Financial Risk

Lecture Notes

UNIVERSITY OF
GOTHENBURG

Sören Christensen

Göteborg Fall 2015

Preface

These lecture notes are based on 5 lectures within the framework of a course at Chalmers University of Technology and the University of Gothenburg in the autumn term 2015. This is only a draft version. It will be updated over the course.

This course is mainly based on [MFE05], but this monograph covers much more material than this course. Further useful lecture notes available online are [Kal14] and [HL07].

Contents

Chapter 1

Introduction and Basic Notations

1.1 Motivation and Aims

Uncertainty is everywhere: all financial institutions are subjected to risks and losses. In most situations, this is not alarming at all. However, large losses may prevent an institution from reaching its goals or even lead to bankruptcy. Therefore, the institutions, on the one hand, try to manage these risks for their own self-interest.On the other hand, extreme losses may also effect third parties and the whole system. To prevent these risks, a buffer capital is required from these institutions, often by law.

Here, risk means an event or action which prevents an institution from meeting its obligations or reaching its goals.

Many sorts of financial risks can be categorized into one of the following three groups:

- *Market risk*: risk that the value of a portfolio changes due to changes of market prices, commodity prices, exchange rates etc.
- *Credit risk*: risk that the value of a portfolio changes because a debtor cannot meet his obligations.
- *Operational risk*: risk caused by problems in internal processes, people, systems.

There are further risks, e.g., liquidity risk, legal risk, ... In this first part of this course, we primarily concentrate on market risk from a quantitative viewpoint. A second part is mainly on credit risk.

The main aims of this course are to develop tools to

- quantify risk,

- measure risk,
- find a suitable capital buffer that is needed to to cover unexpected losses.

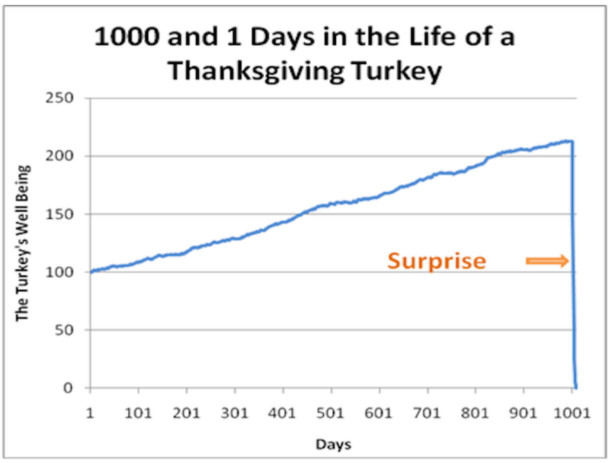
Although important in risk management, this course does not cover

- time-series models (see course on financial time series),
- option pricing (see course on options and mathematics),
- multivariate models (only very basic, perhaps).

1.2 The Turkey-Example

The following parable from Nassim Taleb's book *The Black Swan* ([Tal10]) illustrates one main problem in (quantitative) risk management:

> Consider a turkey that is fed every day. Every single feeding will firm up the bird's belief that it is the general rule of life to be fed every day by friendly members of the human race 'looking out for its best interests,' as a politician would say.
> On the afternoon of the Wednesday before Thanksgiving, something unexpected will happen to the turkey. It will incur a revision of belief.



In the quantitative language, the turkey sets up a model for its well being and uses observations to estimate the parameters to forecast its future well being. The turkey comes to the conclusion that his well being will always increase.

The statistics was fine, BUT his model ignored an important risk factor (Thanksgiving). Please, keep this example in mind over all the course (and beyond).

> If one does not understand the real-world situation well enough, the best quantitative tools will not help.

1.3 Quantitative risk management from a bird's eye view

1.3.1 Main steps

In order to apply quantitative risk management, several steps have to be taken.[1] They can often be summarized as follows (we will give more details in the following).

