset classes like equities, real estate or alternative investments, then one can - in principle - estimate the durations of these assets, but these estimates lack accuracy. This is particularly the case nowadays when interest rates are mainly determined politically.

4. Alternative: loose coupling

Situation with immunization

With the immunizing approaches (to the extent that they are implementable), one is in a position where one can set up a portfolio of securities at time $t = 0$ in such a way that the payment of all subsequent cashflows C_t for all subsequent times $t \in \{1, \ldots, T\}$ ist assured (at least with a very high probability). This is a very comfortable position.

An immunizing portfolio can only be created by using (high-grade) bonds or - in practice - similar fixed income securities such as high-grade loans or private placements or derivatives such as bond futures or swaps.

The discount rate cannot be higher than the overall yield of the immunizing portfolio. Therefore, applicable discount rates tend to be very low, particularly in the current low-interest environment.

This leads to high valuations for some given stream of liability cashflows, and to a low contribution of investment returns to the overall financing of the liability.

Challenges and requirements

In many situations (e.g. in Swiss compulsory accident insurance), the discount rate for the liabilities is exogenously given, and the given level may be well above what is attainable with an immunizing approach.

Or there may be the desire or the requirement of the institution to have a higher investment income than just bond yields. Motivations for this may be:

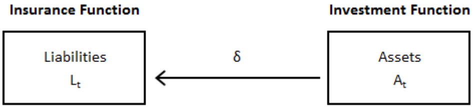
- Higher benefits for the same level of premium or contributions.
- Lower premium or contribution for the same benefits.
- Financing cost increases, particularly due to longevity.

Or the institution may simply have an in-force portfolio that contains substantial amounts of non-fixed income securities like equities, real estate or alternative investments.

In all these situations, a less restrictive coupling between the assets and the liabilities is necessary.

Loose coupling setup

We have a stream of liability cashflows $\mathbf{C} = (C_1, \ldots, C_T)' \in \mathbb{R}^T$ with given discount rate δ, such that the initial value of the liabilities is $L_0 = \mathsf{PV}_0(\mathbf{C}, \delta)$. For covering the liabilities, we hold assets valued A_0. For financing the liabilities, we subdivide the institution into two functions:



- The insurance function services the liabilities. For doing so, it receives a fixed interest δ from the investment function.
- The investment function invests the assets and generates investment proceeds IP_t. It uses these investment proceeds to finance the fixed interest payments δ.

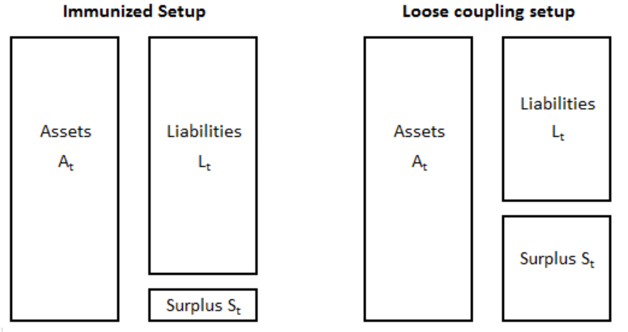
This subdivision into an insurance and an investment function is related to the principle of funds transfer pricing, see e.g. [2].

Now, assets and liabilities are only coupled by the discount rate δ and by their total values A_0 and L_0. Hence the name loose coupling.

Investment risk and the role of surplus

In an immunized setup, one need not care about changes in the value of the investment (bond) portfolio over time, since assets and liabilities are immunized.

In the loose coupling approach, investment proceeds are generally variable, and they may or may not be sufficient to fund the fixed rate δ in any given year. To compensate this uncertainty, the institution must hold a capital buffer, called surplus (or equity, in the corporate world).



In an immunized setup,the discount rate is low, and thus the valuation of the liabilities is high. But there is little to no risk from investing the assets due to the immunized position. Therefore, one needs to hold little or no surplus as a buffer against investment risk.

In a loosely coupled setup, one may set a higher discount rate, leading to lower valuations of the liabilities. But, in order to finance this discount rate, one has to incur a non-negligible investment risk. As a buffer against this, one needs, in turn, a higher surplus. Hence, the choice of discount rate has consequences in the form of investment risk and necessary surplus to compensate it.

Where is the optimum in this dynamic and circular relationship? What is the interplay between discount rate, investment risk and necessary surplus? What is feasible, and what is not feasible?

These questions are of high relevance, since many social and pension insurance institutions work according to the loose coupling principle, in particular Swiss pension funds and compulsory accident insurance.

In the subsequent chapters, we will suitably formalize the problem sketched here, and we will build a thorough understanding.

Bibliography I

[1] Scherer, B.
Liability Hedging and Portfolio Choice.
Risk Books, London, 2005.

[2] Wilson, T.
Value and Capital Management.
John Wiley & Sons, Chichester, 2015.

Financial Risk Management in Social and Pension Insurance

Chapter IV: The Asset / Liability Framework

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 17, 2020

Table of Contents

1. Working principle of capital-based insurance

Promised benefits and liabilities

The institution receives premium income or contributions from its clients or from third parties.

In return, the institution promises - and thus owes - to its clients certain well-specified future benefits, e.g.

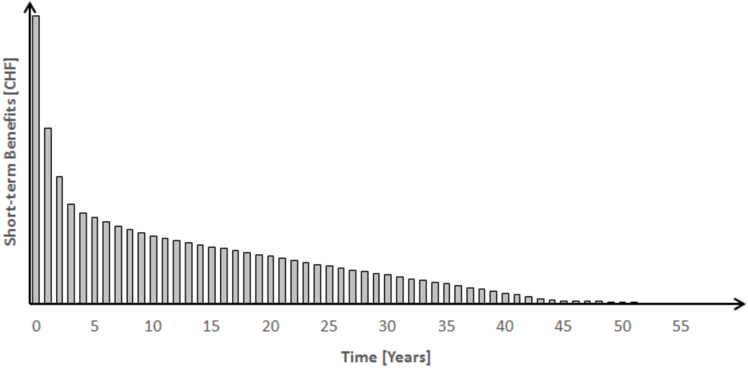
- old-age or disability pensions
- reimbursement of medical costs from injury or illness
- indemnities for foregone salaries
- subsidies for long-term care
- unemployment benefits

These benefits can be expressed as (estimated) future cashflows.

The present value of these future cashflows, possibly augmented by certain reinforcements, is the liability of the institution.

Accident insurance: short-term benefits

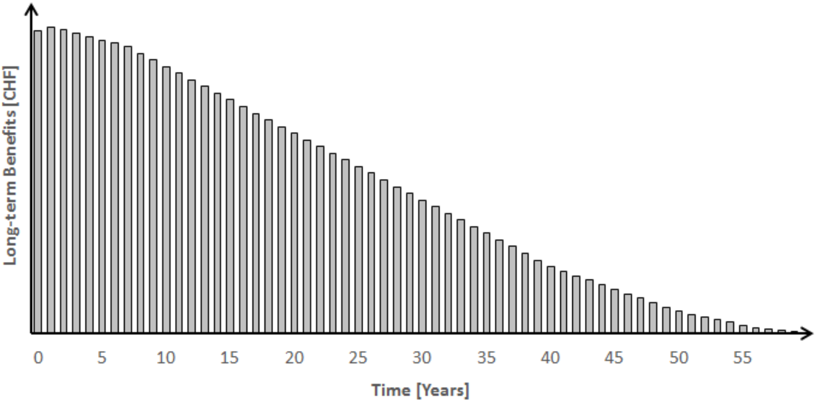
Future payments for medical treatment and income substitution related to accidents already happened.



"Short-term" is relative: After 5 years, 20% of ultimate payments are still open. And after 20 years, 5% of ultimate payments are still open. Some payments occur up to 75 years after the accident.

Accident insurance: long-term benefits

Future pension payments for disabled and next-of-kin from accidents already happened. Average age at beginning of pension is around 42 years.



Accident insurance: liabilities

Excerpt from the balance sheet of the Swiss National Accident Insurance Fund (see suva.ch):

		2019	2018
	Ref. no.	CHF in 1,000s	CHF in 1,000s
Liabilities and equity			
Provisions for short-term benefits		9 253 100	8 872 500
Provisions for long-term benefits		29 392 519	27 312 681
Provisions for cost-of-living allowances		169 961	102 662
Technical provisions	10	38 815 580	36 287 843
Non-technical provisions	11	11 714 976	10 473 557
Equity		3 155 340	3 023 584

Assets and investment proceeds

In order to cover the liabilities, the social insurance institution must hold an appropriate amount of assets.

These assets can be invested, e.g. in bonds, equities, real estate or alternative investments and thus generate investment proceeds.

The investment proceeds are an important part of the funding of the institution. Higher investment proceeds result in lower premia or higher benefits.

In order to facilitate planning and tariffication, a target (required return) is imposed for the investment proceeds.

By investing its assets (e.g. in equities), the institution incurs investment risk, i.e. the investment proceeds may be higher or lower than the target.

The institution must, hence, take appropriate precautions in order to deal with deviations of actual investment proceeds from required ones.

The extent to which the institution can absorb such deviations of actual investment proceeds from required ones is called the risk-taking capability.

Example: assets

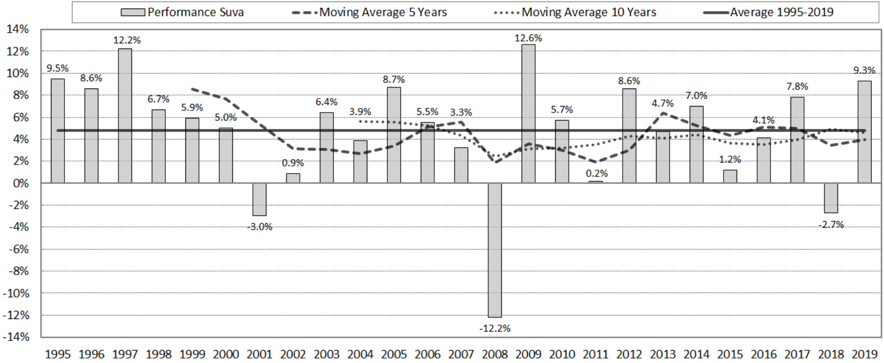
Excerpt from the Suva balance sheet:

Investment categories (including derivatives), at market	31.12.2018 Balance CHF in 1,000s	2019 Changes in value CHF in 1,000s	31.12.2019 Balance CHF in 1,000s
Liquid assets[3]	1 711 104	73 367	1 784 471
Mortgages	692 072	46 157	738 229
Loans and syndicated loans[1]	5 802 340	312 058	6 114 398
Bonds in CHF[3]	9 344 117	312 058	9 656 175
Bonds in foreign currency[3]	5 889 129	480 314	6 369 443
Indirect real estate investments[3]	1 340 837	183 976	1 524 814
Investment properties[2]	5 218 074	248 384	5 466 458
Shares in Switzerland[3]	3 086 220	488 487	3 574 706
Shares outside Switzerland[3]	6 987 752	1 106 474	8 094 226
Alternative investments[3,4]	9 324 095	545 711	9 869 806
Overlays, hedging and opportunities[3]	164 476	314 995	479 471
Total	**49 560 216**	**4 111 982**	**53 672 198**

Substantial portions of the assets are invested in risky asset classes such as equities or alternative investments.

Example: investment proceeds

History of investment returns of Suva:



Due to the relatively high percentage of equity holdings, performance fluctuated considerably from year to year.

On average over several years, however, the performance was sufficient to cover the financing needs.

Asset / Liability Management

The big picture obtained comprises the following elements:

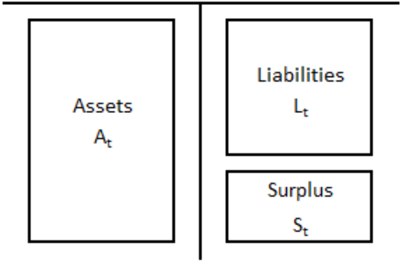
Institution	**Financial markets**
Required return	Realized investment proceeds
Risk-taking capability	Investment risk incurred

Asset / Liability Management (ALM) is the task of reconciling these different and often conflicting aspects such that the promised payouts can be made with high probability.

2. Model framework for social insurance

Balance sheet

The balance sheet describes the state of the institution at some point t in time:



Let t denote time in years; $t \in \{0, \ldots T\}$ for some final time $T < \infty$. Time $t = 0$ is the present, $t > 0$ denotes the future.

Let A_t denote the total value of assets at time t. Specifically, A_t is the market value of the assets at the end of year t.

The total assets can be subdivided into n asset classes $A_{i,t}$, i.e.

$$A_t = \sum_{i=1}^{n} A_{i,t} \quad \text{with } A_{i,t} \geq 0 \text{ for all } t \in \{0, \ldots T\} \text{ and } i \in \{1, \ldots n\}$$

$A_{i,t}$ can denote e.g. equities, bonds, real estate or alternative investments.

By $w_{i,t} := A_{i,t}/A_t$ we denote the portfolio weight of asset class i at time t.

The entire portfolio is specified by $\mathbf{w}_t = (w_{1,t}, \ldots\ w_{n,t})' \in \mathbb{R}^n$. and we have $w_{i,t} \geq 0$ and $\sum_{i=1}^{n} w_{i,t} = 1$.

By L_t we denote the value of the liabilities at time t, i.e. the present value of future cashflows plus necessary reinforcements; see next section for more details.

Surplus and Funding Ratio

Definition 1 (Surplus)

The surplus is the difference between assets and liabilities: $S_t = A_t - L_t$

Surplus is expressed in monetary units and bears limited information; e.g.

- 1 MCHF of surplus against 5 MCHF of liabilities (solid)
- 1 MCHF of surplus against 100 MCHF of liabilities (shaky)

This deficit is overcome by the funding ratio which relates surplus to liabilities:

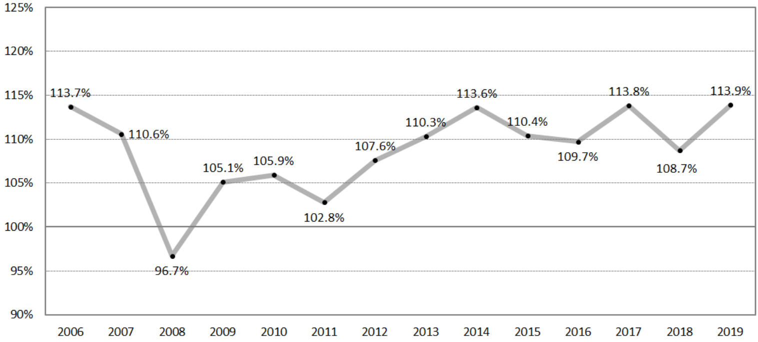
Definition 2 (Funding Ratio)

The funding ratio is the ratio of assets and liabilities:

$$\mathit{FR}_t = \frac{A_t}{L_t} = 1 + \frac{S_t}{L_t}$$

If $\mathrm{FR}_t \geq 1$, i.e. $A_t \geq L_t$ or $S_t \geq 0$, the institution is fully funded or overfunded. Otherwise, the institution is underfunded. The probability or the expected depth of an underfunding are important risk measures in the ALM context.

Example: Funding ratio of Swiss pension funds according to the Swisscanto pension fund studies (see www.swisscanto.ch):



Notice the considerable fluctuations. This development is owed to the financial crises of 2001 and 2008 and to the general downward trend in interest rates. Financial risk management is a serious issue. And it is a long-term issue.

Income statement

The income statement describes the change in the financial position of the institution in the t-th year, i.e. in $(t-1,\, t]$:

Income	**Expense**
Premium / Contribution P_t	Insurance benefits B_t
Investment proceeds IP_t	Revaluation of liabilities RL_t

P_t is the income from premia or contributions; this is an actual cash inflow.

IP_t denotes the proceeds from the investment of the assets in monetary units. This can be an actual cash inflow (e.g. dividends) or a non-cash change in value (positive or negative).

B_t is the expense from paying promised insurance benefits; this is an actual cash outflow.

RL_t is the non-cash expense for the intrinsic revaluation of liabilities; see next section for an in-depth treatment. It can be positive or negative.

We define $C_t := P_t - B_t$ as the net cashflow from insurance operations.

We could also define $\mathsf{NR}_t = P_t + \mathsf{IP}_t - B_t - \mathsf{RL}_t$ as the net result. This is, however, less important in social insurance. The main interest is focused on the balance sheet, in particular on the surplus S_t and the funding ratio FR_t.

Transition equation for assets

The change in the value of assets in year t depends on the investment proceeds obtained in the financial markets and on the net cashflow from insurance operations; the latter is assumed to take place at year-end:

$$A_t = A_{t-1} + \mathsf{IP}_t + C_t$$

Let R_t denote the rate of return from investing the assets, i.e.

$$R_t = \frac{\mathsf{IP}_t}{A_{t-1}} = \frac{A_t - C_t}{A_{t-1}} - 1 = \frac{A_{t^-}}{A_{t-1}} - 1$$

The return R_t only depends on the conditions in the financial markets and on how A_{t-1} is distributed on the different asset classes, but it does not depend on cashflows from insurance operations.

Example: Assume that $A_{t-1} = 100$, $R_t = 5.8\%$ and $C_t = 7.5$. Then we have $A_{t^-} = 100 \cdot (1 + 5.8\%) = 105.8$ and $A_t = A_{t^-} + C_t = 105.8 + 7.5 = 113.3$.

The transition equation can thus be restated as follows:

$$A_t = A_{t-1} + R_t A_{t-1} + C_t = (1 + R_t)A_{t-1} + C_t$$

We can also introduce returns on the level of single asset classes, i.e.

$$R_{i,t} = \frac{A_{i,t^-}}{A_{i,t-1}} - 1 \quad \text{and} \quad \mathbf{R}_t = (R_{1,t}, \, \ldots \, , \, R_{n,t})' \in \mathbb{R}^n$$

The return on the full asset portfolio is then given by the weighted sum of asset class returns, i.e.

$$R_t = \sum_{i=1}^{n} w_{i,t-1} R_{i,t} = \mathbf{w}'_{t-1} \mathbf{R}_t$$

The question of how to construct a suitable portfolio $\mathbf{w}_t$ from the different asset classes will be dealt with in the second part of the course. In the meantime, we will mainly consider the aggregate returns R_t.

Transition equation for liabilities

The change in the value of liabilities can be expressed in the same way as the change in the value of assets:

$$L_t = L_{t-1} + \mathsf{RL}_t + C_t$$

There is a change due to net cashflows from insurance operations, and there is a change due to the intrinsic revaluation. For the latter, we can define

$$\lambda_t = \frac{\mathsf{RL}_t}{L_{t-1}} = \frac{L_t - C_t}{L_{t-1}} - 1 = \frac{L_{t-}}{L_{t-1}} - 1$$

We call λ_t the liability return or the rate of intrinsic change of the liabilities. For details and examples, see the next section The transition equation then becomes

$$L_t = L_{t-1} + \lambda_t L_{t-1} + C_t = (1 + \lambda_t) L_{t-1} + C_t$$

Transition equation for surplus

For the assets and the liabilities, we have:

$$\begin{aligned} A_t - A_{t-1} &= R_t A_{t-1} + C_t \\ L_t - L_{t-1} &= \lambda_t L_{t-1} + C_t \end{aligned}$$

Since $S_t = A_t - L_t$, this leads to

$$\begin{aligned} S_t - S_{t-1} &= (R_t A_{t-1} + C_t) - (\lambda_t L_{t-1} + C_t) \\ &= R_t A_{t-1} - \lambda_t L_{t-1} \\ &= \mathsf{IP}_t - \mathsf{RL}_t \end{aligned}$$

That is, if surplus is to remain at least constant, the investment proceeds must be at least equal to the intrinsic revaluation of liabilities. For the return R_t, this means

$$R_t A_{t-1} - \lambda_t L_{t-1} \geq 0 \quad \text{or} \quad R_t \geq \lambda_t \frac{L_{t-1}}{A_{t-1}} = \frac{\lambda_t}{\mathsf{FR}_{t-1}}$$

Hence, the higher the funding ratio, the lower the required return. In practice, for the reasons stated above, the funding ratio is more important than the surplus.

Transition equation for funding ratio

Inserting the transition equations for assets and liabilities into the funding ratio, we obtain

$$\mathsf{FR}_t = \frac{A_t}{L_t} = \frac{(1+R_t)A_{t-1} + C_t}{(1+\lambda_t)L_{t-1} + C_t}$$

If $C_t = 0$, i.e. if contributions and benefits are in balance, this boils down to

$$\mathsf{FR}_t = \mathsf{FR}_{t-1}\frac{(1+R_t)}{(1+\lambda_t)}$$

That is, to keep the funding ratio at least constant, R_t must be at least equal to λ_t. In this case, λ_t specifies the required return.

For $C_t \neq 0$, this simple relation is distorted; consider

$$1 \overset{!}{=} \frac{\mathsf{FR}_t}{\mathsf{FR}_{t-1}} = \frac{A_t}{L_t}\cdot\frac{L_{t-1}}{A_{t-1}} = \frac{A_{t-1}(1+R_t) + C_t}{L_{t-1}(1+\lambda_t) + C_t}\cdot\frac{L_{t-1}}{A_{t-1}}$$

By rearranging, we obtain

$$A_{t-1}L_{t-1}(1+R_t) + C_t L_{t-1} = A_{t-1}L_{t-1}(1+\lambda_t) + C_t A_{t-1}$$

Dividing by $A_{t-1}L_{t-1}$ and rearranging, we obtain

$$R_t = \lambda_t + \frac{C_t}{L_{t-1}} - \frac{C_t}{A_{t-1}}$$

Using $A_{t-1} = \mathsf{FR}_{t-1}L_{t-1}$, we finally obtain

$$R_t = \lambda_t + \left(1 - \frac{1}{\mathsf{FR}_{t-1}}\right)\frac{C_t}{L_{t-1}}$$

That is, the required return depends not only on λ_t, but also on the current funding ratio and on the net cashflow in relation to the existing liabilities.

Using $L_{t-1} = \frac{A_{t-1}}{\mathsf{FR}_{t-1}}$, we obtain the alternative representation

$$R_t = \lambda_t + (\mathsf{FR}_{t-1} - 1)\frac{C_t}{A_{t-1}}$$

That is, if $\mathsf{FR}_{t-1} > 1$ and $C_t > 0$, this dilutes the surplus, and we must earn an extra return to compensate this dilution.

In general, this yields the following situation:

	underfunded $\mathsf{FR}_{t-1} < 1$	overfunded $\mathsf{FR}_{t-1} > 1$
net cash outflow $C_t < 0$	$R_t \overset{!}{>} \lambda_t$	$R_t \overset{!}{<} \lambda_t$
net cash inflow $C_t > 0$	$R_t \overset{!}{<} \lambda_t$	$R_t \overset{!}{>} \lambda_t$

The extent of the difference must be evaluated case by case.

3. Liabilities and required return

Overview

In the general model, the liabilities are assumed to behave according to the generic transition equation

$$L_t = L_{t-1} + \lambda_t L_{t-1} + C_t = (1 + \lambda_t) L_{t-1} + C_t$$

In the sequel, we explore what this means specifically for different forms of liabilities, in particular with respect to the intrinsic rate of change λ_t.

The most typical form of liabilities is when L_t is the present value of some stream of future cashflows. But there also exist other forms of liabilities that will be considered.

Liability is the present value of a cashflow stream

Assume that the liability is the present value of some stream of promised benefits $\mathbf{B} = (B_1, \ldots, B_T)' \in R^T$, i.e. for a fixed and given discount rate δ, we have

$$L_t = \mathsf{PV}_t(\mathbf{B}, \delta) = \sum_{s=t+1}^{T} \frac{B_s}{(1+\delta)^{s-t}}$$

How does the value of the liability change between $t-1$ and t?

$$\begin{aligned}
L_t - L_{t-1} &= \sum_{s=t+1}^{T} \frac{B_s}{(1+\delta)^{s-t}} - \sum_{s=t}^{T} \frac{B_s}{(1+\delta)^{s-t+1}} \\
&= \sum_{s=t+1}^{T} \frac{B_s}{(1+\delta)^{s-t}} - \sum_{s=t+1}^{T} \frac{B_s}{(1+\delta)^{s-t+1}} - \frac{B_t}{1+\delta} \\
&= \sum_{s=t+1}^{T} \frac{B_s}{(1+\delta)^{s-t}} - \frac{1}{1+\delta} \sum_{s=t+1}^{T} \frac{B_s}{(1+\delta)^{s-t}} - \frac{B_t}{1+\delta} \\
&= L_t \left(1 - \frac{1}{1+\delta}\right) - \frac{B_t}{1+\delta}
\end{aligned}$$

Solving this equation for the value of the liability L_t yields

$$L_t = (1 + \delta)L_{t-1} - B_t$$

That is, we have

- $\lambda_t = \delta$, i.e. the intrinsic rate of change equals the discount rate.
- $C_t = -B_t$, i.e. the cash outflow equals the liability payment.

Recalling the considerations from the previous section, this means that the discount rate δ is the main determinant of the required return. If the cash outflow is relatively small w.r.t. the total liability, and if the funding ratio is not too far away from 1, the required return is approximately equal to the discount rate δ.

Attention: This is based on the tacit assumption that the estimates for the cashflows are unbiased. If the estimates have to be revised from year to year, the intrinsic rate of change can be significantly different from δ.

Liability is a savings account

In a typical Swiss pension fund, the accumulation process of the active members before retirement works as follows:

- Each year t, the employees and the employer make a contribution totaling P_t that is credited to the fund.
- Each year t, an interest ρ_t is granted on the money in the fund.

In particular, all past contributions and all past interest credits are guaranteed and thus a liability with the following transition equation:

$$L_t = L_{t-1} + L_{t-1}\rho_t + P_t = L_{t-1}(1 + \rho_t) + P_t$$

That is, we have:

- $\lambda_t = \rho_t$, the credited interest rate,
- $C_t = P_t$, the contributions.

In a general setup, further cashflows would have to be added (benefits of free passage, retirements).

Attention: This is called "Beitragsprimat", but it is not equal to a defined contribution plan in Anglo-American terminology, but rather to a cash balance plan.

Other liabilities

Besides the above-mentioned elements, the balance sheet of a social insurance institution may also contain other liability elements, e.g.

- Actuarial reserves against fluctuations in the liability cashflows, e.g. due to mortality.
- Reserves for financing future adjustments of the liability parameters, e.g. to cater for longevity.

Most of these liabilities can be modeled in the form

$$L_t = (1 + \lambda_t)L_{t-1} + C_t$$

for some suitable choices of λ_t and C_t.

Portfolio of liabilities

Assume that the total liabilities consist of several positions, i.e.

$$L_t = \sum_{i=1}^{m} L_{i,t} \quad \text{with} \quad L_{i,t} = (1 + \lambda_{i,t}) L_{i,t-1} + C_{i,t} \quad \text{for all } i$$

Then, the consolidated rate of intrinsic change λ_t is a weighted sum of the single rates, i.e.

$$\lambda_t = \sum_{i=1}^{m} \frac{L_{i,t-1}}{L_{t-1}} \lambda_{i,t}$$

The cashflows simply add up: $C_t = \sum_{i=1}^{m} C_{i,t}$.

Hence, the generic model is quite broadly applicable.