1. *Exploratory data analysis and modeling.* Since risk management concerns the unknown future, it is typically based on a mathematical model for the loss. The first step is to determine the structure of this model. More specifically, one needs to identify the relevant risk factors

$$Z_n = (Z_{n,1}, ..., Z_{n,d}),$$

set up a stochastic model for the distribution of the risk factors (more precisely, for their changes), and identify the functional dependence of the asset value V_n on these factors

$$V_n = f(t_n, Z_n),$$

see for Section 3.1 for details. These choices are typically based on an exploratory analysis of comparable data from the past as well as on theoretical considerations.

2. *Data collection and parameter estimation.* The model from Step 1 – in particular the one for the risk factor changes – is typically specified only up to some yet unkown parameters. For concrete applications, these must be estimated. To this end, one needs to dispose of reliable data in the first place.

3. *Stochastic forecast.* Based on the now completely specified stochastic model, one can compute an estimate of the conditional law of the future loss L_{n+1} given the data $(Z_1, ..., Z_n)$ up to the present. Possibly, only a quantile, moment etc. is needed instead of the whole law. For steps 2 and 3 see Chapters 3 and 4.

[1] This subsection is based on ...

4. *Backtesting.* Before these predictions are used in real risk management systems, they should be validated. This is usually done by reviewing whether they would have performed reasonably well in the past.

5. *Draw practical conclusions.* Finally, the prediction or assessment from the model needs to be translated into concrete actions, e.g., concerning buffer capital requirements, see the following Chapter.

1.3.2 Sources of error

In practice, many issues can cause the ultimate assessment of the risk to be faulty.

1. In the modeling step important risk factors may have been overlooked, e.g. counterparty risk, interest rate risk, liquidity risk etc. (Recall the Turkey-Example!) On top, the functional dependence linking risk factors and portfolio value may not always hold. Finally, the stochastic model for the risk factor changes may not be appropriate to describe real data sufficiently well.

2. Some parametric models require an enormous amount of data for reliable estimation, which may not be available in practice. And even if a long history of data is available, it is not clear whether the model from Step 1 is valid with fixed parameters for such a long time. Structural breaks e.g. after crises may lead to changing parameters and hence error-prone estimates.

3. The forecast may be biased due to numerical errors in the computation.

4. Backtesting may suffer from the fact that the model is built and tested with the same data. Events that have not occured in the past and are not allowed for in the model either, may still do so in the future.

5. In most cases the buffer capital will not be enough to cover extreme losses. Therefore it should be taken into account how severe consequences turn out to be if things go wrong.

Chapter 2

Risk Measures

One major question for financial institutions and regulators is:

> How can one determine a suitable capital buffer that is needed to cover unexpected losses?

Here, the aim is to quantify the risk of a (possibly highly complex) portfolio by a single number.

In this chapter, we consider a static setup. This means, we are at time point 0 and fix time horizon T (T time units ahead into the future) and a random variable L and interpret L as the loss of the specific portfolio at time T. Here $L \geq 0$ are losses, $L \leq 0$ are gains. This is a special case of the more general setup introduced in Section 3.1.[1]

2.1 Popular risk measures based on the loss distribution

It seems to be reasonable to use a stochastic model for L and to use this to define a corresponding risk measure. Such measures are presented in this section.
Using such measures, two major points have to be kept in mind:

- There is no guarantee that the model for L is well specified.

[1] In this abstract probabilistic setup it is tempting to think of the situation as having a game-like character. However, I would like to encourage you to also think about the real-world consequences of your models and the related ethical issues.

- For the most important risk measures, the tail of the distribution of L (the large losses) plays an important role. Even if the model for L is fine, it is not straightforward to do statistics for these extreme events, see Chapter 4.

But we ignore these points in this chapter and assume to have a fixed random variable L with known distribution.

2.1.1 Trivial risk measures: Expectation and Standard deviation

Two real numbers associated to each random variable are already well-known from elementary statistics: expectation and variance/standard deviation. The first reason not to consider just the expectation of a random variable as a risk measure is that it only describes an average loss, but gives no information on potential big losses.

Historically, one of the first risk measures used was the standard deviation

$$\rho(L) = c\sqrt{Var(L)},$$

sometimes with some adjustments. One immediately sees that it treats profits and losses the same way ($Var(L) = Var(-L)$). Furthermore, the variance gives only little information about the occurrence of extremely large losses[2]. Furthermore, this risk measure is not monotone: If $L \le L'$ are loss variables, it can happen that $Var(L) > Var(L')$ (for example take L' constant), which is obviously not desirable for a risk measure. The standard deviation is not used as a risk measure in practice nowadays.

2.1.2 Value at Risk (VaR)

Widely used and prescribed in regulations is the value at risk – VaR or V@R – which depends on a parameter $p \in (0, 1)$. For a loss variable L the p-value at risk $VaR_p(L)$ denotes the amount of buffer capital c^* such that the probability of a loss exceeding c^* is at most $1 - p$. More formally:

Definition 2.1. *For $p \in (0, 1)$ and a random variable L we write*

$$VaR_p(L) = \inf\{c : P(L > c) \le 1 - p\}.$$

[2]the best that can be said in general is Chebyshev's inequality

$$P(|L - E(L)| > z) \le \frac{Var(L)}{z^2}.$$

This quantity is called the Value at Risk.

Using the distribution function

$$F_L(c) = P(L \le c) = 1 - P(L > c)$$

we may write equivalently

$$VaR_p(L) = \inf\{c : 1 - F_L(c) \le 1 - p\} = \inf\{c : F_L(c) \ge p\}.$$

Remark 2.1. *1. If F_L is continuous then $P(L > VaR_p(L)) = 1 - p$.*

2. In practice, the risk measure VaR depends both on the parameter p and the time horizon. Typical time horizons are 1 or 10 days for market risk and 1 year for credit risk. Typical values for p are $0.95, 0.99, 0.999$. In the Basel-guidelines for market risk, 10 days and $p = 0.99$ is often used.

3. The VaR can be described very well using the quantile function / generalized inverse

$$F^{-1} = F_L^{\leftarrow} : (0,1) \to \mathbb{R}, \quad F_L^{\leftarrow}(p) = \inf\{c : F_L(c) \ge p\},$$

which is well-known in statistics. In this notation: $VaR_p(L) = F_L^{\leftarrow}(p)$.

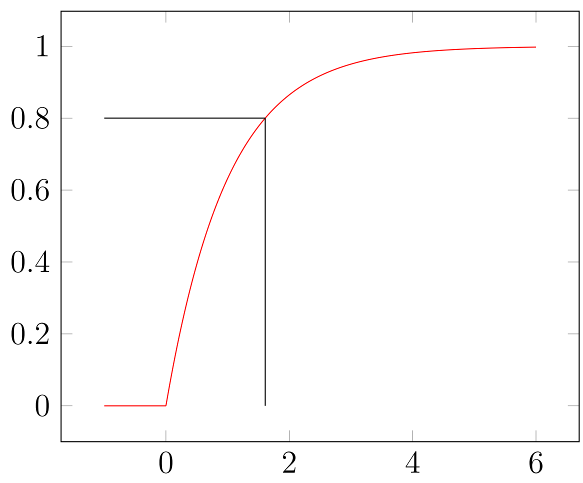


Figure 2.1: For continuous strictly increasing F, the generalized inverse is the usual inverse function