4. Elements of the financial account

Recapitulation of generic model

We have developed the following generic model:

$$\left.\begin{aligned} A_t &= A_{t-1}(1+R_t) + C_t \\ L_t &= L_{t-1}(1+\lambda_t) + C_t \end{aligned}\right\}$$

With given value A_0 for initial assets and L_0 for initial liabilities, and suitable assumptions for the intrinsic growth rate λ_t and for the investment return R_t, the model can always be evaluated by straightforward stochastic simulation. Note:

- In many practical situations, the intrinsic growth rate λ_t of the liabilities can be modeled as deterministic. This is particularly the case if L_t already contains suitable reserve components to deal with fluctuations of actuarial risk.
- Unless we have a full immunization (see previous chapter), the investment returns R_t must be considered as a random variable.

Actuarial risk and required return

If λ were stochastic, $\mathsf{Var}[\lambda_t]$ or some other suitable risk measure like $\mathsf{VaR}_\alpha[\lambda_t]$ or $\mathsf{ES}_\alpha[\lambda_t]$ could be used to denote actuarial risk. However, in the sequel, λ_t will mostly be considered as deterministic.

The required return is the minimum value that the return R_t must attain such that $\mathsf{FR}_t \geq \mathsf{FR}_{t-1}$, i.e. remains at least constant. As seen before, we have:

$$R_t \overset{!}{\geq} \lambda_t + \left(1 - \frac{1}{\mathsf{FR}_{t-1}}\right) \frac{C_t}{L_{t-1}} = \lambda_t + (\mathsf{FR}_{t-1} - 1) \frac{C_t}{A_{t-1}}$$

If the net cashflow $C_t = 0$, this boils down to $R_t \geq \lambda_t$.

Expected return and investment risk

For all $t \in \{1, \ldots, T\}$, we consider R_t as a random variable on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, i.e. $R_t \sim F_t$ with $\mu_t := \mathbf{E}[R_t] < \infty$ and $\sigma_t^2 := \text{Var}[R_t] < \infty$.

- The assumption $\sigma_t^2 < \infty$ (finite second moments) is fairly realistic for annual investment returns. If it comes to higher moments, they may not always be assumed to exist.
- We do not impose any other assumptions at this point, neither Normal distribution nor absence of serial correlation.

We call $\mu_t := \mathbf{E}[R_t]$ the <u>expected return</u>.

The short-term <u>investment risk</u> may be denoted by $\sigma_t^2 = \text{Var}[R_t]$ or by any other suitable risk measure such as $\text{VaR}_\alpha[R_t]$ or $\text{ES}_\alpha[R_t]$

- The notion "short term" is important. This only describes risk for one out of many periods. This is a necessary ingredient, but not sufficient in order to cope with the long-term nature of social and pension insurance.

Risk-taking capability

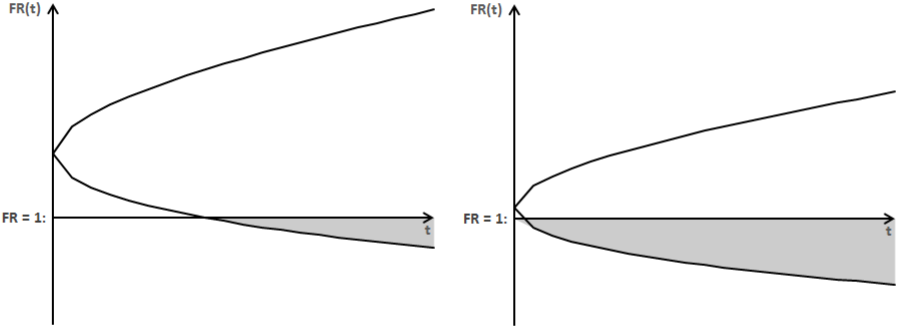
For the time being, we take the initial funding ratio FR_0 as a measure for the risk-taking capability. Intuitively, we have:

- The higher FR_0, the thicker is the cushion for absorbing insufficient investment proceeds without falling into underfunding.
- The lower FR_0, the thinner is the cushion for absorbing insufficient investment proceeds without falling into underfunding.

This is an approximation. Effectively, one would also have to to consider the possibilities that the institution has to increase income or reduce obligations when it falls into underfunding.

Example: High vs. low initial funding ratio.

High initial funding ratio | Low initial funding ratio



The lines represent the 1%- and 99%-quantiles of the funding ratio; the gray area below FR $= 1$ represents the state of underfunding.

Asset / Liability Management: overview

From the above, we obtain the following overall view:

Institution			**Financial markets**
Required return			Expected return
$\lambda_t + (\mathsf{FR}_{t-1} - 1)\frac{C_t}{A_{t-1}}$	$\leftarrow$	$\rightarrow$	μ_t
$\updownarrow$			$\updownarrow$
Risk-taking capability	$\leftarrow$	$\rightarrow$	Investment risk
FR_0			σ_t^2

The challenge is that these quantities are intricately related:

- higher expected return comes at the price of higher investment risk
- lower required return usually also entails a lower risk-taking capability

Asset / Liability Management is the task of reconciling these often-conflicting dimensions such that the payment of the promised insurance benefits is assured with high probability. This task must often be accomplished under additional constraints.

Financial risk

The goal of social an pension insurance is to assure with high probability that all promised payments $\mathbf{C} = (C_1, \ldots, C_T)'$ over the full time period $t = 1, \ldots, T$ can be made.

The proceeds from investing the assets are a source of funding for these promised payments. Financial risk consists of the danger that promised payments cannot be made at any time during the time span $t = 1, \ldots, T$ due to insufficient investment returns R_t.

There is no generally accepted way of quantifying financial risk; possible measures that will be explored subsequently include

- probability of underfunding $\psi_t = \mathbf{P}[\mathsf{FR}_t < 1]$
- Expected Funding Shortfall $\mathsf{EFS}_{\alpha,t}$ or Funding Ratio at Risk $\mathsf{FRaR}_{\alpha,t}$

for one or several time horizons t.

5. General random walk model and net profit condition

General random walk setup

We start from the standard model, i.e. from

$$\mathsf{FR}_t = \frac{A_{t-1}(1+R_t) + C_t}{L_{t-1}(1+\lambda_t) + C_t}$$

We make the following assumptions:

- Contributions equal benefits: $C_t = 0$ for all t.
- Constant intrinsic liability growth rate: $\lambda_t = \lambda$ for all t.
- Investment returns follow a random walk: $R_t = \mu + \varepsilon_t$ where $\varepsilon_t \sim$ iid with $\mathbf{E}[\varepsilon_t] = 0$ and $\mathsf{Var}[\varepsilon_t] < \infty$.

This leads to

$$\frac{\mathsf{FR}_t}{\mathsf{FR}_{t-1}} = \frac{1+\mu+\varepsilon_t}{1+\lambda}$$

Letting $\mathsf{LFR}_t = \log \mathsf{FR}_t$ and taking logarithms, we obtain

$$\mathsf{LFR}_t - \mathsf{LFR}_{t-1} = \log\left(\frac{1+\mu+\varepsilon_t}{1+\lambda}\right)$$

For small-enough values, we have $\log(1+R_t) \approx R_t$ and $\log(1+\lambda) \approx \lambda$, and we can use the approximation

$$\mathsf{LFR}_t = \mathsf{LFR}_0 + \sum_{s=1}^{t} X_s \quad \text{where} \quad X_s := \mu - \lambda + \varepsilon_s$$

or, equivalently

$$Z_t := \mathsf{LFR}_t - \mathsf{LFR}_0 = \sum_{s=1}^{t} X_s$$

That is, we have managed to represent the evolution of the funding ratio as a random walk. The condition $\mathsf{FR}_t > 1$ translates into $\mathsf{LFR}_t > 0$.

The assumption that the investment returns R_t form a random walk is not too unrealistic, at least for the annual returns that we consider here.

This allows us to apply a classic result from the theory of stochastic processes.

Proposition 1 (Random Walk Theorem)

Let $X_s \sim$ *iid with* $\mathbf{P}[X_s = 0] < 1$ *and* $\mathbf{E}[|X_s|] < \infty$*. Then, the random walk* $(Z_t)_{t \in \mathbb{N}_0}$ *with* $Z_t = \sum_{s=1}^{t} X_s$ *has one of the three following behaviors:*

- *if* $\mathbf{E}[X_s] > 0$ *then* $\lim_{t \to \infty} Z_t = +\infty$ $\mathbf{P}$*-a.s.*
- *if* $\mathbf{E}[X_s] < 0$ *then* $\lim_{t \to \infty} Z_t = -\infty$ $\mathbf{P}$*-a.s.*
- *if* $\mathbf{E}[X_s] = 0$ *then* $\liminf_{t \to \infty} Z_t = -\infty$ *and* $\limsup_{t \to \infty} Z_t = +\infty$ $\mathbf{P}$*-a.s.*

Proof: See e.g. the classical book by Resnick [1]. □

If $\mathbf{E}[X_s] < 0$, this may well mean that Z_t dwells in the positive range for quite some time. But sooner or later it will invariably become and remain negative. Hence, in order to assure a sustainable funding, we must have $\mathbf{E}[X_s] > 0$, i.e. $\mu > \lambda$.

Under our assumptions here, λ represents the required return, and μ the expected return. The theorem basically tells us that, irrespective of the initial value, the funding ratio will invariably go below 1 if the expected return is lower than the required return.

Fundamental principle

In order to ensure a sustainable long-term funding of the institution, it is necessary that the expected return be higher than the required return.

If this turns out to be unrealistic, the setup must be adapted such that the required return becomes lower.

General random walk: Lundberg bound

If the log-funding ratio follows a random walk, then we have for the infinite-time ruin probability

$$\psi(\mathsf{LFR}_0) \le e^{-R \cdot \mathsf{FR}_0}$$

where the Lundberg coefficient B is such that $M_{X_s}(R) = 1$ with $M(\cdot)$ being the moment-generating function of the innovations X_s; see e.g. [2].

This only yields an upper bound, and this bound is only valid for the infinite time horizon.

Moreover, this only works for thin-tailed distributions where the moment-generating function actually exists. In our context, this would almost invariably be the Normal one. But for this one, we have a much more powerful framework that we will explore in the sequel.

Therefore, we will not pursue this any further.

General random walk: Spitzer's formula

We define negative ladder heights Y_k as the difference between one local minimum and the next one. It can be shown that the ladder heights are iid with some distribution function $H(\,\cdot\,)$; see e.g. [2]. Spitzer's formula the states that

$$\psi(\mathsf{LFR}_0) = (1 - \psi(0)) \sum_{k \in \mathbb{N}} \psi(0)^k \left(1 - H^{*k}(\mathsf{LFR}_0)\right)$$

This time, we have an equality, but still only for the infinite time horizon. And the n-fold convolutions are rather cumbersome as well.

We see once more why one has to revert to stochastic simulation in most practical setups.

In order to maintain some analytical tractability, we must revert to a different, more restrictive model based on the normal distribution.

Bibliography I

[1] Resnick, S.
Adventures in Stochastic Processes.
Birkhäuser, Basel, 1992.

[2] Wüthrich, M.
Non-life insurance mathematics and statistics.
Lecture notes, ETH Zürich, Zürich, 2015.

Financial Risk Management in Social and Pension Insurance

Chapter V: The Lognormal Model

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 17, 2020

Table of Contents

1. Concept and model

Generic model

In the sequel, we follow the loose coupling approach, i.e. we allow for a general investment strategy and for investment risk and financial risk. That is, there may be a mismatch between required return and realized return.

Recall the generic model:

$$\left.\begin{array}{rcl} A_t & = & A_{t-1}(1+R_t)+C_t \\ L_t & = & L_{t-1}(1+\lambda_t)+C_t \end{array}\right\} \quad \text{for} \quad t \in \{1, \ldots, T\}$$

with the following generic elements:

- the initial value of assets (A_0) and liabilities (L_0)
- the stream of net cashflows C_t from insurance for $t \in \{1, \ldots, T\}$
- the intrinsic rate of change λ_t of the liabilities for $t \in \{1, \ldots, T\}$
- the investment returns $R_t \sim F_t$ for $t \in \{1, \ldots, T\}$ on some $(\Omega, \mathcal{F}, \mathbf{P})$

We assume that A_0 and L_0 are given. The cashflows C_t and the intrinsic rate of change of the liabilities λ_t may be deterministic (usually) or stochastic, wheras the investment returns R_t are always stochastic.

The generic model is fairly flexible. It can cover a wide range of situations and institutions, and it can also be adapted and extended if necessary.

In general, the model can only be evaluated by stochastic simulation. In practice, this is the usual and predominant procedure.

In order to obtain an analytically tractable model, we must make a number of restricting assumptions.

However, this analytically tractable model allows us to define a number of sensible risk measures and to explore their properties and the general interplay between assets and liabilities in an understandable and intuitive manner.

The concepts developed and the insights gained within the analytic framework can then be generalized to the stochastic simulation setup.

Thus, while not necessarily applicable directly, the analytical framework is a very important help for developing better simulation models and for better interpreting the outcomes of simulation models.

The specific model

We assume that the initial values of assets A_0 and liabilities L_0 are given.

Let us assume that the intrinsic growth rate of liabilities λ_t is deterministic and constant over time, i.e. $\lambda_t = \lambda$ for all $t \in \{1, \ldots, T\}$ and some given λ.

We assume that the institution is in equilibrium, i.e. contributions received and benefits paid cancel out such that $C_t = 0$ for all $t \in \{1, \ldots, T\}$.

We assume that the investment returns follow a Normal random walk, that is $R_t = \mu + \varepsilon_t$ for all $t \in \{1, \ldots, T\}$ with $\varepsilon_t \sim$ iid $\mathcal{N}(0, \sigma^2)$ for fixed values of μ and $\sigma^2 < \infty$. Equivalently $R_t \sim \mathcal{N}(\mu, \sigma^2)$.

Under these assumptions, we have

$$\frac{A_t}{L_t} = \frac{A_{t-1}(1 + \mu + \varepsilon_t)}{L_{t-1}(1 + \lambda)} \quad \text{or, equivalently} \quad \mathsf{FR}_t = \mathsf{FR}_{t-1}\, \frac{1 + \mu + \varepsilon_t}{1 + \lambda}$$

For small values of μ, λ and ε_t, we can use the approximation

$$\frac{1 + \mu + \varepsilon_t}{1 + \lambda} \;\approx\; 1 + \mu - \lambda + \varepsilon_t$$

This gives rise to the a modified model

$$\mathsf{FR}_t = \mathsf{FR}_{t-1}\left(1 + \mu - \lambda + \varepsilon_t\right)$$

Letting $\Delta\mathsf{FR}_t = \mathsf{FR}_t - \mathsf{FR}_{t-1}$, this is also equivalent to

$$\frac{\Delta\mathsf{FR}_t}{\mathsf{FR}_{t-1}} = \mu - \lambda + \varepsilon_t$$

And if we went into continuous time, this would translate into

$$\frac{d\,\mathsf{FR}_t}{\mathsf{FR}_t} = (\mu - \lambda)dt + \sigma dW_t$$

for a standard Brownian Motion $(W_t)_{t\in\mathbb{R}_0^+}$. That is, the funding ratio in continuous time follow a geometric Brownian Motion, and we could apply Itô's lemma to obtain

$$\mathsf{FR}_t = \mathsf{FR}_0 \exp\left\{\left(\mu - \lambda - \tfrac{1}{2}\sigma^2\right)t + \sigma W_t\right\}$$

This means that, given FR_0, FR_t is Lognormally distributed, i.e. $\mathsf{FR}_t \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ with $\tilde{\mu} = (\mu - \lambda - \frac{1}{2}\sigma^2)t + \log\mathsf{FR}_0$ and $\tilde{\sigma}^2 = \sigma^2 t$.

Models based on the geometric Brownian Motion, i.e. basically models of the Black-Scholes type, are extensively used in life insurance mathematics for the valuation of unit-linked life insurance contracts; see e.g. [3].

However, we want to remain in discrete time and follow a different avenue. If we consider directly

$$\mathsf{FR}_t = \mathsf{FR}_{t-1} \, \frac{1 + R_t}{1 + \lambda}$$

this leads to nothing analytically tractable. Product distributions are highly cumbersome, even in the simplest settings. Therefore, as in the general random walk setup in Section 5 of Chapter IV, we take logarithms to obtain

$$\log \mathsf{FR}_t - \log \mathsf{FR}_{t-1} = \log(1 + R_t) - \log(1 + \lambda)$$

Applying again the approximations $\log(1+R_t) \approx R_t$ and $\log(1+\lambda) \approx \lambda$, we obtain the the following model, which will turn out to be Lognormal.

The Lognormal model

Definition 1 (Lognormal model)

We assume that the logarithm of the funding ratio follows a Normal random walk, i.e. for a given FR_0 we have

$$\log FR_t - \log FR_{t-1} = \mu - \lambda + \varepsilon_t$$

for all $t \in \{1, \ldots, T\}$ and for given values $\mu, \lambda \in \mathbb{R}$, with $\varepsilon_t \sim$ iid $\mathcal{N}(0, \sigma^2)$ for some given σ^2 with $0 < \sigma^2 < \infty$.

By iterating the definition, we obtain

$$\begin{aligned} \log \mathsf{FR}_t &= \log \mathsf{FR}_{t-1} + (\mu - \lambda) + \varepsilon_t \\ &= \log \mathsf{FR}_{t-2} + (\mu - \lambda) + \varepsilon_{t-1} + (\mu - \lambda) + \varepsilon_t \\ &= \ldots \\ &= \log \mathsf{FR}_0 + (\mu - \lambda)t + \sum\nolimits_{s=1}^{t} \varepsilon_s \end{aligned}$$

Since the sum of independent Normal random variables is again normally distributed, we have: $\log \mathsf{FR}_t \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ with $\tilde{\mu} = \log \mathsf{FR}_0 + (\mu - \lambda)t$ and $\tilde{\sigma}^2 = \sigma^2 t$. This also means that FR_t is Lognormally distributed:

Proposition 1 (Lognormal model)

Under the model set forth in Definitionn 1, the funding ratio has a Lognormal distribution, i.e. $FR_t \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ *with*

- $\tilde{\mu} = \log FR_0 + (\mu - \lambda)t$
- $\tilde{\sigma}^2 = \sigma^2 t$

Before exploring the consequences of this, we study a few essential properties of the Lognormal distribution and of moment generating functions.

2. Moment generating functions

Definition 2 (Moment Generating Function (MGF))

Let $X \sim F$ be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ taking values in $\mathbb{R}$. Let $r \in \mathbb{R}$. The Moment Generating Function (MGF) is given by

$$M_X(r) = \mathbf{E}\left[\exp\left\{rX\right\}\right] = \int_{\mathbb{R}} \exp\left\{rx\right\} dF(x)$$

If X is absolutely continuous with density function $f(x)$, then we also have

$$M_X(r) = \int_{\mathbb{R}} \exp\left\{rx\right\} f(x)dx$$

The MGF is closely related to the Laplace transform of a function.

The MGF is a versatile and useful tool for many situations in probability theory and its applications.

Proposition 2

Assume that there exists some $r_0 > 0$ *such that we have* $M_X(r) < \infty$ *for all* $r \in (-r_0, +r_0)$. *Then we have*

$$M_X(r) = \sum_{k=0}^{\infty} \frac{r^k}{k!} \, \mathbf{E}\left[X^k\right]$$

Proof: Just a sketch; the basic line of argumentation is

$$M_X(r) = \mathbf{E}\left[\exp\{rX\}\right] = \mathbf{E}\left[\sum_{k=0}^{\infty} \frac{(rX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{r^k}{k!} \, \mathbf{E}\left[X^k\right]$$

The last equality is not trivial and takes some additional reasoning [4]. □

Taking derivatives and remembering that $0! = 1$, it immediately follows that

$$\frac{d^k}{dr^k} \, M_X(r)\bigg|_{r=0} = \mathbf{E}\left[X^k\right] < \infty$$

Hence the name moment generating function.

Proposition 3

Let $X \sim F_X$ and assume that there exists some $r_0 > 0$ such that $M_X(r) < \infty$ for all $r \in (-r_0, +r_0)$. Then, F is completely determined by M_X.

That is, if we have X and Y with $M_X = M_Y$, then $F_X = F_Y$.

Proof: See e.g. [1]. $\square$

This is the first main use of the MGF: It allows us to show that two random variables actually have the same probability law.

Proposition 4

Let $X_1, \ldots, X_n$ be independent random random variables. Then we have

$$M_{\sum_{i=1}^n X_i}(r) = \prod_{i=1}^n M_{X_i}(r)$$

Proof: By using the properties of exponential functions and independence, we obtain

$$\begin{aligned} M_{\sum_{i=1}^n X_i}(r) &= \mathbf{E}\left[\exp\left\{r \sum\nolimits_{i=1}^n X_i\right\}\right] \\ &= \mathbf{E}\left[\prod\nolimits_{i=1}^n \exp\{rX_i\}\right] \\ &= \prod\nolimits_{i=1}^n \mathbf{E}\left[\exp\{rX_i\}\right] \\ &= \prod\nolimits_{i=1}^n M_{X_i}(r) \qquad \square \end{aligned}$$

This is the second main use of MGF, i.e. obtaining the law of sums of independent random variables. For more on MGF and their uses, see e.g. [2] or [4].

3. Lognormal distribution

Normal distribution

The Normal distribution and its properties are assumed to be known; otherwise refer e.g. to [2]. This is just for completeness.

Definition 3 (Normal distribution)

Let $\tilde{\mu} \in \mathbb{R}$ *and* $\tilde{\sigma}^2 \in \mathbb{R}^+$ *with* $\tilde{\mu}, \tilde{\sigma}^2 < \infty$*. A random variable* X *has a Normal distribution with parameters* $\tilde{\mu}$ *and* $\tilde{\sigma}^2$*, i.e.* $X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ *if, for all* $x \in \mathbb{R}$*:*

$$\mathbf{P}\left[X \leq x\right] = \Phi\left(\frac{x - \tilde{\mu}}{\tilde{\sigma}}\right)$$

where

$$\Phi(x) = \int_{-\infty}^{x} \varphi(\xi)\, d\xi \quad \textit{and} \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\, x^2\right\}$$

A random variable with $X \sim \mathcal{N}(0, 1)$ is said to have a <u>standard Normal distribution</u>. The function $\Phi(x)$ is the <u>cumulative distribution function</u> of the standard Normal distribution, and $\varphi(x)$ is its <u>density</u>. We will use these two functions frequently in the sequel.

Note and remember that $\varphi(x) = \Phi'(x) > 0$ for all $x \in \mathbb{R}$.

The density of a general Normal random variable $X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ is then

$$f(x) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left\{-\frac{1}{2}\frac{(x-\tilde{\mu})^2}{\tilde{\sigma}^2}\right\}$$

Proposition 5 (Moments of the Normal distribution)

Let $X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$. Then we have:

- *Expectation:* $\mathbf{E}[X] = \tilde{\mu}$
- *Variance:* $\mathit{Var}[X] = \tilde{\sigma}^2$
- *Moment-generating function:* $M_X(r) := \mathbb{E}\left[e^{rX}\right] = \exp\left\{r\tilde{\mu} + \frac{1}{2}r^2\tilde{\sigma}^2\right\}$

Proof: First, we derive the moment-generating function of a Normal random variable:

$$
\begin{aligned}
M_X(r) &= \int_{-\infty}^{+\infty} e^{rx} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left\{-\frac{(x-\tilde{\mu})^2}{2\tilde{\sigma}^2}\right\} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left\{-\frac{x^2 - 2\left(\tilde{\mu} + r\tilde{\sigma}^2\right)x + \tilde{\mu}^2}{2\tilde{\sigma}^2}\right\} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left\{-\frac{x^2 - 2\left(\tilde{\mu} + r\tilde{\sigma}^2\right)x + \left(\tilde{\mu}^2 + 2\tilde{\mu}r\tilde{\sigma}^2 + r^2\tilde{\sigma}^4\right) - 2\tilde{\mu}r\tilde{\sigma}^2 - r^2\tilde{\sigma}^4}{2\tilde{\sigma}^2}\right\} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left\{-\frac{\left(x - \left(\tilde{\mu} + r\tilde{\sigma}^2\right)\right)^2}{2\tilde{\sigma}^2}\right\} dx \; \exp\left\{\frac{2\tilde{\mu}r\tilde{\sigma}^2 + r^2\tilde{\sigma}^4}{2\tilde{\sigma}^2}\right\} \\
&= \exp\left\{r\tilde{\mu} + \tfrac{1}{2}r^2\tilde{\sigma}^2\right\}
\end{aligned}
$$

Then, we use the moment-generating function to derive the moments:

$$
\begin{aligned}
\left.\frac{d}{dr} M_X(r)\right|_{r=0} &= \exp\left\{r\tilde{\mu} + \tfrac{1}{2}r^2\tilde{\sigma}^2\right\} \left.\left(\tilde{\mu} + r\tilde{\sigma}^2\right)\right|_{r=0} \\
&= \tilde{\mu} \\
&= \mathbf{E}\left[X\right]
\end{aligned}
$$

Proof: (continued)

$$\frac{d^2}{dr^2} M_X(r)\bigg|_{r=0} = \exp\left\{r\tilde{\mu} + \tfrac{1}{2}r^2\tilde{\sigma}^2\right\}\left(\tilde{\mu} + r\tilde{\sigma}^2\right)^2 + \exp\left\{r\tilde{\mu} + \tfrac{1}{2}r^2\tilde{\sigma}^2\right\}\tilde{\sigma}^2\bigg|_{r=0}$$

$$= \tilde{\mu}^2 + \tilde{\sigma}^2$$

$$\text{Var}\,[X] = \mathbf{E}\left[X^2\right] - \mathbf{E}\left[X\right]^2$$

$$= \tilde{\sigma}^2 \qquad \square$$

Proposition 6 (Sum of Normal random variables)

Let $X_1, \ldots, X_n \sim$ *iid* $\mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$. *Then* $\sum_{i=1}^n X_i \sim \mathcal{N}(n\tilde{\mu}, n\tilde{\sigma}^2)$.

Proof: Using the properties of the moment-generating function:

$$M_{\sum_{i=1}^n X_i}(r) = \prod_{i=1}^n M_{X_i}(r) = \prod_{i=1}^n \exp\left\{r\tilde{\mu} + \tfrac{1}{2}r^2\tilde{\sigma}^2\right\} = \exp\left\{r(n\tilde{\mu}) + \tfrac{1}{2}r^2(n\tilde{\sigma}^2)\right\}$$

The latter is the moment-generating function of a Normal random variable with parameters $n\tilde{\mu}$ and $n\tilde{\sigma}^2$, and only of this one. □

Lognormal distribution

A Lognormal random variable is a random variable the logarithm of which has a Normal distribution; see also [4].

Definition 4 (Lognormal distribution)

Let $X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$*. Then the random variable* $Y := e^X$ *has a Lognormal distribution; formally* $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$*.*

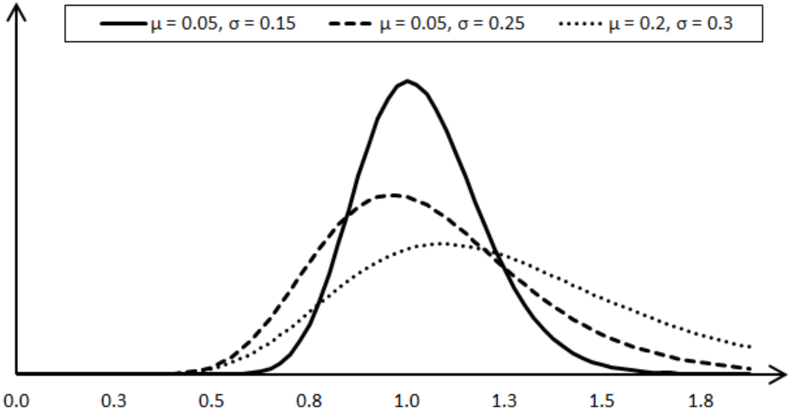
Proposition 7 (Cumulative distribution function and density)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$*. Then we have for* $y > 0$*:*

- *Cumulative distribution function:* $G(Y) = \mathbf{P}\left[Y \leq y\right] = \Phi\left(\frac{\log y - \tilde{\mu}}{\tilde{\sigma}}\right)$
- *Density:* $g(y) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \frac{1}{y} \exp\left\{-\frac{1}{2} \frac{(\log y - \tilde{\mu})^2}{\tilde{\sigma}^2}\right\}$

Proof: If $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$, then $Y = e^X$ with $X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$. But then, we have: $\mathbf{P}\left[Y \leq y\right] = \mathbf{P}\left[e^X \leq y\right] = \mathbf{P}\left[X \leq \log y\right] = \Phi\left(\frac{\log y - \tilde{\mu}}{\tilde{\sigma}}\right)$. The density is obtained by taking the first derivative. $\square$

The Lognormal distribution dwells only on the positive half-axis. Its density can have a wide variety of shapes, for instance:



Therefore, the Lognormal distribution appears well-suited for the modeling of funding ratios which are always positive and will usually cluster around one.

Proposition 8 (Scaling property)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ *and* $\rho > 0$*. Then* $\rho Y \sim LN(\tilde{\mu} + \log\rho, \tilde{\sigma}^2)$*.*

Proof: $\mathbf{P}[\rho Y \le y] = \mathbf{P}\left[Y \le \frac{y}{\rho}\right] = \Phi\left(\frac{\log\frac{y}{\rho} - \tilde{\mu}}{\tilde{\sigma}}\right) = \Phi\left(\frac{\log y - (\tilde{\mu} + \log\rho)}{\tilde{\sigma}}\right) \quad \square$

Proposition 9 (Moments)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$*. Then we have*

- *Expectation:* $\mathbf{E}[Y] = \exp\left\{\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right\}$
- *Variance: Var*$[Y] = \exp\left\{2\tilde{\mu} + \tilde{\sigma}^2\right\}\left(\exp\left\{\tilde{\sigma}^2\right\} - 1\right)$

Proof: Since $Y = e^X$ with $X \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$:
$\mathbf{E}[Y] = \mathbf{E}\left[e^X\right] = M_X(1) = \exp\left\{\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right\}$
Similarly $\mathbf{E}\left[Y^2\right] = M_X(2) = \exp\left\{2\tilde{\mu} + 2\tilde{\sigma}^2\right\}$,
then Var$[Y] =$E$\left[Y^2\right] -$ E$[Y]^2 \quad \square$

Proposition 10 (Loss size index)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ *and* $y > 0$. *Then we have:*

$$I(y) := \frac{\mathbf{E}\left[\mathbf{1}_{\{Y \le y\}} Y\right]}{\mathbf{E}\left[Y\right]} = \Phi\left(\frac{\log y - (\tilde{\mu} + \tilde{\sigma}^2)}{\tilde{\sigma}}\right)$$

Proof: Using the expressions for the density and the expectation:

$$\begin{aligned}
I(y) &= \int_0^y x\, g(x)\, dx \Big/ \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\} \\
&= \int_0^y \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\ \exp\{\log x\}\ \exp\left\{-\frac{(\log x - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right\}\ \exp\left\{-\tilde{\mu} - \frac{\tilde{\sigma}^2}{2}\right\}\ \frac{1}{x}\ dx \\
&= \int_0^y \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\ \exp\left\{-\frac{(\log x)^2 - 2(\tilde{\mu} + \tilde{\sigma}^2)\log x + (\tilde{\mu}^2 + 2\tilde{\mu}\tilde{\sigma}^2 + \tilde{\sigma}^4)}{2\tilde{\sigma}^2}\right\}\ \frac{1}{x}\ dx \\
&= \int_0^y \frac{1}{\sqrt{2\pi}\tilde{\sigma}}\ \exp\left\{-\frac{(\log x - (\tilde{\mu} + \tilde{\sigma}^2))^2}{2\tilde{\sigma}^2}\right\}\ \frac{1}{x}\ dx \\
&= \Phi\left(\frac{\log y - (\tilde{\mu} + \tilde{\sigma}^2)}{\tilde{\sigma}}\right) \qquad \square
\end{aligned}$$

4. Probability of underfunding

Funding ratio: distribution and moments

Under the model as set forth in Definition 1, and according to Proposition 1, given FR_0 and for $t > 0$, we have for the funding ratio:

$$\mathsf{FR}_t \sim LN(\tilde{\mu}, \tilde{\sigma}^2) \quad \text{with} \quad \tilde{\mu} = \log \mathsf{FR}_0 + (\mu - \lambda)t \quad \text{and} \quad \tilde{\sigma}^2 = \sigma^2 t$$

By applying Proposition 9, we obtain:

$$\mathbf{E}\left[\mathsf{FR}_t \middle| \mathsf{FR}_0\right] = \mathsf{FR}_0 \, \exp\left\{(\mu - \lambda + \tfrac{1}{2}\sigma^2)t\right\} \tag{1}$$

$$\mathsf{Var}\left[\mathsf{FR}_t \middle| \mathsf{FR}_0\right] = (\mathsf{FR}_0)^2 \exp\left\{2(\mu - \lambda + \tfrac{1}{2}\sigma^2)t\right\} \left(\exp\left\{\sigma^2 t\right\} - 1\right) \tag{2}$$

We note that the expected value of the funding ratio actually increases as the (short-term) investment risk σ^2 increases. This should, however, be considered with caution and in the context of the probability of underfunding hereafter. Note in particular that also the variance of the funding ratio increases as σ^2 increases.

Underfunding

Recall that an institution is underfunded when its assets A_t are insufficient to cover its liabilities L_t, i.e. $A_t < L_t$, or equivalently $\mathrm{FR}_t = A_t/L_t < 1$.

In the corporate world, this corresponds to an insolvency. However, for a social insurance institution, this does not necessarily mean bankruptcy and liquidation.

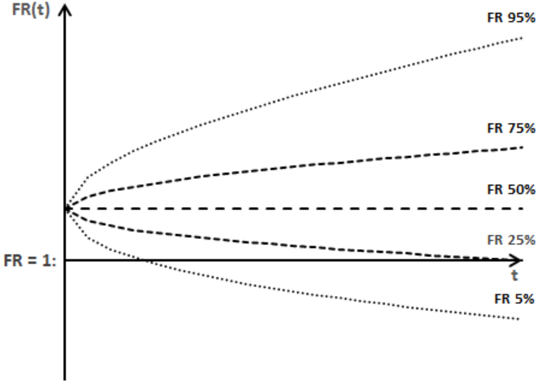
Often, the existence of the institutions is guaranteed by law. This does not imply a financial guarantee, but just the guarantee that the institution will have a client base and premium or contribution revenue over an unlimited period of time.

Therefore, social insurance institutions are often - at least under certain circumstances - allowed to be temporarily underfunded. This may represent an important advantage.

Nevertheless, underfunding is an undesirable state, and the probability of this happening is an important long-term risk measure.

Funding ratio distribution

The graphic shows the evolution of the funding ratio distribution over time, as expressed by some quantiles:



Note the asymmetry in the distribution of the funding ratio.

Probability of underfunding

Definition 5 (Probability of underfunding)

Let $t > 0$. Then the probability of underfunding at time t is defined as

$$\psi_t := \mathbf{P}\left[\, \mathit{FR}_t \leq 1 \middle|\, \mathit{FR}_0 \right]$$

For technical reasons, we use "$\leq$" rather that "$<$". In the Lognormal model, we immediately obtain the following result:

Proposition 11 (Probability of underfunding in the Lognormal model)

Let the Lognormal model according to Definition 1 hold. Then, the probability of underfunding is given by

$$\psi_t = \psi_t(\mathit{FR}_0, \lambda, \mu, \sigma^2) = \Phi\left(-\frac{\log \mathit{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

Proof: Recall that the Lognormal CDF is $\Phi\left(\frac{\log y - \tilde{\mu}}{\tilde{\sigma}}\right)$ and use $y = 1$. $\square$

Sensitivities

Sensitivity is an important concept in risk management. It consists of exploring how a risk measure reacts to changes in one of the input parameters. Specifically, we consider the probability of underfunding:

$$\psi_t = \psi_t(\mathsf{FR}_0, \lambda, \mu, \sigma^2) = \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

and we take partial derivatives w.r.t. the input parameters:

- FR_0: initial funding ratio
- λ: intrinsic rate of growth of the liabilities ($=$ required return)
- μ: expected investment return
- σ^2: (short-term) investment risk

These sensitivities are to be understood "ceteris paribus", i.e. all else remaining equal. In practice, input factors will not change independently from one another. For instance, a higher expected return will come along with a higher investment risk. This will be investigated in the subsequent chapters.

Sensitivity to initial funding ratio

Under the Lognormal model, we have:

$$\begin{aligned}\frac{\partial \psi_t}{\partial \mathsf{FR}_0} &= \frac{\partial}{\partial \mathsf{FR}_0}\, \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right) \\ &= -\frac{1}{\sigma\sqrt{t}}\, \varphi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right) \frac{1}{\mathsf{FR}_0} < 0\end{aligned}$$

Since $\varphi(\,\cdot\,)$ is the density of the standard Normal distribution, it is always positive; the same holds for FR_0 and σ. Therefore, the sensitivity $\frac{\partial \psi_t}{\partial \mathsf{FR}_0}$ is always negative.

That is, a higher initial funding ratio leads to a lower probability of underfunding, and vice versa.

Sensitivity to intrinsic growth rate of liabilities

Under the Lognormal model, we have:

$$\frac{\partial \psi_t}{\partial \lambda} = \frac{\partial}{\partial \lambda}\ \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$
$$= \frac{\sqrt{t}}{\sigma}\ \varphi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right) > 0$$

For the same reasons as above, this sensitivity is always positive.

That is, the higher the intrinsic growth rate of liabilities (and hence the required return), the higher becomes the probability of underfunding, and vice versa.

Sensitivity to expected return

Under the Lognormal model, we have:

$$\begin{aligned}\frac{\partial \psi_t}{\partial \mu} &= \frac{\partial}{\partial \mu}\, \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)\\ &= -\frac{\sqrt{t}}{\sigma}\, \varphi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right) < 0\end{aligned}$$

For the same reasons as above, this sensitivity is always negative.

That is, the higher the expected return, the lower becomes the probability of underfunding, and vice versa.

Sensitivity to short-term investment risk

Under the Lognormal model, we have:

$$\frac{\partial \psi_t}{\partial \sigma} = \frac{\partial}{\partial \sigma}\, \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$
$$= \frac{1}{\sigma^2\sqrt{t}}\, \varphi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)(\log \mathsf{FR}_0 + (\mu - \lambda)t)$$

The first two terms are always positive. The last term on the right-hand side, and hence the entire sensitivity, can be both positive or negative.

If the institution is in a good financial condition, i.e. if $\mathsf{FR}_0 > 1$ and $\mu \geq \lambda$, then $\frac{\partial \psi_t}{\partial \sigma} > 0$, i.e. taking more investment risk increases the probability of underfunding.

There are, however, situations where $\frac{\partial \psi_t}{\partial \sigma}$ can be negative. Consider the condition

$$\log \mathsf{FR}_0 + (\mu - \lambda)t < 0 \quad \text{or, equivalently} \quad \log \mathsf{FR}_0 < -(\mu - \lambda)t$$

That is (assuming $\mu \geq \lambda$), if the institution is sufficiently underfunded, taking higher investment risk may lead to a lower probability of underfunding. Although not obvious at first sight, it makes sense if one studies the situation in-depth.

Probability of underfunding: assessment

The probability of underfunding is a good risk measure, and one often used in practice, for a number of reasons:

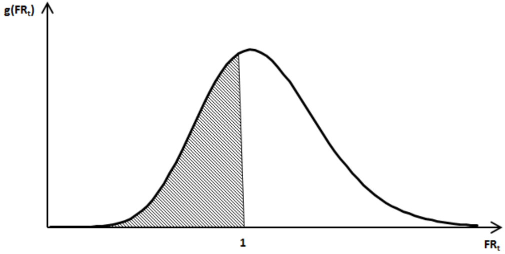
- It is sensible in that it addresses a serious danger.
- It takes a long-term view that is adapted to the nature of social and pension insurance.
- It can be easily calculated in analytical models, and it can be easily estimated by stochastic simulation.
- It lends itself to intuitive interpretation, and it is easy to explain also to non-quantitative audiences.

An important disadvantage is that it only indicates the probability of an underfunding to occur. We would, however, also like to know something about the extent of the underfunding. To this end, we explore a few alternative risk measures in the next section.

5. Alternative risk measures

Concept

Consider the distribution of the funding ratio, here under the Lognormal model:



Contrary to normal practice in insurance risk management, we are not interested in the right tail. The danger consists of funding ratios below one, i.e. lies in the left tail. We consider the following measures for the funding ratio:

1. The left α-quantile: $q_\alpha(\mathsf{FR}_t)$
2. The left expected shortfall: $\mathbf{E}\left[\mathsf{FR}_t \middle| \mathsf{FR}_t \leq q_\alpha(\mathsf{FR}_t)\right]$
3. The conditional expectation: $\mathbf{E}\left[\mathsf{FR}_t \middle| \mathsf{FR}_t \leq 1\right]$

In all three cases, we are ultimately interested in the difference between these three values and 1 (provided this difference is positive), i.e.

$$\max\left(1 - \text{Measure 1, 2 or } 3, 0\right)$$

because this is the funding deficit that must actually be financed in the event of an underfunding. In this way, we obtain the desired measures that also contain information on the extent of an eventual underfunding.

We derive these quantities in the Lognormal framework and explore their properties. Note, however, that they all can be easily estimated also by using stochastic simulation (with some care to be taken such that the tail is sufficiently populated with data points).

For simplicity, we start with a generic Lognormal random variable $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ and insert the specific values for the Lognormal funding ratio model afterwards.

Left quantile

Proposition 12 (Left quantile)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ *and* $\alpha \in (0,1)$. *Then we have:*

$$q_\alpha(Y) = \exp\left\{\tilde{\mu} + \Phi^{-1}(\alpha)\,\tilde{\sigma}\right\}$$

Proof: We are looking for y_0 such that $\mathbf{P}\left[Y \leq y_0\right] = \alpha$, i.e.

$$\begin{aligned} \Phi\left(\frac{\log y_0 - \tilde{\mu}}{\tilde{\sigma}}\right) &= \alpha \\ \frac{\log y_0 - \tilde{\mu}}{\tilde{\sigma}} &= \Phi^{-1}(\alpha) \\ y_0 &= \exp\left\{\tilde{\mu} + \Phi^{-1}(\alpha)\,\tilde{\sigma}\right\} \qquad \square \end{aligned}$$

Typical values of α are e.g. 1% or 5%.

Conditional expectation

Proposition 13 (Conditional expectation)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$. *Then we have:*

$$\mathbf{E}\left[Y \middle| Y \leq 1\right] = \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\} \Phi\left(-\frac{\tilde{\mu} + \tilde{\sigma}^2}{\tilde{\sigma}}\right) \Big/ \Phi\left(-\frac{\tilde{\mu}}{\tilde{\sigma}}\right)$$

Proof: On the one hand, we have according to Proposition 10

$$\begin{aligned}\mathbf{E}\left[\mathbf{1}_{\{Y \leq 1\}} Y\right] &= \mathbf{E}\left[Y\right] \Phi\left(\frac{\log 1 - (\tilde{\mu} + \tilde{\sigma}^2)}{\tilde{\sigma}}\right) \\ &= \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\} \Phi\left(-\frac{\tilde{\mu} + \tilde{\sigma}^2}{\tilde{\sigma}}\right)\end{aligned}$$

On the other hand, we have

$$\mathbf{P}\left[Y \leq 1\right] = \Phi\left(-\frac{\tilde{\mu}}{\tilde{\sigma}}\right)$$

Then $\mathbf{E}\left[Y \middle| Y \leq 1\right] = \mathbf{E}\left[\mathbf{1}_{\{Y \leq 1\}} Y\right] / \mathbf{P}\left[Y \leq 1\right]$. □

Expected shortfall

Proposition 14 (Expected shortfall)

Let $Y \sim LN(\tilde{\mu}, \tilde{\sigma}^2)$ and $\alpha \in (0,1)$. Then we have:

$$\mathbf{E}\left[Y \middle| Y \leq q_\alpha(Y)\right] = \frac{1}{\alpha} \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\}\ \Phi\left(\Phi^{-1}(\alpha) - \tilde{\sigma}\right)$$

Proof: Using Propositions 10 and 12 we have

$$\begin{aligned}
\mathbf{E}\left[\mathbf{1}_{\{Y \leq q_\alpha(Y)\}} Y\right] &= \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\}\ \Phi\left(\frac{\log q_\alpha(Y) - (\tilde{\mu} + \tilde{\sigma}^2)}{\tilde{\sigma}}\right) \\
&= \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\}\ \Phi\left(\frac{\tilde{\mu} + \Phi^{-1}(\alpha)\tilde{\sigma} - (\tilde{\mu} + \tilde{\sigma}^2)}{\tilde{\sigma}}\right) \\
&= \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\}\ \Phi\left(\frac{\Phi^{-1}(\alpha)\tilde{\sigma} - \tilde{\sigma}^2}{\tilde{\sigma}}\right) \\
&= \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\}\ \Phi\left(\Phi^{-1}(\alpha) - \tilde{\sigma}\right)
\end{aligned}$$

Moreover, we have $\mathbf{P}\left[Y \leq q_\alpha(Y)\right] = \alpha$. □

Back to the Lognormal model

Recall that in the Lognormal model according to Definition 1, we have according to Proposition 1

$$\tilde{\mu} = \log \mathsf{FR}_0 + (\mu - \lambda)t \quad \text{and} \quad \tilde{\sigma}^2 = \sigma^2 t$$

Inserting this into the results of Propositions 12, 13 and 14 yields:

1.) For the quantile:

$$\begin{aligned} q_\alpha(\mathsf{FR}_t) &= \exp\left\{\tilde{\mu} + \Phi^{-1}(\alpha)\,\tilde{\sigma}\right\} \\ &= \exp\left\{\log \mathsf{FR}_0 + (\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\} \\ &= \mathsf{FR}_0\, \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\} \end{aligned}$$

2.) For the conditional expectation:

$$\mathbf{E}\left[\mathsf{FR}_t \middle| \mathsf{FR}_t \leq 1\right] = \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\} \Phi\left(-\frac{\tilde{\mu} + \tilde{\sigma}^2}{\tilde{\sigma}}\right) \Big/ \Phi\left(-\frac{\tilde{\mu}}{\tilde{\sigma}}\right)$$

$$= \exp\left\{\log \mathsf{FR}_0 + (\mu - \lambda)t + \tfrac{1}{2}\sigma^2 t\right\} \times$$

$$\Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t + \sigma^2 t}{\sigma\sqrt{t}}\right) \Big/ \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

$$= \mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \tfrac{1}{2}\sigma^2 t\right\} \times$$

$$\Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t + \sigma^2 t}{\sigma\sqrt{t}}\right) \Big/ \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

$$= \mathbf{E}\left[\mathsf{FR}_t\right] \times$$

$$\Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t + \sigma^2 t}{\sigma\sqrt{t}}\right) \Big/ \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

3.) For the expected shortfall:

$$
\begin{aligned}
\mathbf{E}\left[\mathsf{FR}_t \middle| \mathsf{FR}_t \leq q_\alpha(\mathsf{FR}_t)\right] &= \frac{1}{\alpha} \exp\left\{\tilde{\mu} + \tfrac{1}{2}\tilde{\sigma}^2\right\} \Phi\left(\Phi^{-1}(\alpha) - \tilde{\sigma}\right) \\
&= \frac{1}{\alpha} \exp\left\{\log \mathsf{FR}_0 + (\mu - \lambda)t + \tfrac{1}{2}\sigma^2 t\right\} \Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right) \\
&= \frac{1}{\alpha} \mathsf{FR}_0 \exp\left\{\left(\mu - \lambda + \tfrac{1}{2}\sigma^2\right) t\right\} \Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right) \\
&= \frac{1}{\alpha} \mathbf{E}\left[\mathsf{FR}_t\right] \Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right)
\end{aligned}
$$

Expected Funding Shortfall

We introduce a new risk measure that takes into account the extent of an eventual underfunding as desired:

Definition 6 (Expected Funding Shortfall)

Let $\alpha \in (0,1)$ and $t > 0$. The Expected Funding Shortfall is defined as

$$EFS_{\alpha,t} := 1 - \mathbf{E}\left[FR_t \middle| FR_t \leq q_\alpha(FR_t)\right]$$

Proposition 15 (EFS in the Lognormal model)

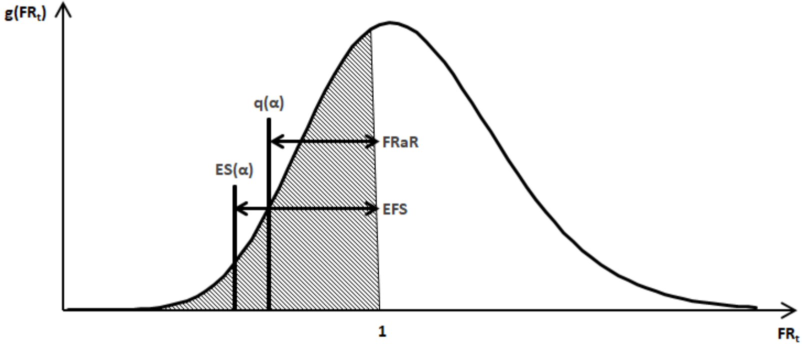
Under the Lognormal model according to Definition 1, the Expected Funding Shortfall is given by

$$\begin{aligned} EFS_{\alpha,t} &= 1 - \tfrac{1}{\alpha}\, FR_0 \exp\left\{ \left(\mu - \lambda + \tfrac{1}{2}\sigma^2\right) t \right\}\, \Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right) \\ &= 1 - \tfrac{1}{\alpha}\, \mathbf{E}\left[FR_t\right]\, \Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right) \end{aligned}$$

Proof: See calculations above. □

Illustration: FRaR and EFS

Graphically, the various measures can be interpreted as follows:



We might as well work directly with the quantile q_α or the Expected Shortfall ES_α. But in practice, one is more interested in the deficit that might have to be made up in order to reach fully-funded status (i.e. $\text{FR}_t = 1$) again. Therefore, it is preferrable to work with the difference to 1.

This is the expected difference to the funding ratio of 1 (i.e. the expected funding shortfall) that is present if the α-event materializes. It is a measure for the amount of money that must be put up in order to bring the institution to fully funded status.

Note that $\mathsf{EFS}_{\alpha,t}$ is here expressed in terms of funding ratio. It can, however, be easily translated into monetary terms.

When doing stochastic simulation, this measure is easy to calculate either in monetary terms or in terms of funding ratio.

For the levels of $\mathsf{EFS}_{\alpha,t}$, we have:

- High value: BAD
- Low value: GOOD

For the changes of $\mathsf{EFS}_{\alpha,t}$, we have:

- Increase: BAD
- Decrease: GOOD

Also with $\mathsf{EFS}_{\alpha,t}$, we explore the sensitivities:

$$\frac{\partial\,\mathsf{EFS}_{\alpha,t}}{\partial\,\mathsf{FR}_0} = -\tfrac{1}{\alpha}\exp\left\{\left(\mu-\lambda+\tfrac{1}{2}\sigma^2\right)t\right\}\,\Phi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right) < 0$$

$$\frac{\partial\,\mathsf{EFS}_{\alpha,t}}{\partial\,\mu} = -\tfrac{t}{\alpha}\,\mathsf{FR}_0\exp\left\{\left(\mu-\lambda+\tfrac{1}{2}\sigma^2\right)t\right\}\,\Phi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right) < 0$$

$$\frac{\partial\,\mathsf{EFS}_{\alpha,t}}{\partial\,\lambda} = \tfrac{t}{\alpha}\,\mathsf{FR}_0\exp\left\{\left(\mu-\lambda+\tfrac{1}{2}\sigma^2\right)t\right\}\,\Phi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right) > 0$$

$$\frac{\partial\,\mathsf{EFS}_{\alpha,t}}{\partial\,\sigma} = \tfrac{1}{\alpha}\,\mathsf{FR}_0\exp\left\{\left(\mu-\lambda+\tfrac{1}{2}\sigma^2\right)t\right\}\times \left[\sqrt{t}\,\varphi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right) - t\,\sigma\,\Phi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right)\right]$$

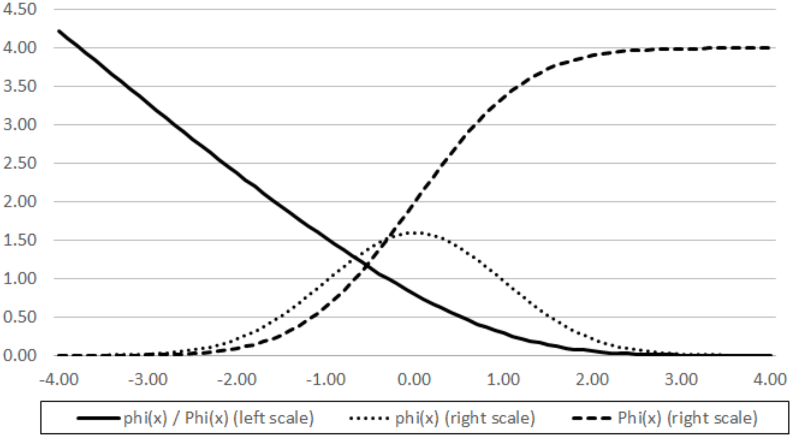
The first three sensitivities are as one would expect. For the fourth expression, the term in the square brackets will be positive, unless

$$\frac{\varphi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right)}{\Phi\left(\Phi^{-1}(\alpha)-\sigma\sqrt{t}\right)} < \sigma\sqrt{t}$$

which is difficult to attain for realistic values of α. This is different from the probability of underfunding, where we can have a negative sensitivity to σ under fairly realistic conditions. That is, $\mathsf{EFS}_{\alpha,t}$ is more conservative in this respect.

Normal density and CDF

Typical values for α are e.g. 1%, 5% or 10%. In these cases $\Phi^{-1}(\alpha)$ is at -2.3, -1.6 and -1.3, respectively. Typical values for σ are (well) below 10%.



Funding Ratio at Risk

In the same manner as before, we can introduce another risk measure, this time inspired by Value at Risk:

Definition 7 (Funding Ratio at Risk)

Let $\alpha \in (0,1)$ and $t > 0$. The Funding Ratio at Risk is defined as

$$\mathit{FRaR}_{\alpha,t} := 1 - q_\alpha(\mathit{FR}_t)$$

Proposition 16 (FRaR in the Lognormal model)

Under the Lognormal model according to Definition 1, the Funding Ratio at Risk is given by

$$\mathit{FRaR}_{\alpha,t} = 1 - \mathit{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\}$$

Proof: See calculations above. □

For the sensitivities, we obtain:

$$\frac{\partial}{\partial \mathsf{FR}_0} \mathsf{FRaR}_{\alpha,t} = -\exp\left\{(\mu-\lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\} < 0$$

$$\frac{\partial}{\partial \mu} \mathsf{FRaR}_{\alpha,t} = -t\,\mathsf{FR}_0 \exp\left\{(\mu-\lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\} < 0$$

$$\frac{\partial}{\partial \lambda} \mathsf{FRaR}_{\alpha,t} = t\,\mathsf{FR}_0 \exp\left\{(\mu-\lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\} > 0$$

$$\frac{\partial}{\partial \sigma} \mathsf{FRaR}_{\alpha,t} = -\,\Phi^{-1}(\alpha)\sqrt{t}\,\mathsf{FR}_0 \exp\left\{(\mu-\lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\} > 0 \text{ if } \alpha < 0.5$$

This is as one would expect. Given that we are in the left tail, sensible values for α are e.g. 1%, 5% or 10%, certainly not more. Also here, there is no realistic situation where the sensitivity to σ is negative, even if the institution is underfunded.