4. It is easily checked that VaR is translation invariant, that is for $b \in \mathbb{R}$

$$VaR_p(L + b) = VaR_p(L) + b.$$

Furthermore, it is positively homogenous, i.e. for $a > 0$

$$VaR_p(aL) = aVaR_p(L).$$

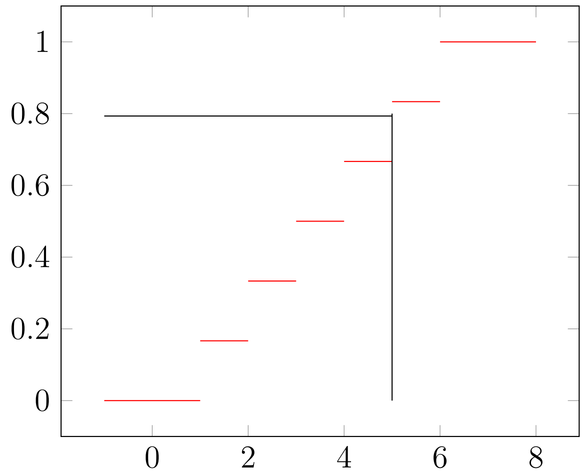


Figure 2.2: For a discrete distribution F the general definition has to be used.

Example 2.2. *Recall that for a standard normal variable Y it holds that $L = aY + b$ is $N(b, a^2)-$distributed. Using this and the previous remark, one obtains*

$$VaR_p(L) = a\Phi^{\leftarrow}(p) + b,$$

where Φ denotes the distribution function of the standard normal.
We will see later that normal variables are often not appropriate to model loss distributions.

The following example, taken from [MFE05], illustrates that the VaR has some non-desirable properties.

Example 2.3. *Consider $d = 100$ corporate bonds with current price 100 for each; default probability is 2% for each bond and the defaults are independent. Each bond pays 105 if no default takes place and no repayment in the case of a default. Each bond i has an associated loss variable L^i with*

$$P(L^i = 100) = 2\%, \quad P(L^i = -5) = 98\%.$$

Consider two portfolios. Portfolio A consists of 100 units of bond 1 (completely concentrated) and Portfolio B is completely diversified: it consists of 1 unit of each bond 1,...,100. From a risk management point of view, it seems to be obvious that one should prefer Portfolio B. But this is not reflected in the VaR:

$$VaR_{0.95}(L^A) = VaR_{0.95}(100L^1) = 100VaR_{0.95}(L^1) = 100(-5) = -500.$$

This means that even after a withdrawal of 500 units risk capital is acceptable under the $VaR_{0.95}$-risk measure. On the other hand we may write

$$Y^i = 100D_i + (-5)(1 - D_i) = 105D_i - 5,$$

where D_i *is Bernoulli-distributed with parameter* 0.02*. Therefore,*

$$VaR_{0.95}(L^B) = VaR_{0.95}\left(\sum_{i=1}^{100} Y^i\right) = VaR_{0.95}\left(105\sum_{i=1}^{100} D_i - 500\right) = 105 VaR_{0.95}\left(\sum_{i=1}^{100} D_i\right) - 500$$

and $\sum_i D_i$ *is* $Bin(100, 0.02)$*-distributed. It holds that*

$$P(\sum_i D_i \le 4) = 0.949..., \quad P(\sum_i D_i \le 5) = 0.984..$$

which yields that $VaR_{0.95}\left(\sum_{i=1}^{100} D_i\right) = 5$ *and therefore*

$$VaR_{0.95}(L^B) = 105 \cdot 5 - 500 = 25 > VaR_{0.95}(L^A). \tag{2.1}$$

This example illustrates that VaR *is not a subadditive in the sense described in Section 2.2 below.*

However, in many important subclasses of loss distributions, VaR *turns out to be subadditive (at least approximately in an appropriate sense).*

2.1.3 Expected Shortfall (ES)

To motivate the next measure, let us return to Example 2.3. One main reason for the observation in (2.1) is that the actual size of large losses was not taken into account. This is circumvented by the expected shortfall:

The expected shortfall at level p stands for the average loss given that the loss exceeds the VaR at the same level p. Formally:

Definition 2.2. *For any random variable* L *with* $E|L| < \infty$ *and any* $p \in (0,1)$ *we define the* expected shortfall *as*

$$ES_p(L) = E(L|L \ge VaR_p(L)) \left(= \frac{E(L1_{\{L \ge VaR_p(L)\}})}{P(L \ge VaR_p(L))}\right).$$

Lemma 2.4. *If* L *has a density* f_L *and distribution function* F_L [3]*, then*

$$\begin{aligned} ES_p(L) &= \frac{1}{1-p}\int_{F_L^{\leftarrow}(p)}^{\infty} x f_L(x)dx \\ &= \frac{1}{1-p}\int_p^1 VaR_y(L)dy. \end{aligned} \tag{2.2}$$

Proof. The first equality directly follows from the definition of the density. For the second one: See project 1. □

[3] The representation (2.2) for the expected shortfall holds under the more general assumption that the distribution function of L is continuous.

Remark 2.5. *1. One immediately sees*

$$ES_p(L) = E(L|L \geq VaR_p(L)) \geq E(VaR_p(L)|L \geq VaR_p(L))) = VaR_p(L).$$

2. *For random variables with continuous distributions, the expected shortfall is monotone, that is, for L, L' with densities it holds that if $L \leq L'$, then $ES_p(L) \leq ES_p(L')$. This follows from Lemma 2.4 as*

$$ES_p(L) = \frac{1}{1-p}\int_p^1 VaR_y(L)dy \leq \frac{1}{1-p}\int_p^1 VaR_y(L')dy = ES_p(L').$$

3. *It is easily seen that the expected shortfall is translation invariant and positive homogenous (use the properties of VaR).*

4. *In many references, formula* (2.2) *is used as a definition for the expected shortfall for all underlying distributions. This is perhaps less intuitive than our definition, but the theory for non-continuous distributions can be developed easier. We keep our definition but only concentrate on continuous distributions in the following.*

2.2 Axiomatic approach to risk measures based on the loss distribution

Rather than considering concrete risk measures, one may also start with desirable properties and investigate their implications.

We consider a general class of loss variables $\mathcal{L}$ and a mapping

$$\rho : \mathcal{L} \to \mathbb{R}.$$

So, what properties should ρ have to be a *good* risk measure?

1. *Translation invariance*:

$$\rho(L + b) = \rho(L) + b \text{ for all } L \in \mathcal{L},\ b \in \mathbb{R}.$$

   If the loss for any position increases by a fixed amount b, we have to alter the required capital by the same amount.

2. *Monotonicity*:

$$L \leq L' \text{ implies } \rho(L) \leq \rho(L') \text{ for all } L, L' \in \mathcal{L}.$$

   Positions with highr losses need more buffer capital.

3. *Positive homogenicity*:

$$\rho(aL) = a\rho(L) \text{ for all } L \in \mathcal{L},\ a \geq 0.$$

   Positions in different currencies need the same buffer capital calculated in those currencies.

4. *Subadditivity*:

$$\rho(L + L') \leq \rho(L) + \rho(L') \text{ for all } L, L' \in \mathcal{L}.$$

   This, in a mathematical formulation, reflects the accepted idea that diversification reduces risk and corresponds to a basic principle of portfolio management. (*Never put all your eggs in one basket.*)

The following definition comes from [ADEH99].

Definition 2.3. *A risk measure ρ satisfying axioms 1.–4. above is called a* coherent risk measure*.*

How do the risk measures above relate to these axioms? Standard deviation fails to be monotone, so it is not a convex risk measure. VaR and ES satisfy the axoims 1.–3. However, subadditivity does not hold for VaR as we observed in Example 2.3. We will study this property for the ES below.

Before coming to this we first show that coherent risk measures can be obtained by using the idea of scenarios.

2.2.1 Scenario-based risk measures and coherency

As before, we consider losses as random variables $L : \Omega \to \mathbb{R}$. In practice, we of course have only imperfect knowledge about the underlying real-world probability measure P.