Concluding remarks

In practice, the probability of underfunding ψ_t is the predominant risk measure, and there is no fundamental objection against this. However, ψ_t does only provide limited information:

- ψ_t just gives the probability of something undesirable, i.e. an underfunding, happening,
- but it bears no information on the extent of the underfunding.

The alternative risk measures $\mathrm{EFS}_{\alpha,t}$ and $\mathrm{FRaR}_{\alpha,t}$ also bear information on the extent of the underfunding. In the case of a turnaround, they can be used as a proxy for the turnaround costs.

This is very useful when weighting the cost of measures for risk reduction against their effect. See the subsequent chapters with the ALM studies for more on this.

Moreover, at least at first sight, $\mathrm{EFS}_{\alpha,t}$ and $\mathrm{FRaR}_{\alpha,t}$ seem to be more conservative when it comes to the treatment of (short-term) investment risk. We will also explore this further in the subsequent chapters.

Bibliography I

[1] BILLINGSLEY, P.
Probability and Measure, third ed.
John Wiley & Sons, Chchester, 1995.

[2] GRIMMETT, G., AND STIRZAKER, D.
Probability and Random Processes, second ed.
Clarendon Press, Oxford, 1992.

[3] KOLLER, M.
Stochastic models in life insurance.
Lecture notes, ETH, Zürich, 2015.

[4] WÜTHRICH, M.
Non-life insurance mathematics and statistics.
Lecture notes, ETH Zürich, Zürich, 2015.

Financial Risk Management in
Social and Pension Insurance

Chapter VI: ALM Study 1
Dealing with the risk / return profile

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 17, 2020

Table of Contents

1. Problem statement

Recall: generic model framework

According to Chapter IV, the generic representation of the problem is

$$\left.\begin{aligned} A_t &= A_{t-1}(1+R_t) + C_t \\ L_t &= L_{t-1}(1+\lambda_t) + C_t \end{aligned}\right\} \quad \text{for} \quad t \in \{1, \ldots, T\}$$

and the relevant influence factors to be reconciled with one another are

Institution	**Financial markets**
Required return	Expected return
$\lambda_t + (\mathsf{FR}_{t-1} - 1)\frac{C_t}{A_{t-1}}$	μ_t
Risk-taking capability	Investment risk
FR_0	σ_t^2

This is a first study of the interplay of these factors, specifically

- interplay between expected return and investment risk
- given required return and risk-taking capability

These considerations can be done in the general case by stochastic simulation. Here, however, we will do them analytically within the Lognormal framework.

Recall: Lognormal model

As in Definition 1 and Proposition 1 of Chapter V, we assume that the institution is in equilibrium (i.e. $C_t \equiv 0$), and we let

$$\log \mathsf{FR}_t - \log \mathsf{FR}_{t-1} = \mu - \lambda + \varepsilon_t \qquad \text{where} \quad \varepsilon_t \sim \text{iid}\,\mathcal{N}\left(0, \sigma^2\right)$$

For given FR_0, this means that $\mathsf{FR}_t \sim LN\left(\log \mathsf{FR}_0 + (\mu - \lambda)t, \sigma^2 t\right)$. This gives rise to the long-term risk measures:

Probability of underfunding (Chapter V, Proposition 8):

$$\psi_t\left(\mathsf{FR}_0, \lambda, \mu, \sigma\right) = \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

Expected Funding Shortfall (Chapter V, Proposition 12):

$$\mathsf{EFS}_{\alpha,t}\left(\mathsf{FR}_0, \lambda, \mu, \sigma\right) = 1 - \tfrac{1}{\alpha}\,\mathsf{FR}_0 \exp\left\{\left(\mu - \lambda + \tfrac{1}{2}\sigma^2\right)t\right\}\Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right)$$

Funding Ratio at Risk (Chapter V, Proposition 13):

$$\mathsf{FRaR}_{\alpha,t}\left(\mathsf{FR}_0, \lambda, \mu, \sigma\right) = 1 - \mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma\sqrt{t}\right\}$$

The goal is to optimize the values of these long-term risk measures under certain assumptions.

Assumptions

We assume throughout this chapter that the required return λ and the risk-taking capability FR_0 are given.

- This corresponds to a situation where the discount rate δ is fixed and given and cannot be changed by the institution.
- Then, also FR_0 directly results from this.

The variables are thus the expected return μ and the short-term investment risk σ^2. In principle, long-term risk is optimized by selecting a high μ and a low σ^2; c.f. the sensitivity analysis in Chapter V.

The problem is that expected return and investment risk are related, i.e. if we select a higher expected return, then we will invariably incur more investment risk.

Therefore, we assume that there is a functional relationship between expected return μ and investment risk σ^2, or σ for the purpose: $\sigma = \sigma(\mu)$.

This relationship $\sigma = \sigma(\mu)$ is called Risk / Return Profile.

Call from practice

In practice, one considers the standard deviation σ of the investment returns rather than the variance σ^2. This is because the standard deviation is on the same scale as the returns themselves and their expectation. Therefore, it lends itself to an easier interpretation, e.g.

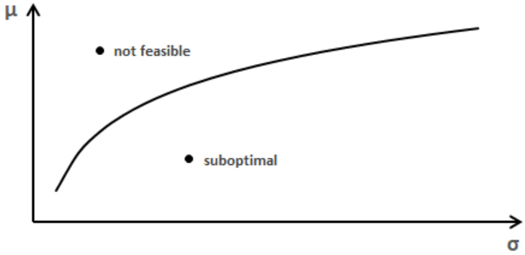
- A standard deviation of 5% means that the average absolute deviation of the returns from their expectation is 5%.
- The resulting variance of 0.0025 has no intuitive interpretation.

2. Risk / return profile

Fundamental situation

It is a fundamental matter of fact that a higher expected return μ comes along with a higher investment risk σ:

- This is theoretically well underpinned, e.g. by the Capital Asset Pricing Model (CAPM); see [1].
- And it is also empirically confirmed; see the subsequent considerations in this course or e.g. [2].



Note that we use σ for the risk, and not σ^2.

We are mainly interested in the points (risk / return combinations) on the line:

- Points below the line are suboptimal, i.e. less return for the same risk or more risk for the same return.
- Points above the line are not feasible.

As risk becomes higher, the extra return obtained becomes lower. This reflects the economic principle of decreasing marginal utility, and it also makes sense empirically.

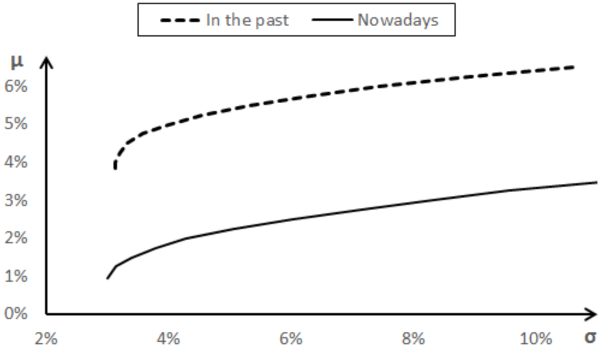
We can safely assume that the function $\mu(\sigma)$ is twice continuously differentiable, so that we can state

$$\mu'(\sigma) > 0 \qquad \text{(increasing return)}$$
$$\mu''(\sigma) \leq 0 \qquad \text{(decreasing marginal return)}$$

For the moment, these facts should be taken as given. We will see in subsequent chapters, when we look at the construction of investment portfolios, why they are well justified.

Risk / return profiles in practice

Risk / return profiles feasible for typical Swiss pension funds; as it used to be around 2005 and as it is nowadays:



Parallel to the decrease in interest rates (see Chapter II), there was a dramatic decline in return expectations over the past ten years. That is:

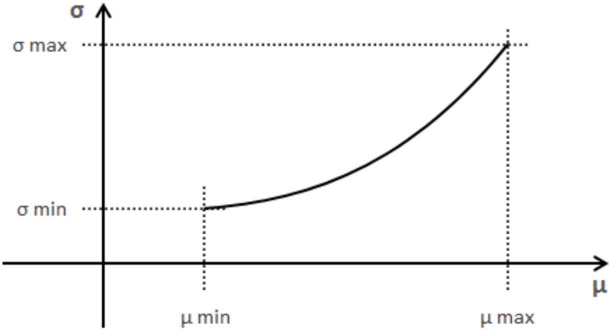
- (Much) more risk for the same return as before (if feasible at all).
- (Much) less return for the same risk as before.

Formalization of risk / return profile

In practice, our policy variable is the expected return μ:

- Required returns are clearly defined, and we must set μ accordingly; e.g. in the Lognormal model this means $\mu \geq \lambda$.
- The requirements for the investment risk σ are generally less clearly defined. This is more of a consequence.

Therefore, we work with a risk / return profile of the form $\sigma = \sigma(\mu)$, which creates the following situation:



We have $\mu \in [\mu_{\min}, \mu_{\max}]$ for some minimum attainable return $\mu_{\min}$ and some maximum attainable return $\mu_{\max} > \mu_{\min}$. Nowadays, we must admit the situation where $\mu_{\min} < 0$.

Consequently, there is a minimum risk $\sigma_{\min} = \sigma(\mu_{\min})$ and a maximum risk $\sigma_{\max} = \sigma(\mu_{\max})$ such that $\sigma \in [\sigma_{\min}, \sigma_{\max}]$. Note that still nowadays $\sigma \geq 0$.

Hence, we let $\sigma(\mu)$ be a twice continuously differentiable function with

$$\left.\begin{array}{r}\sigma(\mu) \geq 0 \\ \sigma'(\mu) > 0 \\ \sigma''(\mu) > 0\end{array}\right\} \quad \text{for all } \mu \in [\mu_{\min}, \mu_{\max}]$$

That is, $\sigma(\mu)$ is a convex function on a compact support.

<u>Example</u>: Let $\sigma(\mu) = \alpha(\mu - \mu_{\min})^2$ for $\mu \geq \mu_{\min}$ and $\alpha > 0$. Then, we have

$$\begin{aligned}\sigma(\mu) &\geq 0 \\ \sigma'(\mu) &= 2\alpha(\mu - \mu_{\min}) > 0 \ \text{ for } \mu > \mu_{\min} \\ \sigma''(\mu) &= 2\alpha > 0\end{aligned}$$

3. Optimizing the probability of underfunding

Setup

Let the Lognormal model according to Definition 1 of Chapter V hold.

Assume that we have a risk / return profile $\sigma(\mu)$ defined on $\mu \in [\mu_{\min}, \mu_{\max}]$ as introduced above.

Assume that FR_0 and λ are given and fixed. Assume moreover that $\lambda \in (\mu_{\min}, \mu_{\max})$, i.e. it is actually possible to attain the required return. (The situation where λ needs to be adapted will be considered in the next chapter.)

Note that we assume the net profit condition to be respected, i.e. $\mu \geq \lambda$. In general, this is indispensable for assuring a sustainable funding.

The question is now: Does it make sense to increase μ beyond λ, i.e. take more risk than absolutely necessary in order to achieve a higher expected return than the minimum. The criterion is long-term risk, e.g. the probability of underfunding. If the latter decreases as we increase μ, then it is worthwhile.

In other words: Can an increase in short-term risk (i.e. σ) lead to a decrease in long-term risk (e.g. ψ_t)? And if so, under what circumstances?

Graphical exploration

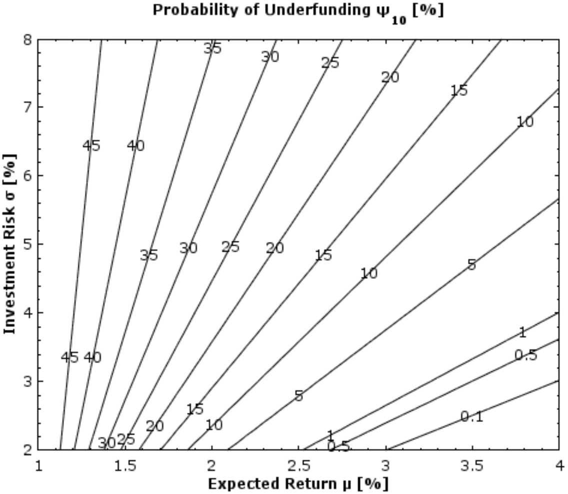
Given λ and FR_0 as parameters and some fixed t, ψ_t becomes a mapping

$$\begin{aligned}\psi_{t,\lambda,\mathsf{FR}_0} : \mathbb{R} \times \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \\ (\mu, \sigma) &\mapsto \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)\end{aligned}$$

We can simply calculate this for each pair $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$, whether feasible or not, and we can plot the results as a risk map in μ/σ-space, e.g. as a line plot.

In a line plot, the lines connect all those combinations (μ, σ) that lead to the same value of ψ_t. That is, the lines in the plot are iso-ψ_t-lines similar to height curves in a topographic map or to isobars in weather chart.

Example: map of ψ_t given $\text{FR}_0 = 110\%$, $\lambda = 2\%$ and $t = 10$ in μ/σ-space:

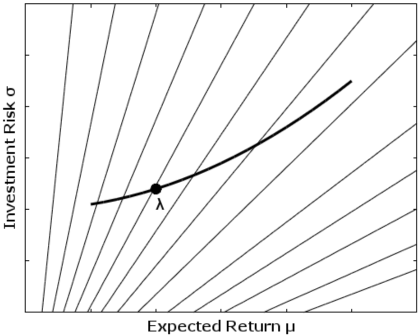


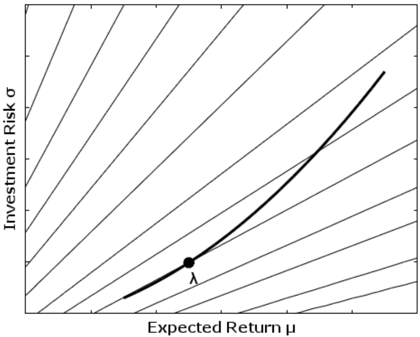
ψ_t becomes lower for higher μ and for lower σ as one would expect from the sensitivities computed in Chapter V.

Call from practice

The good thing about this representation is that in more complicated settings, where there is no analytical tractability, one can always obtain this risk map by stochastic simulation. It requires some computing power, since the simulation of the probability of underfunding must be repeated for a large number of combinations of μ and σ. Moreover, it also requires some brain power from the analyst to get the numerics under control. But it works universally.

Now, we can simply superimpose our risk / return profile of feasible combinations of μ and σ. There are two basic situations:





$\sigma(\mu)$ grows slowly w.r.t. ψ_t, such that we obtain a lower ψ_t by increasing μ beyond λ.

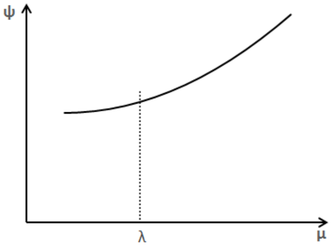
$\sigma(\mu)$ grows too quickly w.r.t. ψ_t, such that we obtain a higher ψ_t by increasing μ beyond λ.

Analytical evaluation

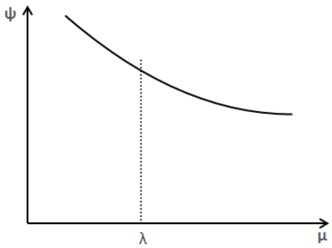
We simply take the formula for ψ_t and insert $\sigma(\mu)$ instead of σ:

$$\psi_{t,\mathsf{FR}_0,\lambda}(\mu) = \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma(\mu)\sqrt{t}}\right)$$

and evaluate over $\mu \in [\mu_{\min}\,,\,\mu_{\max}]$. This representation can also be obtained in a general stochastic simulation setting. We may obtain, for instance:



ψ_t increases as μ increases; not worthwhile to select $\mu > \lambda$.



ψ_t decreases as μ increases; worthwhile to select $\mu > \lambda$ for as long as ψ_t decreases.

Let us now study the derivative $\frac{d}{d\mu}\,\psi_{t,\mathsf{FR}_0,\lambda}(\mu)$; we have:

$$\frac{d}{d\mu}\;\psi_{t,\mathsf{FR}_0,\lambda}(\mu) = \varphi\left(-\frac{\log \mathsf{FR}_0 + (\mu-\lambda)t}{\sigma(\mu)\sqrt{t}}\right)\;(-f'(\mu))$$

where

$$f(\mu) = \frac{\log \mathsf{FR}_0 + (\mu-\lambda)t}{\sigma(\mu)\sqrt{t}}$$

Applying the quotient rule to $f(\mu)$ and rearranging, we obtain

$$\frac{d}{d\mu}\;\psi_{t,\mathsf{FR}_0,\lambda}(\mu) = \varphi\left(-\frac{\log \mathsf{FR}_0 + (\mu-\lambda)\,t}{\sigma(\mu)\sqrt{t}}\right)\;\times$$
$$\frac{1}{\sigma^2(\mu)\,\sqrt{t}}\left[\Big(\log \mathsf{FR}_0 + (\mu-\lambda)\,t\Big)\sigma'(\mu) - \sigma(\mu)\,t\right]$$

We are interested in the situation where $\frac{d}{d\mu}\,\psi_{t,\mathsf{FR}_0,\lambda}(\mu) < 0$. Since both $\varphi(\,\cdot\,)$ and $\sigma(\mu) > 0$, this occurs if and only if the term in square brackets is negative, i.e. if

$$(\log \mathsf{FR}_0 + (\mu-\lambda)\,t)\,\sigma'(\mu)\;<\;\sigma(\mu)\,t$$

We note that the condition is more likely to be satisfied if the institution is underfunded (since then $\log \mathsf{FR}_0 < 0$).

At the point $\mu = \lambda$ the condition boils down to

$$\log \mathsf{FR}_0 \, \sigma'(\lambda) \; < \; \sigma(\lambda) \, t$$

Since $\sigma'(\lambda) > 0$, this condition is always satisfied if the institution is underfunded, i.e. if $\log \mathsf{FR}_0 < 0$. On the other hand, for $\log \mathsf{FR}_0 > 0$, we have

$$\frac{\sigma'(\lambda)}{\sigma(\lambda)} \overset{!}{<} \frac{t}{\log \mathsf{FR}_0}$$

That is, if the relative increase in investment risk is below the bound given by the quotient of t and $\log \mathsf{FR}_0$, then it is worthwhile to take more investment risk and aspire for a higher return than the minimum required.

- The longer the time horizon, the less restrictive the criterion.
- The higher the initial funding ratio, the more restrictive the criterion.

This appears plausible. We should, however, still remain somewhat careful about these findings as long as we have not investigated a risk measure that takes into account the cost of a possible underfunding.

4. Optimizing alternative risk measures

Overview

The same risk / return consideration as in the previous section can also be done based on the Expected Funding Shortfall and on the Funding Ratio at Risk.

As before, let λ and FR_0 be given, and let $\sigma = \sigma(\mu)$ denote the risk / return profile. Then, we have according to Proposition 12 and 13 of Chapter V:

$$\mathsf{EFS}_{\alpha,t}(\mu) = 1 - \tfrac{1}{\alpha}\,\mathsf{FR}_0 \exp\Big\{ \big(\mu - \lambda + \tfrac{1}{2}\sigma^2(\mu)\big)\, t \Big\}\, \Phi\left(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\right)$$

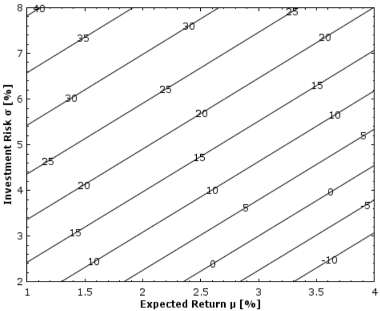
$$\mathsf{FRaR}_{\alpha,t}(\mu) = 1 - \mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma(\mu)\sqrt{t}\right\}$$

The added value of these evaluations is that they also take into account the extent of a possible underfunding, not just the probability of it happening.

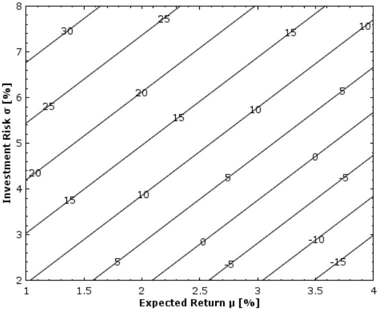
Graphical evaluation

The graphical evaluations can be done in the same manner as for the probability of underfunding. Numbers are in % of funding ratio.

Expected Funding Shortfall



Funding Ratio at Risk



All else being equal, the Funding Ratio at Risk is less conservative than the Expected Funding Shortfall. Otherwise, the basic pattern is not fundamentally different from the probability of underfunding. To explore differences, we have to revert to analytical evaluations.

Analytical evaluation

Based on the equation for $\mathsf{EFS}_{\alpha,t}$ with $\sigma = \sigma(\mu)$ as above, i.e.

$$\mathsf{EFS}_{\alpha,t}(\mu) = 1 - \tfrac{1}{\alpha}\,\mathsf{FR}_0 \exp\left\{ \left(\mu - \lambda + \tfrac{1}{2}\sigma^2(\mu)\right) t \right\} \Phi\left(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\right)$$

we obtain

$$\frac{d}{d\mu}\,\mathsf{EFS}_{\alpha,t}(\mu) = -\frac{1}{\alpha}\,\mathsf{FR}_0\,\left(\frac{d}{d\mu}\,g(\mu)\,h(\mu)\right)$$

where we have

$$g(\mu) = \exp\left\{ \left(\mu - \lambda + \tfrac{1}{2}\sigma^2(\mu)\right) t \right\}$$
$$h(\mu) = \Phi\left(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\right)$$

Observing that

$$g'(\mu) = \exp\left\{ \left(\mu - \lambda + \tfrac{1}{2}\sigma^2(\mu)\right) t \right\}\left(t + \sigma(\mu)\,\sigma'(\mu)\,t\right)$$
$$h'(\mu) = -\sigma'(\mu)\sqrt{t}\;\;\varphi\left(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\right)$$

and ... ./.

./. ... applying the product rule, we obtain

$$\frac{d}{d\mu}\,\mathsf{EFS}_{\alpha,t}(\mu) = \tfrac{1}{\alpha}\,\mathsf{FR}_0 \exp\Big\{ \big(\mu - \lambda + \tfrac{1}{2}\sigma^2(\mu)\big)\, t \Big\} \times \Big[\sqrt{t}\,\sigma'(\mu)\,\varphi\Big(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\Big) - t\,(1+\sigma(\mu)\,\sigma'(\mu))\,\Phi\Big(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\Big)\Big]$$

We are again interested in the situation where $\frac{d}{d\mu}\,\mathsf{EFS}_{\alpha,t}(\mu) < 0$, i.e. where it is worthwhile to increase expected return μ and hence investment risk $\sigma(\mu)$ in order to obtain a lower long-term risk. This is the case if and only if the expression in the square brackets is negative, i.e.

$$t\,(1+\sigma(\mu)\,\sigma'(\mu))\,\Phi\Big(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\Big) \;>\; \sqrt{t}\,\sigma'(\mu)\,\varphi\Big(\Phi^{-1}(\alpha) - \sigma(\mu)\sqrt{t}\Big)$$

We observe that $\sigma(\mu)$, $\sigma'(\mu)$, $\varphi(\,\cdot\,)$ and $\Phi(\,\cdot\,)$ are all positive.

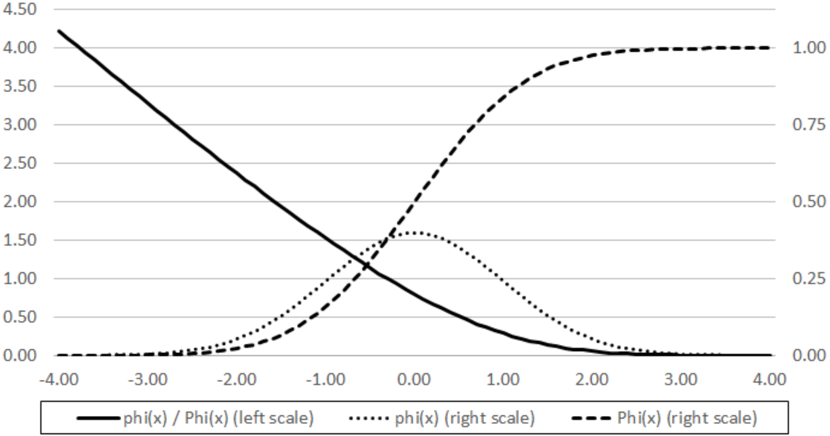
Therefore, we can rearrange the inequality to obtain

$$\sigma'(\mu) \;<\; \frac{1}{\frac{1}{\sqrt{t}} \cdot \frac{\varphi\left(\Phi^{-1}(\alpha)-\sigma(\mu)\sqrt{t}\right)}{\Phi\left(\Phi^{-1}(\alpha)-\sigma(\mu)\sqrt{t}\right)} - \sigma(\mu)}$$

Again, we obtain an upper bound for $\sigma'(\mu)$. That is, if the additional short-term risk $\sigma(\mu)$ that we have to take for an additional unit of expected return is sufficiently low, this reduces long-term risk as expressed by $\mathsf{EFS}_{\alpha,t}$. Moreover, we can observe:

- The bound does not depend on FR_0 here. This seems logical, since FR_0 only appears as a factor in front of everything else for this risk measure.
- However, the bound depends on the level of safety α. If α becomes lower (higher level of safety), then $\varphi(\,\cdot\,)/\Phi(\,\cdot\,)$ becomes higher (see next page) and the bound becomes more restrictive.
- The longer the time horizon, the less restrictive the bound. This also seems logical and is congruent with the probability of underfunding.
- The higher the risk $\sigma(\mu)$ that we already have, the more restrictive the bound becomes. This also appears sensible.

Normal density vs. CDF



Funding Ratio at Risk

Based on the equation for $\mathsf{FRaR}_{\alpha,t}$ and with $\sigma = \sigma(\mu)$ as above, we have

$$\mathsf{FRaR}_{\alpha,t}(\mu) = 1 - \mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma(\mu)\sqrt{t}\right\}$$

For the first derivative w.r.t. μ, we obtain

$$\frac{d}{d\mu}\,\mathsf{FRaR}_{\alpha,t}(\mu) = -\mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\,\sigma(\mu)\sqrt{t}\right\}\left(t + \Phi^{-1}(\alpha)\,\sigma'(\mu)\sqrt{t}\right)$$

For this to be negative, we must have

$$t + \Phi^{-1}(\alpha)\,\sigma'(\mu)\sqrt{t} > 0$$

For reasonable values of α, i.e. $\alpha \ll 0.5$, we have $\Phi^{-1}(\alpha) < 0$, and therefore

$$\sigma'(\mu) \;<\; -\frac{\sqrt{t}}{\Phi^{-1}(\alpha)}$$

That is, we again have an upper bound for the marginal investment risk $\sigma'(\mu)$.

As with the Expected Funding Shortfall, we have the following properties:

- The bound does not depend on FR_0.
- It depends, however, on the level of safety α. The higher the level of safety, i.e. the lower α, the more restrictive the bound becomes.
- The longer the time horizon, the less restrictive the bound becomes.

Contrary to the Expected Funding Shortfall, we have:

- The bound does not depend on the absolute level of investment risk $\sigma(\mu)$ already in force.

Although the expression for the Funding Ratio at Risk is very handy, the Expected Funding Shortfall may be the more comprehensive and also the more prudent decision criterion.

Concluding remarks

In this chapter, we have basically studied the interplay between short-term risk and long-term risk:

- Short-term: single-period investment risk as expressed by $\sigma(\mu)$.
- Long-term: multi-period risk to assets and liabilities as expressed by the measures ψ_t, $\mathsf{EFS}_{\alpha,t}$ or $\mathsf{FRaR}_{\alpha,t}$.

We have seen that, in certain situations, the short-term risk and the long-term risk may be conflicting dimensions, i.e.

- It may make sense to increase short-term investment risk in order to decrease the long-term risk of underfunding.

The condition for this to be the case is generally when the marginal short-term risk is below certain bounds that depend on initial funding ratio, time horizon or desired level of safety.

Bibliography I

[1] Brealy, R., and Myers, S.
Priciples of Corporate Finance, sixth ed.
Irwin McGraw–Hill, Boston, 2000.

[2] Campbell, J., Lo, A., and MacKinlay, A.
The Econometrics of Financial Markets.
Princeton University Press, Princeton, 1997.

Financial Risk Management in Social and Pension Insurance

Chapter VII: ALM Study 2 Incorporating required return

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

September 17, 2020

Table of Contents

1. Problem statement

Recall: generic model framework

According to Chapter IV, the generic representation of the problem is

$$\left.\begin{aligned} A_t &= A_{t-1}(1+R_t)+C_t \\ L_t &= L_{t-1}(1+\lambda_t)+C_t \end{aligned}\right\}$$

and the relevant influence factors to be reconciled with one another are

Institution	**Financial markets**
Required return	Expected return
$\lambda_t + (\mathsf{FR}_{t-1}-1)\frac{C_t}{A_{t-1}}$	μ_t
Risk-taking capability	Investment risk
FR_0	σ_t^2

In this study, we consider the full interplay between assets and liabilities, i.e. liability growth rate λ and initial funding ratio FR_0 are not anymore given constants.

These considerations can be done in the general case by stochastic simulation. Here, however, we will do them analytically within the Lognormal framework.

Recall: Lognormal model

As in Definition 1 of Chapter V, we assume that the institution is in equilibrium (i.e. $C_t = 0$), and we let

$$\log \mathsf{FR}_t - \log \mathsf{FR}_{t-1} = \mu - \lambda + \varepsilon_t \qquad \text{where} \quad \varepsilon_t \sim \text{iid}\, \mathcal{N}\left(0, \sigma^2\right)$$

For given FR_0, this means that $\mathsf{FR}_t \sim LN\left(\log \mathsf{FR}_0 + (\mu - \lambda)t, \sigma^2 t\right)$. This gives rise to the long-term risk measures:

Probability of underfunding:

$$\psi_t = \Phi\left(-\frac{\log \mathsf{FR}_0 + (\mu - \lambda)t}{\sigma\sqrt{t}}\right)$$

Expected Funding Shortfall:

$$\mathsf{EFS}_{\alpha,t} = 1 - \tfrac{1}{\alpha}\, \mathsf{FR}_0 \exp\left\{\left(\mu - \lambda + \tfrac{1}{2}\sigma^2\right) t\right\} \Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right)$$

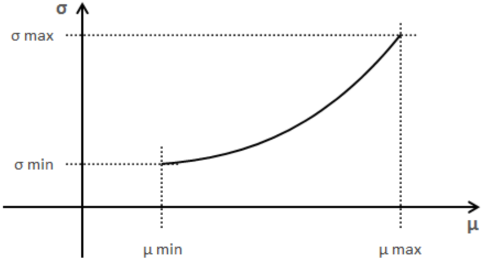
Funding Ratio at Risk:

$$\mathsf{FRaR}_{\alpha,t} = 1 - \mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\, \sigma\sqrt{t}\right\}$$

The goal is to optimize the values of these long-term risk measures under certain assumptions as given below.

Recall: risk / return profile

In the first ALM study in Chapter VI, we have introduced the relationship between the expected return μ and the short-term investment risk σ, i.e. $\sigma = \sigma(\mu)$:



Specifically, we assumed that $\sigma(\mu)$ is twice continuously differentiable with

$$\left.\begin{array}{c} \sigma(\mu) > 0 \\ \sigma'(\mu) > 0 \\ \sigma''(\mu) > 0 \end{array}\right\} \quad \text{for all } \mu \in [\mu_{\min},\, \mu_{\max}]$$

We will maintain this assumption throughout this second study.

The problem

In the previous study, we have assumed that $\lambda \in (\mu_{\min}, \mu_{\max})$, i.e. the required return is within the range of feasible returns.

We have also (tacitly) assumed that the short-term risk σ as well as the long-term risk as expressed by ψ_t, $\mathrm{EFS}_{\alpha,t}$ or $\mathrm{FRaR}_{\alpha,t}$ are acceptable for the institution.

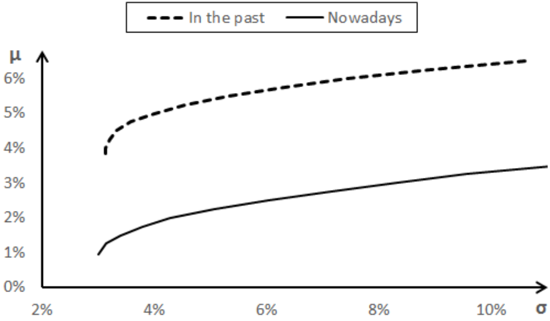
Now, we assume that this is not the case anymore:

- Either λ is simply unfeasible, i.e. $\lambda > \mu_{\max}$.
- Or, at least, the long-term risk caused by letting $\mu = \lambda$ is deemed unacceptably high for the institution.

That is, we must now adapt λ in such a manner that a feasible and acceptable solution becomes possible.

Relevance of the problem

At the beginning of this century, it was commonplace for Swiss pension funds to have discount rates $\delta \geq 4\%$ for their liabilities, giving rise to required returns $\lambda \geq 4\%$. With the then-prevailing risk / return profiles, this was fairly reasonable.



However, since then, risk / return profiles have deteriorated dramatically. Formerly feasible and acceptable required returns are not sensible anymore.

Therefore, reducing discount rates has become a standard task for Swiss (and many other) pension funds.

The consequences

Think of the liabilities as the present value of future promised cashflows, c.f. Chapters II and III:

$$L_0(\delta) = \mathsf{PV}_0(\mathbf{C}, \delta) = \sum_{t=1}^{T} \frac{C_t}{(1+\delta)^t}$$

Now, if we reduce δ (and hence λ), this also means that the present value L_0 increases. For instance, using the approximation introduced earlier, we have

$$L_0(\delta + \Delta\delta) - L_0(\delta) \approx -D^L L_0(\delta)(\Delta\delta) + \tfrac{1}{2}\, K^L L_0(\delta)(\Delta\delta)^2$$

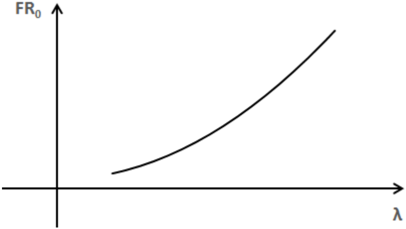
Given that the funding ratio is defined as $\mathsf{FR}_t = A_t/L_t$ this also means that the initial funding ratio FR_0, and hence the risk-taking capability decrease.

That is, reducing the discount rate δ reduces the required return λ and the short-term investment risk σ that must be taken. But it also reduces the risk-taking capability so that the long-term risk may eventually increase. We somehow must find an optimal balance in this situation.

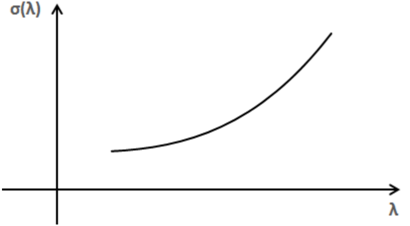
The optimization task

If we put these influences together, we have two different developments:

Institutional side:



Financial markets side:



If the liability growth rate λ is reduced, then also FR_0, i.e. the risk-taking capability, becomes lower, and vice versa.

If the required return λ is reduced, then also the investment risk $\sigma(\lambda)$ is reduced, and vice versa.

In principle, these developments are compatible with one another, but we must explore which one dominates under what conditions.

2. Liability profile

Liability values

We take a closer look at the relationship between the initial value of the liabilities L_0 and the intrinsic growth rate of liabilities λ. This consideration builds on Section 1 of Chapter II. Assume for the moment that L_0 can be expressed as the present value of a stream of cashflows:

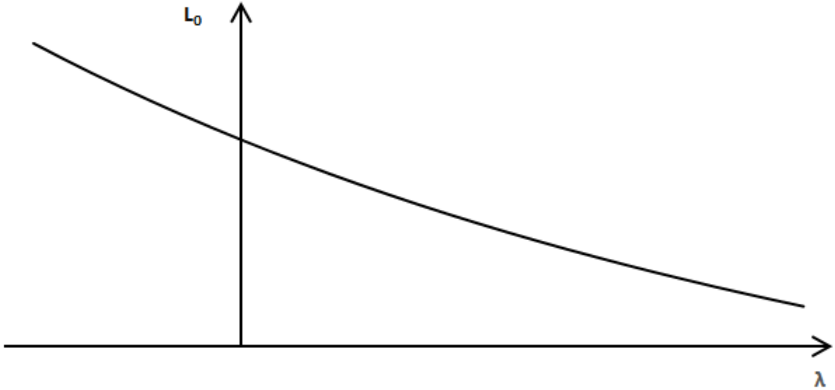
$$L_0(\lambda) = \sum_{t=1}^{T} \frac{C_t}{(1+\lambda)^t}$$

Here, we have inserted the intrinsic growth rate λ instead of the discount rate δ. This is justified by the equality of the two, $\lambda = \delta$, as it was shown in Section 3 of Chapter III. And, according to Propositions 2 and 3 of Chapter II, we have for non-negative cashflows C_t:

$$L_0'(\lambda) = \frac{\partial}{\partial \lambda} \mathsf{PV}_0(\mathbf{C}, \lambda) < 0$$
$$L_0''(\lambda) = \frac{\partial^2}{\partial \lambda^2} \mathsf{PV}_0(\mathbf{C}, \lambda) > 0$$

That is, the value of the liabilities decreases as λ increases, and vice versa.

Illustration: Relationship between initial value of liabilities and intrinsic liability growth rate (= discount rate):



Economic interpretation

Now that we know the relationship between the liability growth rate λ and the required value for the expected investment return, we can also give an economic interpretation to the relationship with the value of the liabilities. For some given stream of promised cashflows, we have:

- If we choose a high value for λ (i.e. a high discount rate δ), then the initial value of the liability L_0 is low. I.e. we have to put up less money at the beginning, but we have to earn a higher investment return μ over time, and we must bear a higher investment risk $\sigma(\mu)$ for doing so.
- If we choose a low value for λ (i.e. a low discount rate δ), we only have to earn a lower investment return μ, and we incur a lower investment risk $\sigma(\mu)$ for doing so. But the initial value L_0 is higher, i.e. we have to put up a higher amount of money at the beginning.

In principle, it is desirable to maximize λ, and hence the contribution from investment returns. But this must be done without compromising long-term security as expressed by ψ_t (or $\text{EFS}_{\alpha.t}$ or $\text{FRaR}_{\alpha.t}$).

This study is basically about finding a suitable balance.

Duration and Convexity

We have already introduced the concepts of Duration and Convexity in Section 1 of Chapter II, and we can apply them here.

Let λ_0 denote the liability growth rate currently in force, i.e. our starting point. Then we have for the Duration D^L and for the Convexity K^L of the liabilities:

$$D^L = -\frac{L_0'(\lambda_0)}{L_0(\lambda_0)} > 0 \qquad \text{and} \qquad K^L = \frac{L_0''(\lambda_0)}{L_0(\lambda_0)} > 0$$

In principle, we can usually calculate $L_0(\lambda)$ directly for any value of λ. However, if necessary, we can use the second-order Taylor approximation

$$L_0(\lambda) \approx L_0(\lambda_0) - D^L L_0(\lambda_0)(\lambda - \lambda_0) + \tfrac{1}{2} K^L L_0(\lambda_0)(\lambda - \lambda_0)^2$$

Up until now, we have motivated these relationships for a situation where $L_0(\lambda)$ is the present value of a stream of cashflows $\mathbf{C}$. In the sequel, we also assume that they hold when the liabilities contain other elements. For typical social insurance liabilities, this is usually the case. However, in the context of unit-linked life insurance products, these assumptions may no longer hold; see e.g. [1]. This is particularly the case if $C_t = C_t(\lambda)$ which we exclude here.

Initial funding ratio

The initial funding ratio is defined as $\mathsf{FR}_0 = A_0/L_0$. In our setup, where λ is deliberately set by the institution, the asset value A_0, i.e. the market value, is independent of λ, so that we have:

$$\mathsf{FR}_0(\lambda) = \frac{A_0}{L_0(\lambda)}$$

Note: In an immunized setup (c.f. Sections 2 and 3 of Chapter IV), both A_0 and L_0 would move with λ which, in turn, is determined by bond yields, and these movements would cancel out if the immunization is good.

In the setup prevailing here, however, we have

$$\mathsf{FR}_0'(\lambda) = \frac{d}{d\lambda}\mathsf{FR}_0(\lambda) = -\frac{A_0}{(L_0(\lambda))^2}L_0'(\lambda) = -\mathsf{FR}_0(\lambda)\,\frac{L_0'(\lambda)}{L_0(\lambda)} > 0$$

The expression is positive since we have $L_0'(\lambda) < 0$. Specifically for $\lambda = \lambda_0$, since $D^L = -L_0'(\lambda_0)/L_0(\lambda_0)$, we have

$$\mathsf{FR}_0'(\lambda_0) = D^L\,\frac{A_0}{L_0(\lambda_0)} = D^L\,\mathsf{FR}_0(\lambda_0)$$

For the second derivative, we obtain (by applying the quotient rule):

$$\begin{aligned}
\mathsf{FR}_0''(\lambda) &= \frac{d}{d\lambda}\,(-A_0)\,\frac{L_0'(\lambda)}{\left(L_0(\lambda)\right)^2} \\
&= (-A_0)\left[\frac{L_0''(\lambda)\left(L_0(\lambda)\right)^2 - L_0'(\lambda)\,2\,L_0(\lambda)\,L_0'(\lambda)}{\left(L_0(\lambda)\right)^4}\right] \\
&= (-A_0)\left[\frac{L_0''(\lambda)}{\left(L_0(\lambda)\right)^2} - 2\,\frac{\left(L_0'(\lambda)\right)^2}{\left(L_0(\lambda)\right)^3}\right] \\
&= \frac{A_0}{L_0(\lambda)}\left[2\left(\frac{L_0'(\lambda)}{L_0(\lambda)}\right)^2 - \frac{L_0''(\lambda)}{L_0(\lambda)}\right] \\
&= \mathsf{FR}_0(\lambda)\left[2\left(\frac{L_0'(\lambda)}{L_0(\lambda)}\right)^2 - \frac{L_0''(\lambda)}{L_0(\lambda)}\right]
\end{aligned}$$

Specifically for $\lambda = \lambda_0$, the latter expression is equivalent to

$$\mathsf{FR}_0''(\lambda_0) = \mathsf{FR}_0(\lambda_0) \left[2 \left(D^L \right)^2 - K^L \right]$$

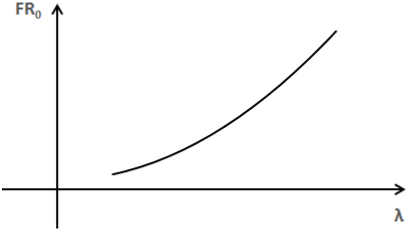
The second derivative $\mathsf{FR}_0''(\lambda)$ may thus be positive or negative, depending on the sign of the expression in the square brackets. Specifically, we have $\mathsf{FR}_0''(\lambda) > 0$ if

$$\frac{L_0''(\lambda)}{L_0(\lambda)} < 2 \left(\frac{L_0'(\lambda)}{L_0(\lambda)} \right)^2 \quad \text{or, at } \lambda = \lambda_0\text{:} \quad K^L < 2 \left(D^L \right)^2$$

That is, the convexity of the liabilities should be bounded with respect to their duration. We refrain from evaluating this condition analytically, but we observe:

- For realistic cashflow patterns, we will usually have $\mathsf{FR}_0''(\lambda) > 0$.
- $\mathsf{FR}_0''(\lambda) < 0$ is mainly achieved by degenerate cashflow patterns, e.g. with high cashflows at the beginning and at the end, with nothing in-between.

That is, in most practical settings, the second derivative $\mathsf{FR}_0''(\lambda)$ will usually be non-negative. Therefore, the funding ratio as a function of the liability growth rate has the following generic (convex) shape:



That is, if we increase the value of λ, then also FR_0, and thus also the risk-taking capability, increases (and even the marginal risk-taking capability increases). This is, in principle, compatible with the fact that we have to take more investment risk in order to finance a higher λ.

Finally, we take a brief look at the log-funding ratio. On the one hand, we have

$$\log \mathsf{FR}_0(\lambda) = \log A_0 - \log L_0(\lambda)$$

and hence

$$\frac{d}{d\lambda} \log \mathsf{FR}_0(\lambda) = -\frac{L_0'(\lambda)}{L_0(\lambda)} \qquad \big(= D^L > 0 \text{ if } \lambda = \lambda_0\big)$$

And on the other side, we have

$$\frac{d}{d\lambda} \log \mathsf{FR}_0(\lambda) = \frac{\mathsf{FR}_0'(\lambda)}{\mathsf{FR}_0(\lambda)}$$

This may prove to be useful.

3. Optimizing the probability of underfunding

Setup

The liability profile is the relationship $\lambda \mapsto \mathsf{FR}_0(\lambda)$ as introduced above. We also have the risk / return profile $\mu \mapsto \sigma(\mu)$ as introduced in the last chapter. Moreover, we equate liability growth rate and expected return, i.e. $\mu = \lambda$. Putting all this together, we obtain for the probability of underfunding

$$\psi_t(\lambda) = \Phi\left(-\frac{\log \mathsf{FR}_0(\lambda)}{\sigma(\lambda)\sqrt{t}}\right)$$

We want to find the value $\lambda \in [\mu_{\min}, \mu_{\max}]$ that minimizes the probability of underfunding. If necessary, λ_0 denotes the current value of λ, which may not be the optimum.

Notice the difference in the setup: In the previous chapter, the institutional side, i.e. λ and FR_0, was given, and only the position on the risk / return profile was up for optimization. Here, also the institutional side is under review.

In practice, we would simply compute $\psi_t(\lambda)$ over the full range $\lambda \in [\mu_{\min}, \mu_{\max}]$ and look up the optimum. This also works well in a general setup where stochastic simulation must be applied. Here, however, we make a few analytical considerations.

General situation

We take the first derivative

$$\frac{d}{d\lambda}\,\psi_t(\lambda) = \frac{d}{d\lambda}\,\Phi\left(-\frac{\log \mathsf{FR}_0(\lambda)}{\sigma(\lambda)\sqrt{t}}\right)$$

$$= -\varphi\left(-\frac{\log \mathsf{FR}_0(\lambda)}{\sigma(\lambda)\sqrt{t}}\right)\left(\frac{g(\lambda)}{h(\lambda)}\right)'$$

where $g(\lambda) = \log \mathsf{FR}_0(\lambda)$ and $h(\lambda) = \sigma(\lambda)\sqrt{t}$. We then have

$$g'(\lambda) = \frac{\mathsf{FR}_0'(\lambda)}{\mathsf{FR}_0(\lambda)} \qquad \big(\text{and, in particular: } g'(\lambda_0) = D^L\big)$$

$$h'(\lambda) = \sigma'(\lambda)\sqrt{t}$$

This yields then

$$\left(\frac{g(\lambda)}{h(\lambda)}\right)' = \frac{\frac{\mathsf{FR}_0'(\lambda)}{\mathsf{FR}_0(\lambda)}\,\sigma(\lambda)\sqrt{t} - \log \mathsf{FR}_0(\lambda)\,\sigma'(\lambda)\sqrt{t}}{\big(\sigma(\lambda)\big)^2\,t}$$

Noting that $\frac{\mathsf{FR}_0'(\lambda)}{\mathsf{FR}_0(\lambda)} = \frac{d}{d\lambda}\log \mathsf{FR}_0(\lambda)$ and putting everything together, we obtain

$$\frac{d}{d\lambda}\,\psi_t(\lambda) = \varphi\left(-\frac{\log \mathsf{FR}_0(\lambda)}{\sigma(\lambda)\sqrt{t}}\right)\frac{\log \mathsf{FR}_0(\lambda)\sigma'(\lambda) - \left(\frac{d}{d\lambda}\log \mathsf{FR}_0(\lambda)\right)\sigma(\lambda)}{\left(\sigma(\lambda)\right)^2\sqrt{t}}$$

We note that $\varphi(\,\cdot\,)$, $\sigma(\lambda)$, $\sigma'(\lambda)$ and $\frac{d}{d\lambda}\log \mathsf{FR}_0(\lambda)$ are all positive. Only $\log \mathsf{FR}_0(\lambda)$ can have both signs.

Again, we are interested in the situation where $\frac{d}{d\lambda}\psi_t(\lambda) < 0$, i.e. where it makes sense to increase λ, and hence μ and hence also the short-term risk $\sigma(\lambda)$ in order to reduce the long-term risk $\psi_t(\lambda)$. This is the case if and only if

$$\log \mathsf{FR}_0(\lambda)\,\sigma'(\lambda) - \left(\frac{d}{d\lambda}\log \mathsf{FR}_0(\lambda)\right)\sigma(\lambda) < 0$$

or, equivalently

$$\log \mathsf{FR}_0(\lambda)\,\sigma'(\lambda) < \left(\frac{d}{d\lambda}\log \mathsf{FR}_0(\lambda)\right)\sigma(\lambda)$$

Now, if $\log \mathsf{FR}_0(\lambda) < 0$, i.e. if $\mathsf{FR}_0(\lambda) < 1$ (equivalent to an underfunding), this is always the case, irrespective of the values of $\sigma(\lambda)$ and $\sigma'(\lambda)$.

That is, if the institution is already underfunded, it makes no sense to reduce the required return λ (i.e. the discount rate δ), because this further exacerbates the current underfunding, and thus increases the risk of future underfunding, at least according to this risk measure. We remain, however, a bit suspicious, because this assertion is completely independent of short-term investment risk $\sigma(\lambda)$.

Assume now that $\log \mathsf{FR}_0(\lambda) > 0$ (i.e. $\mathsf{FR}_0(\lambda) > 1$, overfunding). Then, the above inequality transforms into

$$\frac{\sigma'(\lambda)}{\sigma(\lambda)} < \frac{\frac{d}{d\lambda} \log \mathsf{FR}_0(\lambda)}{\log \mathsf{FR}_0(\lambda)}$$

That is, if the relative increase in investment risk $\frac{\sigma'(\lambda)}{\sigma(\lambda)}$ is lower than the relative increase in (log-) risk-taking capability $\frac{d}{d\lambda} \log \mathsf{FR}_0(\lambda) / \log \mathsf{FR}_0(\lambda)$, then it makes sense to increase λ in order to decrease the long-term risk of underfunding ψ_t.

If, on the other hand, the increase in investment risk outgrows the increase in risk-taking capability, then it makes sense to decrease λ in order to decrease the long-term risk of underfunding ψ_t (provided that $\mathsf{FR}_0(\lambda) > 1$).

This criterion appears entirely plausible and is easy to evaluate. In the situation where $\lambda = \lambda_0$, it boils further down to

$$\frac{\sigma'(\lambda)}{\sigma(\lambda)} < \frac{D^L}{\log \mathsf{FR}_0(\lambda_0)}$$

That is, if this criterion is satisfied, it makes sense to set $\lambda > \lambda_0$. If the contrary is true, it makes sense to set $\lambda < \lambda_0$.

Global considerations

As we have seen, we can choose $\lambda \in [\mu_{\min},\, \mu_{\max}]$. The global optimization task consists of choosing λ^* such that the risk of underfunding is minimized, i.e.

$$\lambda^* = \underset{\lambda \in [\mu_{\min},\, \mu_{\max}]}{\arg\min}\ \psi_t(\lambda)$$

Given that we have functional representations for $\log \mathsf{FR}_0(\lambda)$ and $\sigma(\lambda)$, we can easily calculate

$$\psi_t(\lambda) = \Phi\left(-\frac{\log \mathsf{FR}_0(\lambda)}{\sigma(\lambda)\sqrt{t}}\right) \qquad \text{for} \quad \lambda \in [\mu_{\min},\, \mu_{\max}]$$

and look up the optimum λ^*. In order to further characterize λ^* analytically, we would have to take the second derivative $\frac{d^2}{d\lambda^2}\psi_t(\lambda)$. This is, of course, possible, but the expression is rather unwieldy and does not lead to any intuitive criteria.

Given the easy direct consideration of $\psi_t(\lambda)$, we need not be unhappy about this.