A *scenario* is another probability measure Q which represents possible outcomes against which we want to precautions (e.g., what happens if Greece defaults, if the Yen goes down by 20% against the Euro,...), and we assume that Q is absolutely continuous w.r.t. P, i.e. for all events $A \in \mathcal{A}$ it holds that if $P(A) = 0$, then $Q(A) = 0$. This means that what is impossible in the real world is also impossible under the scenario risk measure.

Our construction for a risk measure is now the following: We consider a set of scenarios, calculate the expected loss under all scenarios, and take the maximum of these losses as the buffer capital. Formally:

Definition 2.4. *For some set $\mathcal{C}$ of scenarios, we define the corresponding* scenario risk measure $\rho_{\mathcal{C}}$ *by*

$$\rho_{\mathcal{C}}(L) = \sup_{Q \in \mathcal{C}} E_Q(L),$$

where $E_Q(L)$ is the expectation under Q.

This approach indeed leads to coherent risk measure:

Proposition 2.6. *Each scenario risk measure $\rho_{\mathcal{C}}$ a coherent risk measure.*

Proof.

- $\rho_{\mathcal{C}}(L + c) = \sup_{Q\in\mathcal{C}} E_Q(L + c) = \sup_{Q\in\mathcal{C}} E_Q(L) + c = \rho_{\mathcal{C}}(L) + c.$
- If $L \le L'$ then
$$\rho_{\mathcal{C}}(L) = \sup_{Q\in\mathcal{C}} E_Q(L) \le \sup_{Q\in\mathcal{C}} E_Q(L') = \rho_{\mathcal{C}}(L').$$
- for $c > 0$
$$\rho_{\mathcal{C}}(cL) = \sup_{Q\in\mathcal{C}} E_Q(cL) = \sup_{Q\in\mathcal{C}} cE_Q(L) = c\rho_{\mathcal{C}}(L').$$
- $$\rho_{\mathcal{C}}(L + L') = \sup_{Q\in\mathcal{C}}(E_Q(L) + E_Q(L')) \le \sup_{Q\in\mathcal{C}} E_Q(L) + \sup_{Q\in\mathcal{C}} +E_Q(L') = \rho_{\mathcal{C}}(L) + \rho_{\mathcal{C}}(L').$$

□

A mathematical result (using classical functional analysis; Riesz representation) states that – under suitable technical assumptions – also the converse is true: Any coherent risk measure is a scenario risk measure for some suitable set of scenarios. More details are discussed in the following section, see als [ADEH99]. Unfortunately, for general coherent risk measures, this is only an existence result, but there is no algorithm how to find the set of scenarios explicitly.

2.2.2 ES is a coherent risk measure

Now, we will prove that – at least for continuous distributions – the expected shortfall is a coherent risk measure. To this end, we will show that it can be represented explicitly as a scenario risk measure.

Proposition 2.7. *Let $p \in (0, 1)$ and*

$$\mathcal{C} = \{P^B : \ B \in \mathcal{A} \text{ with } P(B) = 1 - p\},$$

where $P^B = P(\cdot|B)$ denotes the conditional probability given B. Then

$$ES_p(L) = \rho_{\mathcal{C}}(L) = \sup_{Q \in \mathcal{C}} E_Q(L)$$

for all random variables L with a continuous distribution.
In particular, ES_p is a coherent risk measure.

Proof. 1. We first prove that

$$(1-p)ES_p(L) = \sup_{B, \mathbb{P}(B) \leq 1-p} E(1_B L).$$

So, let B be an event with $\mathbb{P}(B) \leq 1-p$. By a case analysis, one immediately checks that [4]

$$(L - VaR_p(L))(1_{\{L \geq VaR_p(L)\}} - 1_B) \geq 0,$$

hence

$$E(L1_{\{L \geq VaR_p(L)\}}) - E(L1_B) \geq VaR_p(L)(\mathbb{P}(\{L \geq VaR_p(L)\}) - \mathbb{P}(B)) \geq 0,$$

where the last inequality holds as $\mathbb{P}(\{L \geq VaR_p(L)\}) = 1 - p \geq \mathbb{P}(B)$. This proves

$$E(L1_{\{L \geq VaR_p(L)\}}) \geq \sup_{B, \mathbb{P}(B) \leq 1-p} E(1_B L)$$

with = for $B = \{L \geq VaR_p(L)\}$.