4. Optimizing alternative risk measures

Expected Funding Shortfall

We do the same reasoning as before for the Expected Funding Shortfall according to Definition 5 and Proposition 12 in Chapter V:

$$\mathsf{EFS}_{\alpha,t} = 1 - \frac{1}{\alpha}\,\mathsf{FR}_0 \exp\left\{\left(\mu - \lambda + \tfrac{1}{2}\sigma^2\right) t\right\}\,\Phi\left(\Phi^{-1}(\alpha) - \sigma\sqrt{t}\right)$$

Letting $\mu = \lambda$, $\mathsf{FR}_0 = \mathsf{FR}_0(\lambda)$ and $\sigma = \sigma(\lambda)$, we have

$$\mathsf{EFS}_{\alpha,t}(\lambda) = 1 - \frac{1}{\alpha}\,\mathsf{FR}_0(\lambda) \exp\left\{\tfrac{1}{2}\,\left(\sigma(\lambda)\right)^2 t\right\}\,\Phi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{t}\right)$$

The first derivative is given by

$$\frac{d}{d\lambda}\,\mathsf{EFS}_{\alpha,t}(\lambda) = -\frac{1}{\alpha}\,\frac{d}{d\lambda}\,\left(f(\lambda)g(\lambda)h(\lambda)\right)$$

with component functions

$$\begin{aligned} f(\lambda) &= \mathsf{FR}_0(\lambda) \\ g(\lambda) &= \exp\left\{\tfrac{1}{2}\,\left(\sigma(\lambda)\right)^2 t\right\} \\ h(\lambda) &= \Phi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{t}\right) \end{aligned}$$

The derivatives of these components are

$$\begin{aligned} f'(\lambda) &= \mathsf{FR}_0'(\lambda) > 0 \\ g'(\lambda) &= \exp\left\{\tfrac{1}{2}\ (\sigma(\lambda))^2\, t\right\} \sigma(\lambda)\sigma'(\lambda)\, t > 0 \\ h'(\lambda) &= -\varphi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{t}\right) \sigma'(\lambda)\sqrt{t} < 0 \end{aligned}$$

Applying the three-factor version of the product rule, i.e.

$$\begin{aligned} (fgh)' &= (fg)'h + (fg)h' \\ &= (f'g + fg')h + (fg)h' \\ &= f'gh + fg'h + fgh' \end{aligned}$$

and taking account of the minus signs, we obtain:

$$\begin{aligned} \frac{d}{d\lambda}\, \mathsf{EFS}_{\alpha,t}(\lambda) = \tfrac{1}{\alpha}\ \exp\left\{\tfrac{1}{2}\ (\sigma(\lambda))^2\, t\right\} \times \\ \Big[\mathsf{FR}_0(\lambda)\, \varphi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{t}\right) \sigma'(\lambda)\sqrt{t} \\ - \mathsf{FR}_0'(\lambda)\, \Phi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{t}\right) \\ - \mathsf{FR}_0(\lambda)\, \Phi\left(\Phi^{-1}(\alpha) - \sigma(\lambda)\sqrt{t}\right) \sigma(\lambda)\, \sigma'(\lambda)\, t\Big] \end{aligned}$$

We note that all single terms within the square brackets are positive. Hence, for $\frac{d}{d\lambda}\,\mathsf{EFS}_{\alpha,t}(\lambda)$ to be negative, we must have

$$\mathsf{FR}_0(\lambda)\,\varphi\left(\Phi^{-1}(\alpha)-\sigma(\lambda)\sqrt{t}\right)\sigma'(\lambda)\sqrt{t} < \Phi\left(\Phi^{-1}(\alpha)-\sigma(\lambda)\sqrt{t}\right)\left[\mathsf{FR}_0'(\lambda)+\mathsf{FR}_0(\lambda)\sigma(\lambda)\sigma'(\lambda)\,t\right]$$

which is equivalent to

$$\frac{\varphi\left(\Phi^{-1}(\alpha)-\sigma(\lambda)\sqrt{t}\right)}{\Phi\left(\Phi^{-1}(\alpha)-\sigma(\lambda)\sqrt{t}\right)} < \frac{\mathsf{FR}_0'(\lambda)+\mathsf{FR}_0(\lambda)\,\sigma(\lambda)\,\sigma'(\lambda)\,t}{\mathsf{FR}_0(\lambda)\,\sigma'(\lambda)\sqrt{t}}$$

which is also equivalent to

$$\frac{\varphi\left(\Phi^{-1}(\alpha)-\sigma(\lambda)\sqrt{t}\right)}{\Phi\left(\Phi^{-1}(\alpha)-\sigma(\lambda)\sqrt{t}\right)} < \frac{\mathsf{FR}_0'(\lambda)}{\mathsf{FR}_0(\lambda)\,\sigma'(\lambda)\sqrt{t}}+\sigma(\lambda)\sqrt{t}$$

That is, a situation with $\frac{d}{d\lambda}\,\mathsf{EFS}_{\alpha,t}(\lambda) < 0$ can, in principle, arise, depending on the specific parametrizations of $\sigma(\lambda)$ and $\mathsf{FR}_0(\lambda)$.

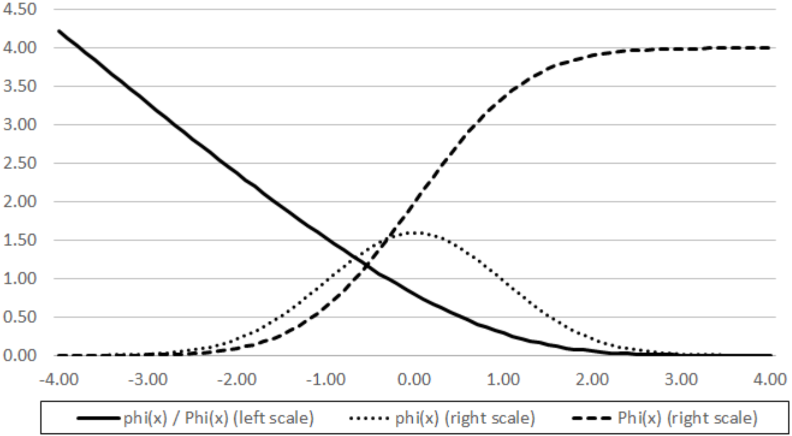
The criterion depends on the chosen level of safety α. The higher the level of safety, i.e. the lower α (e.g. 1%), the more restrictive the criterion becomes; see the graphic on the next page.

The right-hand side depends on the relative growth of the risk-taking capability, i.e. $\mathsf{FR}_0'(\lambda)/\mathsf{FR}_0(\lambda)$ as opposed to the growth and level of investment risk, i.e. $\sigma'(\lambda)$ and $\sigma(\lambda)$, although the relationship is less handy than the one for ψ_t.

In any case, a high growth of investment risk, i.e. a high value of $\sigma'(\lambda)$, reduces the term on the right-hand side and makes the criterion less likely to be satisfied, which is plausible.

Here again, the global optimum $\lambda^* = \arg\min_{\lambda \in [\mu_{\min},\, \mu_{\max}]} \mathsf{EFS}_{\alpha,t}(\lambda)$ is best obtained by evaluating the function $\mathsf{EFS}_{\alpha,t}(\lambda)$ over all feasible values $\lambda \in [\mu_{\min},\, \mu_{\max}]$ and looking up λ^*.

Standard Normal density and CDF and their ratio



Funding Ratio at Risk

According to Definition 6 and Proposition 13 in Chapter V, we have

$$\mathsf{FRaR}_{\alpha,t} = 1 - \mathsf{FR}_0 \exp\left\{(\mu - \lambda)t + \Phi^{-1}(\alpha)\sigma\sqrt{t}\right\}$$

For $\mu = \lambda$, $\mathsf{FR}_0 = \mathsf{FR}_0(\lambda)$ and $\sigma = \sigma(\lambda)$, this becomes

$$\mathsf{FRaR}_{\alpha,t}(\lambda) = 1 - \mathsf{FR}_0(\lambda) \exp\left\{\Phi^{-1}(\alpha)\,\sigma(\lambda)\sqrt{t}\right\}$$

Taking the first derivative is straightforward:

$$\frac{d}{d\lambda}\mathsf{FRaR}_{\alpha,t}(\lambda) = -\exp\left\{\Phi^{-1}(\alpha)\,\sigma(\lambda)\sqrt{t}\right\} \times \left[\mathsf{FR}_0'(\lambda) + \mathsf{FR}_0(\lambda)\,\Phi^{-1}(\alpha)\,\sigma'(\lambda)\sqrt{t}\right]$$

We note that $\Phi^{-1}(\alpha) < 0$ for all sensible values, i.e. $\alpha \ll 0.5$. Therefore, also in this case, there can be situations where $\frac{d}{d\lambda}\mathsf{FRaR}_{\alpha,t}(\lambda)$ is negative, i.e. where it makes sense to increase λ, and hence $\sigma(\lambda)$ in order to decrease long-term risk as expressed by $\mathsf{FRaR}_{\alpha,t}$.

For the derivative to be negative, we must thus have:

$$\mathsf{FR}_0'(\lambda) + \mathsf{FR}_0(\lambda)\,\Phi^{-1}(\alpha)\,\sigma'(\lambda)\sqrt{t} \;>\; 0$$

This is equivalent to

$$\mathsf{FR}_0'(\lambda) \;>\; -\mathsf{FR}_0(\lambda)\,\Phi^{-1}(\alpha)\,\sigma'(\lambda)\sqrt{t}$$

which, in turn, is equivalent to

$$\frac{\mathsf{FR}_0'(\lambda)}{\mathsf{FR}_0(\lambda)} \;>\; -\Phi^{-1}(\alpha)\,\sigma'(\lambda)\sqrt{t}$$

That ist, the relative growth of the risk-taking capability, i.e. $\mathsf{FR}_0'(\lambda)/\mathsf{FR}_0(\lambda)$ must be greater than the growth of the investment risk, i.e. $\sigma'(\lambda)$, weighted by the chosen level of safety α.

The higher the level of safety, i.e. the lower α (e.g. 1%), the more restrictive the criterion becomes. We also note that a lower initial funding ratio $\mathsf{FR}_0(\lambda)$ increases the left-hand side, so that the criterion is more easily satisfied.

Conclusions

Recall that the liability growth rate λ is a policy variable that the institution can choose. If the liabilities only consist of the present value of some promised cashflows, then λ is mainly determined by the discount rate δ. This is also predominantly the case in more general situations.

The consequence of the choice of λ is, however, that one must attain a corresponding expected investment return μ. In the Lognormal model, we specifically have $\mu = \lambda$. In particular, one must also incur the related investment risk $\sigma(\lambda)$. In a myopic setup, one would simply set λ as low as possible in order to incur the lowest possible investment risk.

This may, however, not be optimal in the long term. Choosing a higher liability growth rate and thus incurring a higher short-term investment risk may actually lead to a lower risk of underfunding in the long run, relative to all considered risk measures ψ_t, $\mathsf{EFS}_{\alpha,t}$ and $\mathsf{FRaR}_{\alpha,t}$.

Whether or not this is the case, depends on the specific liability profile $(\lambda \mapsto \mathsf{FR}_0(\lambda))$ and risk / return profile $(\mu \mapsto \sigma(\mu))$. In the simple Lognormal framework used here, the interplay between these profiles can be studied and optimized analytically. In more general settings, the same logic can still be applied directly, based on stochastic simulation models structured accordingly.

5. Example

Situation and task

We consider an institution that might e.g. be a Swiss Pension fund. We assume that the institution is in equilibrium, i.e. cash inflows and outflows from insurance operations cancel out, so that we can directly apply the Lognormal model.

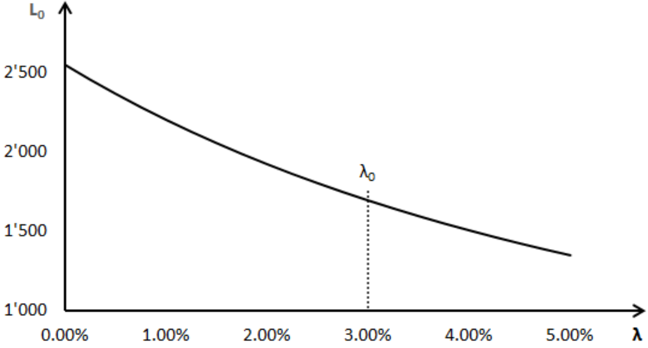
The current discount rate / liability growth rate / required return λ_0 equals 3.0%. At this rate, the liabilities have a value L_0 of 1'695 MCHF; their duration D^L equals 12.2. The institution holds assets A_0 of 2'035 MCHF, so that the funding ratio FR_0 at λ_0 equals 120%.

We assume that the risk / return profile of nowadays, as shown in Chapter VI, is applied for the institution. Under this risk / return profile, an expected return $\mu = \lambda_0$ of 3.0% is feasible. However, it comes at a high volatility $\sigma(\lambda_0)$ of 8.33%.

Question: Is λ_0 still optimal in terms of long-term risk, or would a different λ^* lead to a lower long-term risk. We consider all risk measures ψ_t, $\mathsf{EFS}_{\alpha,t}$ and $\mathsf{FRaR}_{\alpha,t}$ with a time horizon of $t = 10$ years and a safety level of $\alpha = 5\%$.

Liability valuations

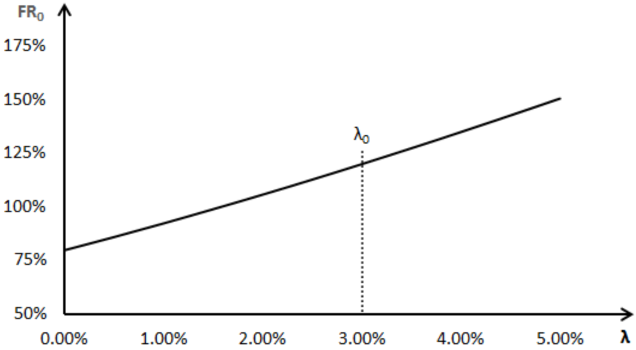
Valuations of the liabilities L_0 resulting from the application of different possible discount rates / liability growth rates λ:



Depending of the value of λ chosen, the valuation of the liabilities can vary quite considerably. Note that cashflows are up to 40 years in the future, and the duration of the liabilities is around 12.

Liability profile

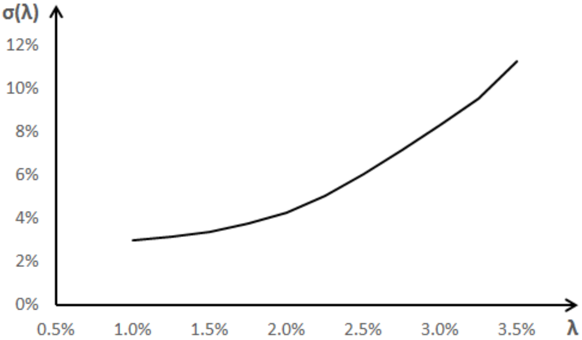
The initial value of the assets is unaffected by the choice of λ. Therefore, the following liability profile results from the liability values according to the last page:



As expected, higher values of λ lead to higher values of FR_0 and thus to a higher risk-taking capability. Does this relation keep pace with the risk / return profile?

Risk / return profile

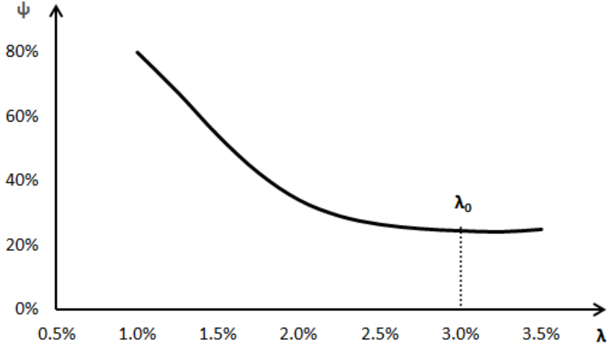
The risk / return profile in the representation adapted to our needs, i.e. with $\mu = \lambda$ as the independent policy variable and $\sigma(\lambda)$ as the dependent variable:



As required return λ is increased, the incurred investment risk $\sigma(\lambda)$ increases over-proportionally. Does this increase outpace the increase in risk-taking capability?

Probability of underfunding

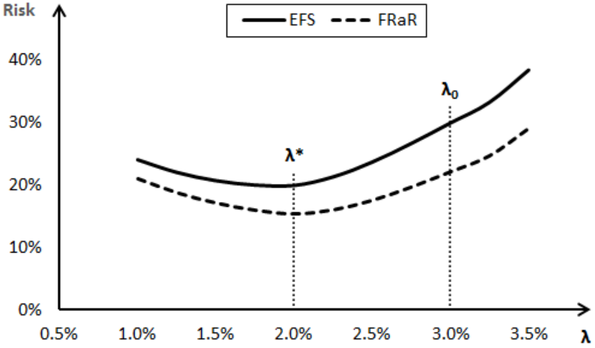
Probability of underfunding ψ_t (for $t = 10$ years) resulting from the liability profile and risk / return profile given above for different feasible values of λ:



For values $\lambda < 2.5\%$, ψ_t increases dramatically. Therefore, based on the consideration of ψ_t alone, it is advisable not to decrease λ below 2.5%. Even the current value $\lambda_0 = 3.0\%$ could be considered as optimal.

Alternative risk measures

$\text{EFS}_{\alpha,t}$ and $\text{FRaR}_{\alpha,t}$ (for $t = 10$ years and $\alpha = 5\%$) resulting from the liability profile and risk / return profile given above for different feasible values of λ:



Taking into account also the extent of an eventual underfunding, it appears advisable to select $\lambda^* \approx 2\%$, significantly lower than what is suggested by ψ_t. Values considerably higher or lower than 2% lead to significantly higher long-term risk.

Observations

As we have conjectured from the analytical considerations, the alternative risk measures that take into account the extent of an eventual underfunding lead to more conservative conclusions than the mere probability of an underfunding.

In real-world situations, it is, indeed, advisable to take into account the extent of an eventual underfunding. This extent represents the potential turnaround costs for the underfunding if the latter turns out to be intolerable.

Form among the alternative risk measures, Expected Funding Shortfall turns out to be consistently more conservative than Funding Ratio at Risk, as one would expect from the analytical construction of the two measures.

The conclusions drawn from these evaluations are fairly robust against variations in the time horizon t and the safety level α. The actual values of the risk measures differ considerably, but the derived optimal values λ^* are fairly similar.

In any real-world decision making, one should rely on some risk measure that takes into account the extent of an underfunding and its financial consequences, not just of the probability of it happening.

Bibliography I

[1] Koller, M.
Stochastic models in life insurance.
Lecture notes, ETH, Zürich, 2015.

Financial Risk Management in Social and Pension Insurance

Chapter VIII: Portfolio Construction and the Risk / Return Profile

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

November 14, 2020

Table of Contents

1. Introduction

Recall: model framework

Recall our generic model framework from Chapter IV:

$$\left.\begin{aligned} A_t &= A_{t-1}(1+R_t) + C_t \\ L_t &= L_{t-1}(1+\lambda_t) + C_t \end{aligned}\right\} \quad \text{for all } t \in \{1, \ldots, T\} \text{ and given } A_0, L_0$$

Here, our interest is on R_t, which denotes the rate of return from investing the assets A_{t-1} over the time period $(t-1, t]$, i.e.

$$R_t = \frac{A_t - C_t}{A_{t-1}} - 1 = \frac{A_{t-}}{A_{t-1}} - 1$$

Up until now, we have only cared about the total return R_t from investing the entire assets A_{t-1}. In the generic model, we did not make specific assumptions on the stochastic law of R_t, we simply put $R_t \sim F_t$ with expectation $\mathbf{E}[R_t] = \mu_t$ and finite variance $\mathbf{Var}[R_t] = \sigma_t^2 < \infty$.

In the Lognormal model of Chapter V, we assumed moreover that $R_t \sim \text{iid } \mathcal{N}(\mu, \sigma^2)$, i.e. $\mu_t \equiv \mu$ and $\sigma_t^2 \equiv \sigma^2$. We will maintain this assumption for the time being and discuss some limitations later.

Recall: risk / return profile

In Chapter VI, we assumed that by investing the assets A_t in different manners, we can attain different levels of expected returns $\mu \in [\mu_{\min},\, \mu_{\max}]$. And by doing so, we will incur an investment risk σ that depends on the level of expected return, i.e. $\sigma = \sigma(\mu)$ with

$$\left.\begin{array}{r}\sigma(\mu) > 0 \\ \sigma'(\mu) > 0 \\ \sigma''(\mu) \geq 0\end{array}\right\} \quad \text{for all } \mu \in [\mu_{\min},\, \mu_{\max}]$$

That is, if we want a higher level of return, we must invariably incur a higher level of investment risk.

We called the relationship $\mu \mapsto \sigma(\mu)$ the risk / return profile and took it as given.

In this chapter, we will look at how the assets A_t can be invested into different asset classes and how the risk / return profile can be derived from doing so. In particular, we will see that the postulated properties of the risk / return profile are actually sensible.

Terminology: asset classes

We assume that there are n different asset classes $i \in \{1, \ldots, n\}$. Asset classes can be e.g. foreign bonds, domestic equities, real estate or hedge funds. In practice, this set of available asset classes is often called the investment universe. We denote by $A_{i,t}$ the amount of money that is invested in asset class i, so that we have

$$A_t = \sum_{i=1}^{n} A_{i,t} \qquad \text{where } A_{i,t} \geq 0 \text{ for all } i \text{ and } t$$

Note that we could also allow for $A_{i,t} < 0$. This corresponds to a so-called short position. However, in a social and pension insurance context, such short positions are usually neither allowed nor desirable.

The portfolio weights $\mathbf{w}_t \in \mathbb{R}^n$ denote the relative share that each asset class i hold of the total portfolio, i.e.

$$\mathbf{w}_t = (w_{1,t}, \ldots, w_{n,t})' \in \mathbb{R}^n \quad \text{with} \quad w_{i,t} = \frac{A_{i,t}}{A_t}$$

Under the assumptions above, we have

$$\mathbf{w}_t \geq 0 \qquad \text{and} \qquad \mathbf{1}'\mathbf{w}_t = 1$$

The first expression is to be understood component-wise, and in the second expression $\mathbf{1} = (1, , \ldots , 1)' \in \mathbb{R}^n$ such that $\mathbf{1}'\mathbf{w}_t = \sum_{i=1}^{n} w_{i,t}$.

Note that $\mathbf{w}_t$ is a policy variable, i.e. in principle we can set $\mathbf{w}_t$ to values according to our needs and wishes at any time t by buying and selling assets of the different asset classes. In practice, market liquidity or other constraints may limit this ability to some extent.

Example: asset classes

Distribution of the investment assets of The Swiss National Accident Insurance Fund (Suva) onto various asset classes:

Investment categories (including derivatives), at market	31.12.2018 Balance CHF in 1,000s	2019 Changes in value CHF in 1,000s	31.12.2019 Balance CHF in 1,000s
Liquid assets[3]	1 711 104	73 367	1 784 471
Mortgages	692 072	46 157	738 229
Loans and syndicated loans[1]	5 802 340	312 058	6 114 398
Bonds in CHF[3]	9 344 117	312 058	9 656 175
Bonds in foreign currency[3]	5 889 129	480 314	6 369 443
Indirect real estate investments[3]	1 340 837	183 976	1 524 814
Investment properties[2]	5 218 074	248 384	5 466 458
Shares in Switzerland[3]	3 086 220	488 487	3 574 706
Shares outside Switzerland[3]	6 987 752	1 106 474	8 094 226
Alternative investments[3,4]	9 324 095	545 711	9 869 806
Overlays, hedging and opportunities[3]	164 476	314 995	479 471
Total	**49 560 216**	**4 111 982**	**53 672 198**

The investment universe of Suva is fairly vast.

Asset class returns

As for the total portfolio return R_t, each asset class has its return $R_{i,t}$, i.e.

$$R_{i,t} := \frac{A_{i,t} - C_{i,t}}{A_{i,t-1}} - 1 = \frac{A_{i,t^-}}{A_{i,t-1}} - 1$$

with $\mu_{i,t} = \mathbf{E}\left[R_{i,t}\right]$ and $\sigma_{i,t}^2 = \mathbf{Var}\left[R_{i,t}\right] < \infty$. The return vector is defined as $\mathbf{R}_t = (R_{1,t}, \ldots, R_{n,t})' \in \mathbb{R}^n$. For the time being, we assume that returns are iid multivariate Normal, i.e.

$$\mathbf{R}_t \sim \text{iid}\, \mathcal{N}_n\left(\boldsymbol{\mu}, \Sigma\right)$$

with expectation $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' = \mathbf{E}\left[\mathbf{R}_t\right] \in \mathbb{R}^n$ and with covariance matrix $\mathbf{Cov}\left[R_t\right] \in \mathbb{R}^{n \times n}$. The covariance matrix is symmetric and positive (semi-) definite. Its elements can be represented as

$$\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j \qquad \text{for all} \quad i, j \in \{1, \ldots, n\}$$

where $\sigma_i^2 = \mathbf{Var}\left[R_{i,t}\right]$, $\sigma_j^2 = \mathbf{Var}\left[R_{j,t}\right]$ and $\rho_{ij} = \mathbf{Corr}\left[R_{i,t}, R_{j,t}\right]$, i.e. the correlation which describes the dependence between asset classes i and j.

We always have $\rho_{ii} = 1$, and in the case of the multivariate Normal distribution, we also have $\rho_{ij} \in [-1, +1]$. Note that for other multivariate distributions, ρ_{ij} may only take values in a subset of $[-1, +1]$, and the correlation may not be a meaningful measure of dependence; see [4].

By $R = (\rho_{ij})_{i,j \in \{1, \ldots, n\}} \in \mathbb{R}^{n \times n}$, we denote the correlation matrix.

In asset management, it is fairly commonplace to consider the correlation matrix R and the standard deviations σ_i (called volatilities) separately, because this lends itself to a more intuitive interpretation than the covariance matrix.

If we have $\mathbf{R}_t \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and if the portfolio weights at the beginning of the period are $\mathbf{w}_{t-1}$, then we have

$$\mathbf{E}\left[R_t\right] = \mathbf{E}\left[\mathbf{w}'_{t-1}\mathbf{R}_t\right] = \mathbf{w}'_{t-1}\,\boldsymbol{\mu}$$

$$\mathbf{Var}\left[R_t\right] = \mathbf{Var}\left[\mathbf{w}'_{t-1}\mathbf{R}_t\right] = \mathbf{w}'_{t-1}\Sigma\,\mathbf{w}_{t-1}$$

This is a general property of multivariate distributions; see e.g. [1].

Example: investment universe

For our subsequent considerations, we will use the following, somewhat simplified investment universe:

	Expected Return		Volatility	Correlation Matrix					
	As of 2005	As of 2015		FI CHF	FI GL	CH EQ	GL EQ	RE	HF
Fixed Income in CHF	3.1%	0.5%	3.5%	1.00	0.71	-0.06	-0.14	0.26	-0.07
Fixed Income Global	4.4%	0.8%	4.3%	0.71	1.00	-0.14	-0.21	0.15	-0.01
Swiss Equities	6.5%	3.5%	15.2%	-0.06	-0.14	1.00	0.77	0.28	0.45
Global Equities	7.8%	4.1%	17.5%	-0.14	-0.21	0.77	1.00	0.25	0.54
Real Estate	5.3%	3.0%	7.8%	0.26	0.15	0.28	0.25	1.00	0.23
Hedge Funds	7.1%	2.3%	6.8%	-0.07	-0.01	0.45	0.54	0.23	1.00

Note that the expected returns had to be revised dramatically over time although they were meant to be long-term... (c.f. Chapters II, VI and VII). The volatilities and correlations, however, turned out to be fairly stable.

2. Investment strategy

Setup

We assume that we have n asset classes with returns $\mathbf{R}_t = (R_{1,t}, \ldots, R_{n,t})' \in R^n$ for all $t \in \{1, \ldots, T\}$. Specifically, we assume

$$\mathbf{R}_t \sim \text{ iid } \mathcal{N}_n(\boldsymbol{\mu}, \Sigma) \quad \text{for all} \quad t \in \{1, \ldots, T\}$$

That is, investment returns on asset class level have an iid multivariate Normal distribution with

- Expected return vector $\boldsymbol{\mu} = \mathbf{E}[\mathbf{R}_t] \in \mathbb{R}^n$,
- Covariance matrix $\Sigma = \mathbf{Cov}[\mathbf{R}_t] \in \mathbb{R}^{n \times n}$.

We specifically assume that Σ is positive definite instead of just semi-definite, and we also assume that $\boldsymbol{\mu}$ is linearly independent of $\mathbf{1}$, i.e. not all asset classes have the same expected return.

Moreover, we assume constant portfolio weights over time, i.e.

$$\mathbf{w}_t \equiv \mathbf{w} \quad \text{for all} \quad t \in \{1, \ldots, T\}$$

This is a so-called rebalancing strategy, i.e. at each time $t > 0$, the portfolio is rebalanced to its original weights $\mathbf{w}$ by buying and selling assets.

Illustration: rebalancing

Two asset classes: $i \in \{1, 2\}$ (think of bonds and equities; total assets $A_0 = 100$).

At $t = 0$: $w_{1,0} = 60\%$ and $w_{2,0} = 40\%$, corresponding to $A_{1,0} = 60$ and $A_{2,0} = 40$.

Between $t = 0$ and $t = 1$: $R_{1,1} = -2\%$ and $R_{2,1} = +10\%$

This leads to $A_{1,1^-} = 58.8$ and $A_{2,1^-} = 44.0$ for $A_{1^-} = 102.8$.

We have $R_1 = A_{1^-}/A_0 - 1 = 2.8\% = (60\%, 40\%)'(-2\%, +10\%)$.

Before rebalancing we have portfolio weights $w_{1,1^-} = 58.8/102.8 = 57.2\%$ and $w_{2,1^-} = 44.0/102.8 = 42.8\%$

Rebalancing: Sell 2.9 of Asset 2 and buy 2.9 of Asset 1 to obtain $A_{1,1} = 61.7$ and $A_{2,1} = 41.1$, corresponding to $w_{1,1} = 60\%$ and $w_{2,1} = 40\%$.

Basically, we sell the winners and buy the losers. This works well in a cyclical market.

More on rebalancing

A rebalancing strategy corresponds to an anti-cyclical behavior, i.e. assets with high returns in the previous period are sold, and those with low returns in the previous period are bought. If investment returns are mean-reverting, this is a reasonable strategy. Moreover, it keeps the risk profile of the portfolio constant.

Rebalancing strategies maintained over several years are fairly commonplace for long-term institutional investors.

- There may be different disciplines for rebalancing, e.g. calendar-based (each year / quarter / month), or threshold-based, i.e. when actual weights are too far away from required ones.
- There may or may not be some active overlay that allows for limited opportunistic deviations from the strategic portfolio weights $\mathbf{w}$.

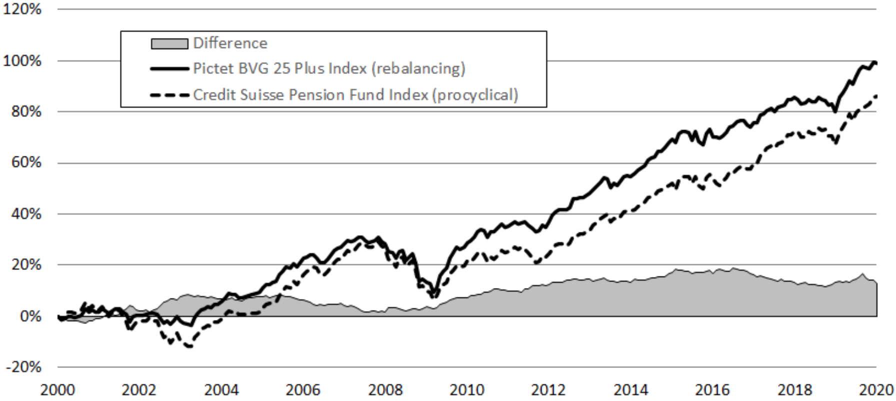
There exist more sophisticated, dynamic approaches to investment strategy design. They basically take into account the specific path of the returns and adjust weights in an adaptive manner.

- A simple example is portfolio insurance. This approach tends not to work when one needs it most.

Treatment of more sophisticated dynamic strategies requires stochastic optimization methodology that is beyond the scope of this course.

Illustration: value of rebalancing

Cumulative performance of Pictet BVG 25 Plus Index (strictly rebalancing) vs. Credit Suisse Pension Fund Index (containing procyclical behavior).



Rebalancing strategy outperforms sustainably. Outperformance is particularly high after losses sustained in financial crises.

Constraints

Form now on, let $\mathbf{w} = (w_1, \ldots, w_n)' \in \mathbb{R}^n$ denote the static portfolio weights that are valid for all times under consideration (not necessarily up to time T).

As already mentioned, we need to have

$$\sum_{i=1}^{n} w_i = 1 \qquad \text{or} \qquad \mathbf{1}'\mathbf{w} = 1$$

That is, we require that all assets are fully invested in the asset classes $1, \ldots, n$.

Moreover, we require that $w_i \geq 0$ for all $i \in \{1, \ldots, n\}$, abbreviated $\mathbf{w} \geq \mathbf{0}$. That is, no short positions are allowed. This makes sense for a pension fund, and it is often (e.g. in Switzerland) also mandated by law.

Moreover, there may be other constraints in place, either given by law or regulation, or self-imposed by the institution, or simply by necessity for practical reasons. These constraints may have two generic forms:

- Equality constraints: $f(\mathbf{w}) = 0$
- Inequality constraints: $f(\mathbf{w}) \leq 0$

Depending on how well-behaved the functions f and g are, these constraints may be more or less easily incorporated into an optimization problem. It is most convenient if constraints are linear, i.e.

- Equality constraints: $A\mathbf{w} = \mathbf{b}$ (or $A\mathbf{w} - \mathbf{b} = \mathbf{0}$)
- Inequality constraints: $A\mathbf{w} \leq \mathbf{b}$ (or $A\mathbf{w} - \mathbf{b} \leq \mathbf{0}$)

for some $A \in \mathbb{R}^{k \times n}$ and $\mathbf{b} \in \mathbb{R}^k$.

Example: constraints

Swiss pension fund regulation (BVV 2, SR 831.441.1) specifically imposes the following constraints:

- No short positions.
- Max. 50% in mortgages or mortgage-backed securities incl. Pfandbriefe.
- Max. 50% in equities.
- Max. 30% in real estate, incl. max. 10% in foreign real estate.
- Max. 15% in alternative investments.
- Max. 30% in unhedged foreign-currency investments of all kinds.

All these constraints can be represented as linear inequality constraints.

The optimization task

In an asset / liability setup, we are always given some required return. And we must have an expected return $\mu = \boldsymbol{\mu}'\mathbf{w} = \mathbf{E}[R_t]$ that is at least equal to this required return in order to assure a sustainable funding, provided that this is feasible; c.f. Chapter VI and VII. In the general model, this amounts to

$$\mu \geq \lambda_t + (\mathsf{FR}_{t-1} - 1)\frac{C_t}{A_{t-1}}$$

In the simplified Lognormal model, we simply have $\mu \geq \lambda$.

In the sequel, we consider μ as the given target return. We must find portfolio weights $\mathbf{w}$ such that $\boldsymbol{\mu}'\mathbf{w} \geq \mu$. There is no unique solution to this problem.

From among all possible solutions, it makes sense to choose the one that produces the minimal short-term investment risk $\sigma^2 = \mathbf{w}'\Sigma\mathbf{w}$. Including our other constraints, this amounts to

$$\min_{\mathbf{w}} \mathbf{w}'\Sigma\mathbf{w} \quad \text{s.t.} \quad \left.\begin{aligned} \mathbf{1}'\mathbf{w} &= 1 \\ \boldsymbol{\mu}'\mathbf{w} &\geq \mu \\ \mathbf{w} &\geq \mathbf{0} \end{aligned}\right\}$$

A portfolio $\mathbf{w}^*$ that is the solution of this optimization problem is then called mean / variance - efficient, because from all portfolios that satisfy the constraints, it is the one with the lowest variance, i.e. the lowest risk.

Optimization problems of this kind are called quadratic convex optimization problems (because the objective function is both quadratic and convex, and the constraints are convex).

There exist many numerical optimizers to solve such optimization problems efficiently. This is also the case if further constraints are added, provided that they are sufficiently well-behaved. For more information on mathematical optimization and its applications, see e.g. [3].

We will, however, follow a different path hereafter, i.e. we will try to explore some properties of mean / variance - efficient portfolios analytically so as to gain a deeper understanding.

3. Portfolio optimization in the mean / variance case

Task: properties of efficient portfolios

We use the setup just introduced, and we assume the following elements to be given:

- Expected return vector $\boldsymbol{\mu} \in \mathbb{R}^n$, linearly independent of $\mathbf{1}$.
- Covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$, positive definite.
- Target return $\mu \in \mathbb{R}$.

The mean / variance efficient portfolio $\mathbf{w}^*$ is then the solution of the problem

$$\min_{\mathbf{w}} \tfrac{1}{2}\,\mathbf{w}'\Sigma\mathbf{w} \quad \text{s.t.} \quad \left.\begin{aligned} \mathbf{1}'\mathbf{w} &= 1 \\ \boldsymbol{\mu}'\mathbf{w} &\geq \mu \\ \mathbf{w} &\geq \mathbf{0} \end{aligned}\right\}$$

Note that we have added the factor $\frac{1}{2}$ for mathematical convenience. While $\boldsymbol{\mu}$ and Σ are given by the financial markets, μ is a policy variable that we can set to different values.

For the different values of μ, different optimal portfolios $\mathbf{w}^*(\mu)$ will result,with different optimal portfolio variances $\sigma^2(\mu) = \mathbf{w}^*(\mu)'\Sigma\mathbf{w}^*(\mu)$. That is, we have a functional relationship $\sigma = \sigma(\mu)$, determined by $\boldsymbol{\mu}$ and Σ and the constraints. This is actually the risk / return profile that we already used.

We are now going to explore the properties of this risk / return profile in order to justify the assumptions that we simply made in the previous chapters.

- We do this first in the most basic setup with just the constraint $\mathbf{1}'\mathbf{w} = 1$.
- Then we add further constraint and look at their influence.

Prerequisite: Lagrange multipliers

This should be known from an undergraduate calculus course; otherwise see some appropriate textbook, e.g. [2].

Let $\mathcal{X}$ be an open subset of $\mathbb{R}^n$, and let

$$f : \mathcal{X} \to \mathbb{R} \quad \text{and} \quad \mathbf{g} = (g_1, \ldots, g_m)' : \mathcal{X} \to \mathbb{R}^m$$

be continuously differentiable; $m \leq n$. Let f have a local extremum in $\boldsymbol{\xi} \in \mathcal{X}$ that satisfies the constraint $\mathbf{g}(\boldsymbol{\xi}) = \mathbf{0}$. Assume moreover that the matrix

$$\mathbf{g}'(\boldsymbol{\xi}) = \left(\frac{\partial g_i(\boldsymbol{\xi})}{\partial x_j} \right)_{i \in \{1, \ldots, m\},\, j \in \{1, \ldots, n\}}$$

has full rank. Then, there exist m numbers $\lambda_1, \ldots, \lambda_m$ (the Lagrange multipliers) such that we have

$$f'(\boldsymbol{\xi}) + \sum_{i=1}^{m} \lambda_i \, g_i'(\boldsymbol{\xi}) = \mathbf{0}$$

Important: This only holds for equality constraints $g_i(\boldsymbol{\xi}) = 0$. For inequality constraints, the similar but more complicated Karush-Kuhn-Tucker conditions must be applied and evaluated; see [3].

In practice, this means that, if we want to solve the problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g_1(\mathbf{x}) = 0, \ldots \; g_m(\mathbf{x}) = 0$$

we first establish the Lagrange function or Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_m) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \, g_i(\mathbf{x})$$

Then, we take derivatives with respect to $\mathbf{x}$ and $\lambda_1, \ldots, \lambda_m$, set them all to zero and solve the resulting system of equations

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \mathbf{0}\,; \qquad \frac{\partial \mathcal{L}}{\partial \lambda_1} = 0\,, \quad \ldots\,, \quad \frac{\partial \mathcal{L}}{\partial \lambda_m} = 0$$

If f is twice continuously differentiable, and if the Hessian Matrix

$$(Hf)(\boldsymbol{\xi}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{\xi}) \right)_{i,j \in \{1, \ldots, n\}}$$

is positive definite in $\boldsymbol{\xi}$, then $\boldsymbol{\xi}$ is actually a minimum. For our typical target function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}'\Sigma\mathbf{x}$, we have $(Hf)(\mathbf{x}) = \Sigma$, which is positive definite by assumption.

Global minimum variance portfolio

To begin with, we drop the short constraint $(\mathbf{w} \geq 0)$ as well as the return requirement $(\boldsymbol{\mu}'\mathbf{w} \geq \mu)$ and consider the simplified problem

$$\min_{\mathbf{w}} \tfrac{1}{2}\, \mathbf{w}'\Sigma\mathbf{w} \quad \text{s.t.} \quad \mathbf{1}'\mathbf{w} = 1 \qquad (\Leftrightarrow \mathbf{1}'\mathbf{w} - 1 = 0)$$

The resulting optimal portfolio $\mathbf{w}_{\text{min}}$ is called global minimum variance portfolio. It is the least risky portfolio given that one must be fully invested.

The Lagrangian is $\mathcal{L}(\mathbf{w}, \lambda) = \frac{1}{2}\, \mathbf{w}'\Sigma\mathbf{w} + \lambda(\mathbf{1}'\mathbf{w} - 1)$.

We must solve the following system

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \Sigma\mathbf{w} + \lambda\mathbf{1} = \mathbf{0}$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{1}'\mathbf{w} - 1 \;\; = 0$$

The first equation yields

$$\mathbf{w} = -\lambda\Sigma^{-1}\mathbf{1}$$

Inserting this into the second equation yields

$$-\lambda \, \mathbf{1}'\Sigma^{-1}\mathbf{1} = 1 \qquad \Rightarrow \qquad -\lambda = \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

Thus, the global minimum variance portfolio $\mathbf{w}_{\mathsf{min}}$ is given by

$$\mathbf{w}_{\mathsf{min}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

For its return μ_{min} and its variance σ^2_{min}, we obtain

$$\mu_{\mathsf{min}} = \boldsymbol{\mu}'\mathbf{w}_{\mathsf{min}} = \frac{\boldsymbol{\mu}'\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

$$\begin{aligned} \sigma^2_{\mathsf{min}} &= \mathbf{w}'_{\mathsf{min}}\Sigma\mathbf{w}_{\mathsf{min}} \\ &= \frac{1}{(\mathbf{1}'\Sigma^{-1}\mathbf{1})^2} \, \mathbf{1}'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{1} \\ &= \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \end{aligned}$$

Example: simple investment universe

For the simple investment universe as introduced in Section 1, the global minimum variance portfolio and its properties look as follows:

	Minimum Variance Portfolio Weights
Fixed Income in CHF	65.0%
Fixed Income Global	12.2%
Swiss Equities	0.2%
Global Equities	0.8%
Real Estate	1.5%
Hedge Funds	20.2%
Total	100.0%

Expected Return μ as of 2005	4.15%
Expected Return μ as of 2015	0.96%

Variance σ^2	0.000895
Standard Deviation σ	2.99%

Portfolios with defined returns

Let us assume that $\mu \geq \mu_{\text{min}}$. And, given this, let us change the requirement $\boldsymbol{\mu}'\mathbf{w} \geq \mu$ into $\boldsymbol{\mu}'\mathbf{w} = \mu$. (One could show that this must be the case anyway for optimal portfolios.) This leads to the optimization problem (see also [6]):

$$\min_{\mathbf{w}} \tfrac{1}{2}\,\mathbf{w}'\Sigma\mathbf{w} \quad \text{s.t.} \quad \left.\begin{aligned} \mathbf{1}'\mathbf{w} - 1 &= 0 \\ \boldsymbol{\mu}'\mathbf{w} - \mu &= 0 \end{aligned}\right\}$$

The Lagrangian for this optimization problem is

$$\mathcal{L}(\mathbf{w}, \lambda_1, \lambda_2) = \tfrac{1}{2}\,\mathbf{w}'\Sigma\mathbf{w} + \lambda_1(\mathbf{1}'\mathbf{w} - 1) + \lambda_2(\boldsymbol{\mu}'\mathbf{w} - \mu)$$

This leads to the optimality conditions

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \Sigma\mathbf{w} + \lambda_1\mathbf{1} + \lambda_2\boldsymbol{\mu} = \mathbf{0} \qquad (1)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \mathbf{1}'\mathbf{w} - 1 = 0 \qquad (2)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = \boldsymbol{\mu}'\mathbf{w} - \mu = 0 \qquad (3)$$

From condition (1) we obtain

$$\mathbf{w} = -\lambda_1 \Sigma^{-1}\mathbf{1} - \lambda_2 \Sigma^{-1}\boldsymbol{\mu}$$

Provided that $\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu} \neq 0$, we can rewrite this as

$$\mathbf{w} = -\lambda_1 \left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right) \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} - \lambda_2 \left(\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}\right) \frac{\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}$$

From condition (2), we obtain moreover

$$\mathbf{1}'\mathbf{w} = -\lambda_1 \,\mathbf{1}'\Sigma^{-1}\mathbf{1} - \lambda_2 \,\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu} = 1$$

This is equivalent to

$$-\lambda_2 \,\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu} = 1 - (-\lambda_1 \,\mathbf{1}'\Sigma^{-1}\mathbf{1})$$

Letting $\nu := -\lambda_1 \,\mathbf{1}'\Sigma^{-1}\mathbf{1}$, we then have $1 - \nu = -\lambda_2 \,\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}$.

That is, we can represent the optimal portfolio $\mathbf{w}$ as

$$\mathbf{w} = \nu\, \mathbf{w}_{\mathsf{min}} + (1-\nu)\, \mathbf{w}_{\mathsf{risk}}$$

where we have

$$\mathbf{w}_{\mathsf{min}} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \qquad \text{and} \qquad \mathbf{w}_{\mathsf{risk}} = \frac{\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}$$

We note that $\mathbf{w}_{\mathsf{min}}$ is actually the minimum risk portfolio that we have introduced before. That is, the optimal portfolio $\mathbf{w}$ is a convex combination of the minimum risk portfolio $\mathbf{w}_{\mathsf{min}}$ and some risky portfolio $\mathbf{w}_{\mathsf{risk}}$. For the latter, we have

$$\mu_{\mathsf{risk}} = \boldsymbol{\mu}'\mathbf{w}_{\mathsf{risk}} = \frac{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}$$

$$\begin{aligned} \sigma^2_{\mathsf{risk}} &= \mathbf{w}'_{\mathsf{risk}}\Sigma\mathbf{w}_{\mathsf{risk}} \\ &= \frac{1}{(\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu})^2}\,(\Sigma^{-1}\boldsymbol{\mu})'\Sigma(\Sigma^{-1}\boldsymbol{\mu}) \\ &= \frac{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}{\left(\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}\right)^2} \end{aligned}$$

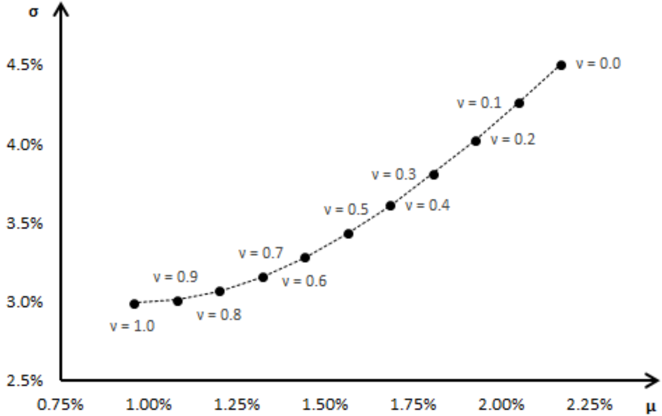
Example: simple investment universe

For the simple investment universe as introduced in Section 1, the risky portfolio and its properties look as follows:

	Minimum Variance Portfolio Weights	Risky Portfolio Weights As of 2005	Risky Portfolio Weights As of 2015
Fixed Income in CHF	65.0%	22.0%	-8.9%
Fixed Income Global	12.2%	39.3%	38.2%
Swiss Equities	0.2%	-0.4%	1.7%
Global Equities	0.8%	1.2%	2.8%
Real Estate	1.5%	6.3%	35.4%
Hedge Funds	20.2%	31.6%	30.9%
Total	100.0%	100.0%	100.0%
Expected Return μ as of 2005	4.15%	5.05%	
Expected Return μ as of 2015	0.96%		2.17%
Variance σ^2	0.000895	0.001091	0.002029
Standard Deviation σ	2.99%	3.30%	4.50%

Example: simple investment universe

Using $\mathbf{w}_{\text{min}}$ and $\mathbf{w}_{\text{risk}}$ from the simple investment universe (version 2015), for different values of ν, we obtain the following risk / return profile:



One might think that this is a hyperbola.

Relation to expected return μ

We now must further investigate ν. To this end, we evaluate constraint (3), i.e. $\boldsymbol{\mu}'\mathbf{w} = \mu$ by inserting the expression for $\mathbf{w}$:

$$\begin{aligned}
\mu &= \boldsymbol{\mu}' \left(\nu\, \mathbf{w}_{\mathsf{min}} + (1-\nu)\mathbf{w}_{\mathsf{risk}}\right) \\
&= \nu \left(\boldsymbol{\mu}'\mathbf{w}_{\mathsf{min}}\right) + \boldsymbol{\mu}'\mathbf{w}_{\mathsf{risk}} - \nu \left(\boldsymbol{\mu}'\mathbf{w}_{\mathsf{risk}}\right) \\
&= \nu\, \mu_{\mathsf{min}} + \mu_{\mathsf{risk}} - \nu\, \mu_{\mathsf{risk}}
\end{aligned}$$

This can be solved to yield

$$\nu = \frac{\mu_{\mathsf{risk}} - \mu}{\mu_{\mathsf{risk}} - \mu_{\mathsf{min}}}$$

That is, if $\mu = \mu_{\mathsf{risk}}$, then $\nu = 0$. And if $\mu = \mu_{\mathsf{min}}$, then $\nu = 1$. We also have:

$$1 - \nu = \frac{\mu - \mu_{\mathsf{min}}}{\mu_{\mathsf{risk}} - \mu_{\mathsf{min}}}$$

Shape of the risk / return profile

We want to establish a functional relationship between σ^2 and μ. To this end, we explore

$$\begin{aligned}\sigma^2 &= \mathbf{w}'\Sigma\mathbf{w} \\ &= (\nu\,\mathbf{w}_{\text{min}} + (1-\nu)\mathbf{w}_{\text{risk}})'\,\Sigma\,(\nu\,\mathbf{w}_{\text{min}} + (1-\nu)\mathbf{w}_{\text{risk}}) \\ &= \nu^2\,\mathbf{w}'_{\text{min}}\Sigma\mathbf{w}_{\text{min}} + 2\nu(1-\nu)\,\mathbf{w}'_{\text{min}}\Sigma\mathbf{w}_{\text{risk}} + (1-\nu)^2\,\mathbf{w}'_{\text{risk}}\Sigma\mathbf{w}_{\text{risk}}\end{aligned}$$

Recall that $\mathbf{w}'_{\text{min}}\Sigma\,\mathbf{w}_{\text{min}} = \sigma^2_{\text{min}} > 0$ and $\mathbf{w}'_{\text{risk}}\Sigma\,\mathbf{w}_{\text{risk}} = \sigma^2_{\text{risk}} > 0$. We also have $\sigma^2_{\text{risk}} > \sigma^2_{\text{min}}$. Moreover, we have

$$\begin{aligned}\mathbf{w}'_{\text{min}}\Sigma\,\mathbf{w}_{\text{risk}} &= \frac{1}{(\mathbf{1}'\Sigma^{-1}\mathbf{1})(\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu})}(\mathbf{1}\Sigma^{-1})'\Sigma(\Sigma^{-1}\boldsymbol{\mu}) \\ &= \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = \sigma^2_{\text{min}}\end{aligned}$$

That is, we can rewrite the expression for σ^2 as

$$\begin{aligned}\sigma^2 &= \nu^2\sigma^2_{\text{min}} + \left(2\nu - 2\nu^2\right)\sigma^2_{\text{min}} + \left(1 - 2\nu + \nu^2\right)\sigma^2_{\text{risk}} \\ &= \nu^2\left(\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}\right) - 2\nu\left(\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}\right) + \sigma^2_{\text{risk}}\end{aligned}$$

This, in turn, can be rewritten as (noting that $\sigma^2_{\text{risk}} - \sigma^2_{\text{min}} > 0$)

$$\begin{aligned}\frac{\sigma^2}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}} &= \nu^2 - 2\nu + \frac{\sigma^2_{\text{risk}}}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}} \\ &= (\nu - 1)^2 + \frac{\sigma^2_{\text{risk}}}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}} - 1 \\ &= (\nu - 1)^2 + \frac{\sigma^2_{\text{min}}}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}}\end{aligned}$$

This is, indeed, the equation of a hyperbola in ν. And since

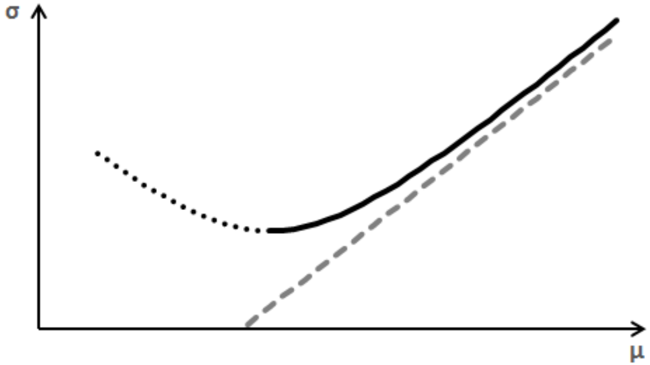
$$\nu = \frac{\mu_{\text{risk}} - \mu}{\mu_{\text{risk}} - \mu_{\text{min}}}$$

i.e. ν is just an affine coordinate transform of μ, this also describes a north - south opening hyperbola in $\mu - \sigma$ - space:

$$\frac{\sigma^2}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}} = \frac{(\mu - \mu_{\text{min}})^2}{(\mu_{\text{risk}} - \mu_{\text{min}})^2} + \frac{\sigma^2_{\text{min}}}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}}$$

Illustration: risk / return profile

The risk / return profile resulting from the optimization problem on p.20 (portfolio with defined returns) is a hyperbola in $\mu - \sigma$ - space.



Note that the hyperbola asymptotically tends towards a straight line.

Properties of the risk / return profile

Since the risk / return profile (under the given constraints in this optimization) is a hyperbola in $\mu - \sigma$ - space, it has the following generic representation:

$$\frac{\sigma^2}{a^2} - \frac{(\mu - c)^2}{b^2} = 1$$

For the positive branch with $a > 0$, this means

$$\sigma(\mu) = a \sqrt{1 + \frac{(\mu - c)^2}{b^2}}$$

$$\text{for} \quad a^2 = 1; \quad b^2 = \frac{\sigma^2_{\text{min}}}{\sigma^2_{\text{risk}} - \sigma^2_{\text{min}}} \left(\mu_{\text{risk}} - \mu_{\text{min}}\right)^2; \quad c = \mu_{\text{min}}$$

Therefore, we have

- $\sigma(\mu) > 0$ for all $\mu \geq \mu_{\text{min}}$
- $\sigma'(\mu) > 0$ for all $\mu > \mu_{\text{min}}$ (with $\sigma'(\mu_{\text{min}}) = 0$)
- $\sigma''(\mu) > 0$ for all $\mu \geq \mu_{\text{min}}$ (with $\lim_{\mu \to \infty} \sigma''(\mu) = 0$)

as we have postulated in the previous chapters.