2. Now, the claim follows from the definition of the (elementary) conditional probability.

□

2.3 More on the theory of coherent risk measure*

The aim of this section is to make the statement *Any coherent risk measure is a scenario risk measure for some suitable set of scenarios* precise. To this end, we consider the space

$$\mathcal{M} := L_\infty$$

of all a.s. bounded random variables as our space of loss variables. It is well known from functional analysis that the *topological dual space* of $\mathcal{M}$ is the space b_a of all finitely additive, finite, signed measures, which are absolutely continuous with respect to P. Together with the Hahn-Banach theorem, we obtain the following Lemma (if you do not know about functional analysis, just accept the result; I will give a geometric interpretation in the lecture):

[4] Note that the following argument is very similar to the one in the Neyman-Pearson-Lemma in mathematical statistics.

Lemma 2.8. *Let $\mathcal{U} \subseteq \mathcal{M} = L_\infty$ be convex and open, and $L_0 \notin \mathcal{U}$. Then, there exists $Q \in b_a$ such that*

$$E_Q(L) < E_Q(L_0) \text{ for all } L \in \mathcal{U},$$

where E_Q denotes the integral w.r.t. the finitely additive measure Q[5].

Proof. This follows from standard facts in functional analysis, see, for example, Dunford-Schwartz, Linear Operators I, page p. 214 and page 296. □

This abstract mathematical result now leads to the representation result for all coherent risk measures on $\mathcal{M}$ as scenario risk measures:

Theorem 2.1. *Let ρ be any coherent risk measure on $\mathcal{M}$. Then there exists a set $\mathcal{C} \subset b_a$ of finitely additive probability measures*[6] *(*scenarios*) such that for all $L \in \mathcal{M}$*

$$\rho(L) = \sup_{Q \in \mathcal{C}} E_Q(L).$$

Proof.

1. It is enough to prove the following: For all $L_0 \in \mathcal{M}$ exists a finitely additive probability measure $Q \in b_a$ such that

$$E_Q(L) \leq \rho(L) \text{ for all } L, \quad E_Q(L_0) \leq \rho(L_0),$$

because then the choice

$$\mathcal{C} = \{Q : E_Q(L) \leq \rho(L) \text{ for all } L\}$$

is as desired.

2. So let $L_0 \in \mathcal{M}$ be arbitrary. By considering $L_0 - \rho(L_0) + 1$ instead of L_0, we can assume w.l.o.g. that $\rho(L_0) = 1$. Now write

$$\mathcal{U} = \{L \in \mathcal{M} : \rho(L) < 1\}.$$

As ρ is a convex function (by positive homogenicity and subadditivity), the set $\mathcal{U}$ is convex. It is furthermore open. For a proof of this fact, let $L \in \mathcal{U}$ and choose $\epsilon := 1 - \rho(L)$ and let $Z \in \mathcal{M}$ be such that $\|Z - L\|_\infty < \epsilon$. Then,

$$\rho(Z) = \rho(Z - L + L) \leq \rho(\|Z - L\|_\infty + L) = \|Z - L\|_\infty + \rho(L) < 1,$$

i.e. $Z \in \mathcal{U}$.