Digression: adding a risk-free asset

Assume we have an additional, risk-free asset 0 with deterministic return μ_0. Let $\tilde{\boldsymbol{\mu}} = (\mu_0, \boldsymbol{\mu})' \in \mathbb{R}^{n+1}$ and $\tilde{\mathbf{w}} = (w_0, \mathbf{w})' \in \mathbb{R}^{n+1}$. We solve the modified problem

$$\min_{\tilde{\mathbf{w}}} \frac{1}{2} \, \mathbf{w}'\Sigma\mathbf{w} \quad \text{s.t.} \quad \left.\begin{aligned} \tilde{\mathbf{1}}'\tilde{\mathbf{w}} &= 1 \\ \tilde{\boldsymbol{\mu}}'\tilde{\mathbf{w}} &= 1 \end{aligned}\right\}$$

Then, provided that $\tilde{\mathbf{1}}$ and $\tilde{\boldsymbol{\mu}}$ are linearly independent, one can show (see e.g. [6]) that optimal portfolios are still of the form

$$\tilde{\mathbf{w}} = \nu \, \tilde{\mathbf{w}}_1 + (1 - \nu) \, \tilde{\mathbf{w}}_2$$

but this time with

$$\tilde{\mathbf{w}}_1 = (1, 0, \ldots, 0)' \in \mathbf{R}^{n+1} \quad \text{and} \quad \tilde{\mathbf{w}}_2 = (0, w_1, \ldots, w_n)' \in \mathbf{R}^{n+1}$$

That is, each optimal portfolio is a combination of the risk-free asset and some fixed risky reference portfolio $\tilde{\mathbf{w}}_2$. One can also show that in this case, the risk / return profile $\sigma = \sigma(\mu)$ boils down to a straight line, i.e. we should have $\sigma'(\mu) > 0$ and $\sigma''(\mu) = 0$.

This is, in principle, more favorable than the situation with only risky assets. However, is is not realistic for all cases where $\tilde{\mathbf{w}}$ is at some distance from $\tilde{\mathbf{w}}_2$, since it is not possible to hold a large portion of the portfolio in cash.

Where do we stand?

Recall that in the optimization problem that we have just solved analytically, we have made two omissions:

- We made the constraint $\boldsymbol{\mu}'\mathbf{w} = \mu$ instead of $\boldsymbol{\mu}'\mathbf{w} \geq \mu$. For any $\mu \geq \mu_{\text{min}}$ this is not important.
- We have not required $\mathbf{w} \geq 0$, i.e. we still allow for short positions. This, however, is an important omission.

The optimizations problem including the additional inequality constraint $\mathbf{w} \geq 0$ can be solved numerically without any problem; c.f. also the risk / return profile examples already shown.

However, it is not tractable analytically. Thus, in order to gain some qualitative insights, we have to follow a different path.

Reformulation of the optimization problem

The generic optimization problem can be formulated as follows:

$$\min_{\mathbf{w}} f(\mathbf{w}) \quad \text{s.t.} \quad \left.\begin{array}{ll} g_i(\mathbf{w}) = 0, & i = 1, \dots, m \\ g_j(\mathbf{w}) \leq 0, & j = m+1, \dots, r \end{array}\right\}$$

Let us now consider the set $\mathcal{A}_r \subseteq \mathbb{R}^n$ containing all points $\mathbf{w} \in \mathbb{R}^n$ that satisfy all r constraints, i.e.

$$\mathcal{A}_r := \{\, \mathbf{w} \in \mathbb{R}^n \mid g_i(\mathbf{w}) = 0 \; \forall\, i = 1, \dots, m;\; g_j(\mathbf{w}) \leq 0 \; \forall\, j = m+1, \dots, r \}$$

The original optimization problem is thus equivalent to

$$\min_{\mathbf{w} \in \mathcal{A}_r} f(\mathbf{w})$$

Let $\mathbf{w}_r \in \mathcal{A}_r$ denote the optimal solution.

Adding further constraints

Now, we can add a further constraint, irrespective of whether this is an equality or inequality constraint:

$$\left.\begin{aligned} \min_{\mathbf{w}} f(\mathbf{w}) \quad \text{s.t.} \quad & g_i(\mathbf{w}) = 0, && i = 1, \ldots, m \\ & g_j(\mathbf{w}) \leq 0, && j = m+1, \ldots, r \\ & g_{r+1}(\mathbf{w}) \leq 0 && \end{aligned}\right\}$$

This can again be formulated in set notation, i.e.

$$\mathcal{A}_{r+1} := \{\, \mathbf{w} \in \mathbb{R}^n \mid \; g_i(\mathbf{w}) = 0 \;\forall\, i = 1, \ldots, m; \ldots$$
$$\ldots\; g_j(\mathbf{w}) \leq 0 \;\forall\, j = m+1, \ldots, r;\; g_{r+1}(\mathbf{w}) \leq 0\}$$

The only difference between $\mathcal{A}_r$ and $\mathcal{A}_{r+1}$ is the additional constraint. This also means that

$$\mathcal{A}_{r+1} \subseteq \mathcal{A}_r$$

That is, the additional constraint leads to an equal or smaller set of admissible points.

Now, there can be exactly two situations:

- $\mathbf{w}_r \in \mathcal{A}_{r+1}$, i.e. the solution of the original problem also satisfies the additional constraint. Then we have $\mathbf{w}_{r+1} = \mathbf{w}_r$, and also $f(\mathbf{w}_{r+1}) = f(\mathbf{w}_r)$.
- $\mathbf{w}_r \in \mathcal{A}_r \setminus \mathcal{A}_{r+1}$, i.e. the solution of the original problem does not satisfy the additional constraint. Then we have $\mathbf{w}_{r+1} \neq \mathbf{w}_r$. And, due to the optimality of $\mathbf{w}_r$ on all of $\mathcal{A}_r$, we also have $f(\mathbf{w}_{r+1}) > f(\mathbf{w}_r)$

Translated back to our portfolio optimization problem, this means:

- For the same level of expected return μ, adding a further constraint leads to at least as much or more risk than without the additional constraint.
- That is, the hyperbola that we have obtained before is a minorant for all the risk / return profiles where more constraints apply.
- In particular, the hyperbola is a minorant for the risk / return profile of the original optimization problem that also incorporates the non-negativity constraint.

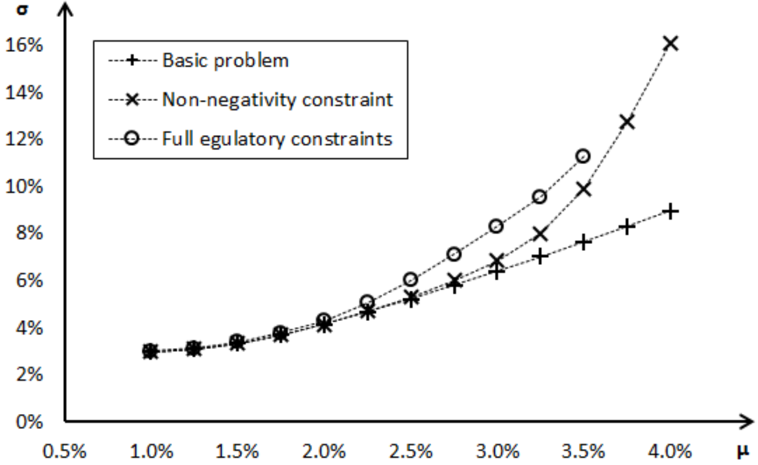
Example: simple investment universe

We use the simple investment universe as introduced in Section 1. Relative to the setup already explored analytically, we investigate (numerically) also the two following setups featuring additional constraints:

- Add only the non-negativity constraint $\mathbf{w} \geq 0$.
- Add also the additional constraints imposed by Swiss pension fund legislation and supervision (c.f. p. 19).

Example: influence of constraints

Risk / return profiles under various sets of constraints:



Under full constraints, returns above 3.5% are not feasible.

4. Conclusions and observations

Conclusions on the risk / return profile

The properties of the risk / return profile $\sigma = \sigma(\mu)$, i.e.

$$\left.\begin{aligned} \sigma(\mu) > 0 \\ \sigma'(\mu) > 0 \\ \sigma(\mu) \geq 0 \end{aligned}\right\} \qquad \text{for all} \quad \mu \in (\mu_{\mathsf{min}},\, \mu_{\mathsf{max}})$$

are confirmed to be sensible.

The more constraints one adds, the more risk one has to take in order to achieve the same expected return.

Therefore, investment constraints should only be imposed where they are really necessary and sensible. Freedom has a value and regulation has a cost.

Relation between return and short-term investment risk

So-called risk-adjusted performance measures put expected return in relation with the risk that must be incurred to earn the return. Let $\mu_{\mathbf{w}} = \boldsymbol{\mu}'\mathbf{w}$ and $\sigma^2_{\mathbf{w}} = \mathbf{w}'\Sigma\mathbf{w}$ denote the expected value and the variance of the portfolio with weights $\mathbf{w}$.

The best-known example is the Sharpe Ratio which assumes the presence of a risk-free asset with deterministic return μ_0. Then we have:

$$\mathsf{SR}_{\tilde{\mathbf{w}}} = \frac{\mu_{\tilde{\mathbf{w}}} - \mu_0}{\sigma_{\tilde{\mathbf{w}}}}$$

In our context, where the availability of a risk-free asset is limited, it makes sense to consider the risk-adjusted return above the minima given by the global minimum variance portfolio (c.f. p.27):

$$\mathsf{RAR}_{\mathbf{w}} = \frac{\mu_{\mathbf{w}} - \mu_{\mathsf{min}}}{\sigma_{\mathbf{w}} - \sigma_{\mathsf{min}}}$$

The Information Ratio considers unadjusted risk and return. Its application is, however, more suitable in the context of active asset management:

$$\mathsf{IR}_{\mathbf{w}} = \frac{\mu_{\mathbf{w}}}{\sigma_{\mathbf{w}}}$$

Contributions to risk and return

We have $\mu_{\mathbf{w}} = \boldsymbol{\mu}'\mathbf{w} = \sum_{i=1}^{n} w_i \mu_i$. Therefore, we can easily identify the absolute and relative contribution of some asset class i to the overall return:

$$\mathsf{AC}_i^{\mathsf{Return}} = w_i\, \mu_i \qquad \text{resp.} \qquad \mathsf{RC}_i^{\mathsf{Return}} = \frac{w_i\, \mu_i}{\mu_{\mathbf{w}}}$$

For the risk, such an easy additive decomposition is less straightforward. First, we have to take $\sigma_{\mathbf{w}}^2$ rather than $\sigma_{\mathbf{w}}$. Then, we can state:

$$\sigma_{\mathbf{w}}^2 = \mathbf{w}'\Sigma\mathbf{w} = \sum\nolimits_{i=1}^{n} w_i \left(\sum\nolimits_{j=1}^{n} \sigma_{ij}\, w_j \right)$$

The term in the parentheses contains both the risk from the variance of the asset class itself ($\sigma_{ii} w_i$) as well as the risk from the covariance with the other asset classes ($\sigma_{ij} w_j$, $j \neq i$). Thus, we can state

$$\mathsf{AC}_i^{\mathsf{Risk}} = \sum\nolimits_{j=1}^{n} \sigma_{ij} w_j \qquad \text{and} \qquad \mathsf{RC}_i^{\mathsf{Risk}} = \frac{1}{\sigma_{\mathbf{w}}^2} \sum\nolimits_{j=1}^{n} \sigma_{ij} w_j$$

Marginal contributions to risk and return

Here, we consider the effect that a marginal change in the portfolio weight of asset class i has on overall risk and return.

For the return:

$$\frac{d}{dw_k}\,\mu_{\mathbf{w}} = \mu_k$$

For the risk:

$$\frac{d}{dw_k}\,\sigma^2_{\mathbf{w}} = \frac{d}{dw_k}\,\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\,w_i\,w_j = 2\left(w_k\,\sigma_{kk} + \sum_{i=1}^{n} w_i\sigma_{ik}\right)$$

It is worthwhile to increase the weight of an asset class if it has a favorable relationship between high marginal contribution to return and low marginal contribution to risk.

Limitations of the mean / variance optimization

Mean / variance optimization is an excellent and simple tool for studying the interplay between risk, return and constraints in investment portfolios. However, in real-world settings, it has a number of limitations.

In particular, asset class returns are assumed to follow a multivariate Normal distribution. This may be relaxed somewhat to general elliptical distributions, but this is still a restrictive assumption.

- It is well-known that many financial time series exhibit significantly non-Normal behavior; see [4]. The higher the frequency, the more so.
- Moreover, dependence structures tend to be non-linear. In particular, there may be correlation breakdowns in times of financial crises. That is, diversification may not work when it is needed most urgently; see also [4].
- This may be less of an issue with yearly data, but care must still be taken!

In any case, the underlying markets must be analyzed carefully. If Normality cannot be assumed to hold, either different models should be used, or the Normal model should be complemented with suitable stress tests.

The portfolios obtained from mean / variance optimization tend to be rather sensitive to changes in estimates for expected returns $\boldsymbol{\mu}$ and covariances Σ. Given the high estimation uncertainty, particularly for $\boldsymbol{\mu}$, this is an issue.

Example: sensitivities

Simple investment universe as on p.11. Slight changes in return estimates lead to rather sizable changes in optimal portfolio weights.

Base case		CH a bit better		Bonds a bit better	
mu	w	mu	w	mu	w
0.50%	0.00%	0.75%	10.72%	1.00%	0.00%
0.75%	35.85%	0.75%	25.25%	1.25%	42.99%
3.50%	6.95%	3.75%	10.73%	3.50%	2.86%
4.10%	12.21%	4.10%	8.31%	4.10%	9.14%
3.00%	30.00%	3.00%	30.00%	3.00%	30.00%
2.25%	15.00%	2.25%	15.00%	2.25%	15.00%

The same effects would be obtained with changes in volatility or correlation assumptions.

The risk / return profile, i.e. $\sigma = \sigma(\mu)$ is less sensitive against such changes, which is an advantage in the present context.

Alternatives for portfolio construction

For portfolio optimization, there exist many alternatives to the simple mean / variance optimization introduced here; see e.g. [5]:

- Portfolio bootstrapping, which does not rely on any parametric model for the asset returns.
- Use of different risk measures, e.g. mean / expected shortfall optimization.
- Bayesian approaches that explicitly incorporate the estimation uncertainty for μ in the Black / Litterman framework.

In practice, the construction of the strategic portfolio $\mathbf{w}$ of an institutional investor will involve both qualitative and quantitative steps. Operational restrictions also play an important role.

If it simply comes to determining the risk / return profile $\sigma = \sigma(\mu)$, the solution obtained from mean / variance optimization is always a reference: actual risk $\sigma(\mu)$ for some give level of return μ will likely not be lower that what is suggested by the mean / variance optimization.

Irrespective of the optimization method: The generic characteristics of the risk / return profile, i.e. $\sigma(\mu) > 0$, $\sigma'(\mu) > 0$ and $\sigma''(\mu) \geq 0$ are certainly plausible.

Bibliography I

[1] GRIMMETT, G., AND STIRZAKER, D.
Probability and Random Processes, second ed.
Clarendon Press, Oxford, 1992.

[2] HEUSER, H.
Lehrbuch der Analysis.
Teubner, Stuttgart, 1993.

[3] LUPTACIK, M.
Mathematical Optimization and Economic Analysis.
Springer, Berlin, 2010.

[4] MCNEIL, A., FREY, R., AND EMBRECHTS, P.
Quantitative Risk Management, revised ed.
Princeton University Press, Princeton, 2015.

[5] MEUCCI, A.
Risk and Asset Allocation.
Springer, Berlin, 2005.

Bibliography II

[6] MÜLLER, H.
Modern portfolio theory: Some main results.
ASTIN Bulletin 18, 2 (1988), 127–145.

Financial Risk Management in Social and Pension Insurance

Chapter IX: Empirical Considerations

ETH Zurich, Fall Semester, 2020

Peter Blum

Suva, The Swiss National Accident Insurance Fund, Lucerne

November 14, 2020

Problem statement

In the Lognormal model, we assume that investment returns follow a Normal distribution $\mathcal{N}(\mu, \sigma^2)$. And in mean / variance portfolio optimization, we assume that the asset class returns follow a multivariate Normal distribution $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$.

Also with any other model or method, we make certain (implicit or explicit) assumptions on the stochastic law of the investment returns and other stochastic variables. It is extremely important to verify whether the underlying markets satisfy these assumtions. Otherwise, the model is invalid.

Therefore, empirical data analysis is an indispensable part of the modelling process. One must understand the relevant empirical properties of the data in order to make valid models and draw valid conclusions: Let the data tell us their story!

The first step is always a descriptive analysis by preparing graphical evaluations and by computing summary statistics. This is assumed to be known and not treated any further here.

Then, inferential analysis can be made; see e.g. [1] or [4]. We will limits ourselves here to some evaluations in order to check Normality.

Descriptive statistics

Everyone according to his or her own needs and preferences...

(Data and chart: Bloomberg)

Multivariate Normal distribution

A random vector $\mathbf{X} \in \mathbb{R}^n$ is said to have a multivariate Normal distribution, written $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, if its probability density function is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \, \Sigma^{-1} \, (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

for some vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and some symmetric and positive definite matrix $\Sigma \in \mathbb{R}^{n \times n}$. We then have (refer e.g. to [2] for all this):

$$\mathbf{E}\left[\mathbf{X}\right] = \boldsymbol{\mu} \qquad \text{and} \qquad \mathbf{Cov}\left[\mathbf{X}\right] = \Sigma$$

Moreover, for some $\mathbf{w} \in \mathbb{R}^n$ (think of portfolio weights), we have

$$\mathbf{E}\left[\mathbf{w}'\mathbf{X}\right] = \mathbf{w}'\boldsymbol{\mu} \qquad \text{and} \qquad \mathbf{Var}\left[\mathbf{w}'\mathbf{X}\right] = \mathbf{w}'\Sigma\mathbf{w}$$

If $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, then $X_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$ for all $i \in \{1, \ldots, n\}$. That is, the marginal distributions are univariate Normal.

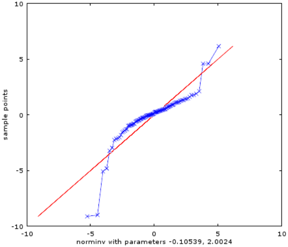
On the other hand, if $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ for $i \in \{1, \ldots, n\}$, this does not necessarily mean that $\mathbf{X} = (X_1, \ldots X_n)' \in \mathbb{R}^n$ is multivariate Normal. There must also be a very specific linear dependence structure; see [3].

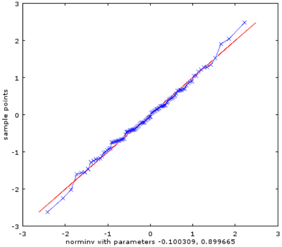
Quantile-Quantile plot (QQ plot)

Let $x_1, \ldots, x_K$ be K independent observations from some random variable X, and let $x_{(1)}, \ldots, x_{(K)}$ be the respective order statistics. $x_{(k)}$ corresponds to the empirical $\frac{k}{K}$- quantile.

For $X \sim F$ (e.g. $F = \mathcal{N}(\mu, \sigma^2)$), we can also compute the theoretical quantiles $F^{-1}\left(\frac{k-\frac{1}{2}}{K}\right)$. The correction $k - \frac{1}{2}$ is in order to avoid $F^{-1}(1)$.

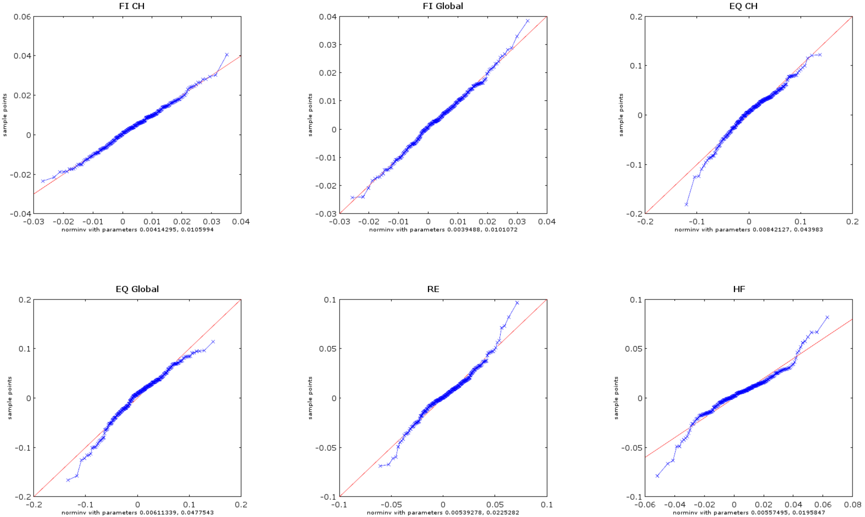
If $x_1, \ldots, x_K$ do actually come from $X \sim F$, then the scatter plot of empirical against theoretical quantiles should form approximately a straight line.





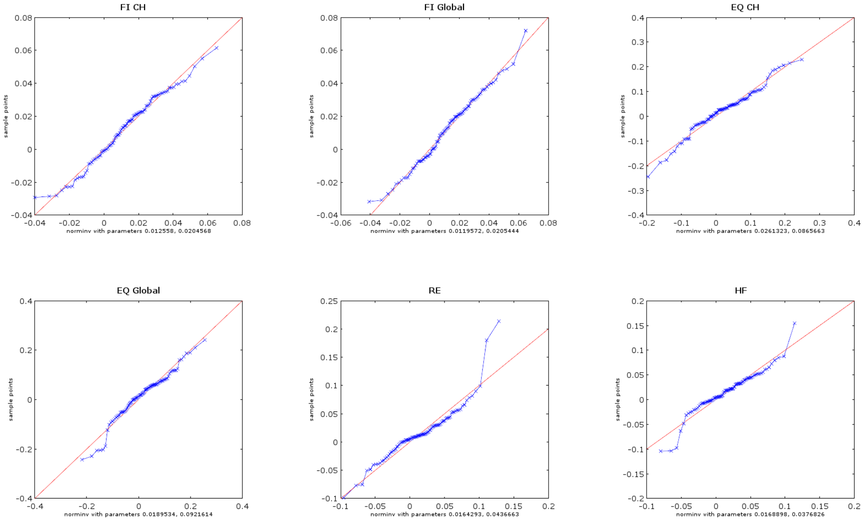
QQ plots for monthly returns

Clear deviations from normality in some cases. (Data: Bloomberg)



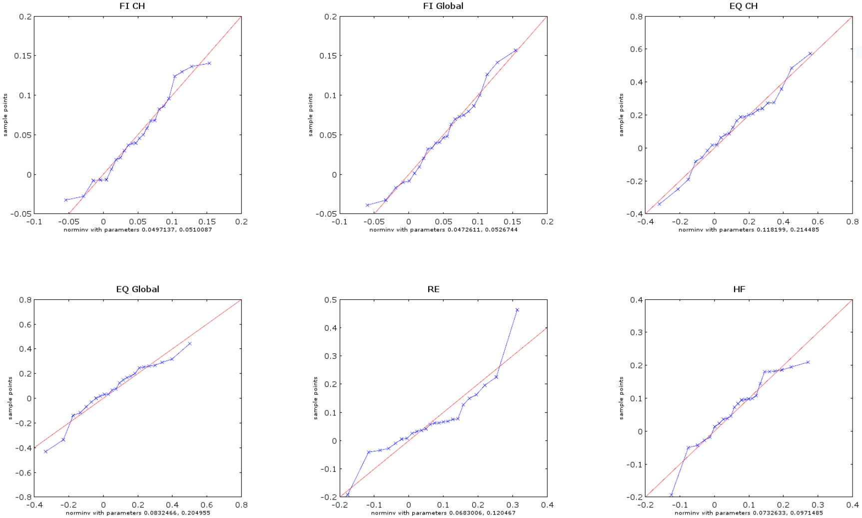
QQ plots for quarterly returns

Predominantly normal, but with exceptions and outliers. (Data: Bloomberg)



QQ plots for yearly returns

Usually normal, but occasional outliers occur. (Data: Bloomberg)



Kolmogorov-Smirnov test

Let $x_1, \ldots, x_K$ be K independent observations from some random variable X. The empirical cumulative distribution function is then given by

$$\hat{F}(x) = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}_{\{x_k \leq x\}}$$

If actually $X \sim F$, then the empirical distribution function $\hat{F}(x)$ should not be too different from the theoretical one. The difference can be expressed as

$$D_n := \|\hat{F} - F\|_\infty = \sup_x |\hat{F}(x) - F(x)|$$

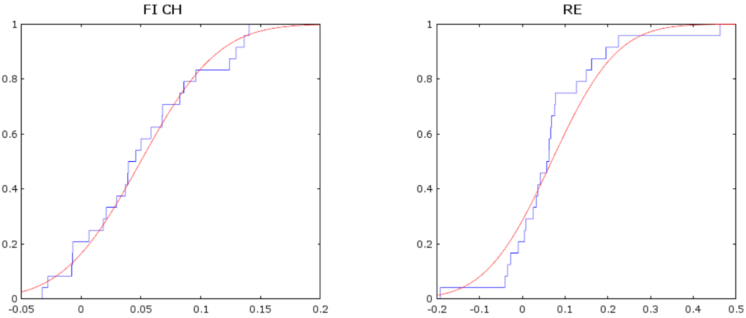
Under the null hypothesis $X \sim F$ against a two-sided alternative, we must have

$$\sqrt{n}\, D_n \xrightarrow{D} K \; (n \to \infty) \quad \text{where} \quad K(x) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp\left\{-2j^2 x^2\right\}$$

K is called Kolmogorov-Smirnov distribution (see e.g. [5]). The null hypothesis is rejected on significance level $\alpha \in (0,1)$ if $D_n > \frac{1}{\sqrt{n}} K^{\leftarrow}(1-\alpha)$.

Illustration: cumulative distribution functions

Empirical (jagged, blue) and theoretical (smooth, red) cumulative distribution functions for two data series:



The Kolmogorov-Smirnov test is based on the greatest (vertical) difference between theoretical and empirical distribution function.

Kolmogorov-Smirnov test: results

Data Series	Frequency	p-Value of KS test	Hypothesis of normality
FI CH	monthly	0.9543	by no means rejected
FI Global	monthly	0.6472	by no means rejected
EQ CH	monthly	0.0544	rejected
EQ Global	monthly	0.0126	clearly rejected
RE	monthly	0.0851	weakly rejected
HF	monthly	0.0324	rejected
FI CH	quarterly	0.8363	by no means rejected
FI Global	quarterly	0.8729	by no means rejected
EQ CH	quarterly	0.4151	by no means rejected
EQ Global	quarterly	0.3790	by no means rejected
RE	quarterly	0.2300	not rejected
HF	quarterly	0.2912	not rejected
FI CH	yearly	0.9842	by no means rejected
FI Global	yearly	0.9985	by no means rejected
EQ CH	yearly	0.9462	by no means rejected
EQ Global	yearly	0.9848	by no means rejected
RE	yearly	0.1949	not rejected
HF	yearly	0.9069	by no means rejected

Concluding remarks

For financial times series with frequencies of monthly and higher, non-normal return distributions are the rule rather than the exception (see [3]), and the use of models based on the Normal distribution is problematic in such cases.

For yearly investment returns, however, the hypothesis of a Normal distribution of the returns cannot be rejected in many cases. Models based on the Normal distribution may thus be viable in these instances.

We should, however, notice that even in the most well-behaved yearly time series, there are occasional outliers that cannot be reconciled with the hypothesis of normality.

When using Normal models, they should, therefore, always be complemented with some stress scenarios (historical ones or constructed ones).

Bibliography I

[1] Campbell, J., Lo, A., and MacKinlay, A.
The Econometrics of Financial Markets.
Princeton University Press, Princeton, 1997.

[2] Grimmett, G., and Stirzaker, D.
Probability and Random Processes, second ed.
Clarendon Press, Oxford, 1992.

[3] McNeil, A., Frey, R., and Embrechts, P.
Quantitative Risk Management, revised ed.
Princeton University Press, Princeton, 2015.

[4] Stahel, W.
Statistische Datenanalyse, third ed.
Vieweg, Braunschweig, 2000.

[5] Wüthrich, M.
Non-life insurance mathematics and statistics.
Lecture notes, ETH Zürich, Zürich, 2015.