[5] One can define the expectation for these finitely-additive measures in the usual way

[6] i.e. each $Q \in \mathcal{C}$ fulfills: $Q \geq 0$, $Q(\Omega) = 1$

3. Now, we can apply Lemma 2.8 and obtain $Q \in b_a$ such that $E_Q(L) < E_Q(L_0)$ for all $L \in \mathcal{U}$. We show now, that this is as desired in 1:

4. First, note that as $0 \in \mathcal{U}$, we obtain that $0 = E_Q(0) < E_Q(L_0)$. By considering $Q/Q(L_0)$ instead, we can assume w.l.o.g. that $E_Q(L_0) = 1$. This already shows that $E_Q(L_0) = \rho(L_0)$.

5. We now prove that $E_Q(L) \leq \rho(L)$ for arbitrary L. Let c be such that $\rho(L) < c$. Then

$$\rho(L - c + 1) \leq \rho(L) - c + 1 < 1,$$

i.e. $L - c + 1 \in \mathcal{U}$. By 3., we obtain $E_Q(L) - c + 1 = E_Q(L - c + 1) < E_Q(L_0) = 1$, i.e. $E_Q(L) < c$. As c was arbitrary, this shows $E_Q(L) \leq \rho(L)$.

6. It remains to be proved that Q is a non negative probability measure. We first show that Q is nonnegative by showing that for each $L \geq 0$ it holds that $E_Q(L) \geq 0$. So, let $L \geq 0$ and $c > 0$. Then $\rho(-cL) \leq \rho(0) = 0 < 1$, hence as above $-cE_Q(L) = E_Q(-cL) < 1$, so that $E_Q(L) > -1/c$. As c was arbitrary, we get the positivity of Q.
It remains to be proved that $E_Q(1) = 1$. Let $c > 0$. If $c < 1$, then $\rho(c) = c < 1$ and therefore $cE_Q(1) = E_Q(c) < 1$. As c was arbitrary, $E_Q(1) \leq 1$. On the other hand, if $c > 1$, then

$$\rho(2L_0 - c) = 2\rho(L_0) - c = 2 - c < 1,$$

hence $2 - cE_Q(1) = E_Q(2 - cL_0) < 1$, i.e. $E_Q(1) \geq 1/c$. As above, as c was arbitrary, $E_Q(1) \geq 1$, and we have equality as claimed.

□

Chapter 3

Data-driven Risk Management

In the previous chapter, we have studied risk measures basically from a purely probabilistic point of view. To use risk measures in practice, it is of course essential to know the distribution of the loss variable L in our model explicitly, or at least to know relevant quantities that are necessary to calculate the risk measure. Here, statistics comes into play.

Our main examples of risk measures – VaR_p and ES_p – are by definition based on the large losses (typically, p close to 1), i.e. on extremal events. As such events are rare we have little data material on them, so estimation will only have limited accuracy. This is the major challenge in this field of statistics, Extreme-Value (EV) statistic. The aim of this field is to develop a theory to make best possible use of the available data to provide estimates for VaR_p and ES_p.

But first, we will introduce the general notational framework and summarize some general statistical background:

3.1 Notations and Framework

We start with a portfolio value $V(t)$ at time t. Looking Δt time units into the future, the value is $V(t+\Delta t)$, so the loss from t to $t+\Delta t$ is

$$L_{(t,t+\Delta t]} = V(t) - V(t+\Delta t).$$

We consider a discrete time model with equidistant time points starting in 0, we have time points $t_m = m\Delta t$ (Δt could be one day, 10 days, a month,...). Then we have losses

$$L_{m+1} := L_{(m\Delta t,(m+1)\Delta t]} = V(m\Delta t) - V((m+1)\Delta t).$$

3.1.1 Functional modeling

We model the portfolio value as a function of time and the risk factors

$$V_m := V(t_m) = f(t_m, Z_m)$$

with random *risk factors* $Z_m = (Z_{m,1}, ..., Z_{m,d})$. Here, Z_m is a random variable with values in $\mathbb{R}^d$. Writing $X_{m+1} = Z_{m+1} - Z_m$ for the risk factor changes, we have

$$L_{m+1} = f(t_m, Z_m) - f(t_{m+1}, Z_m + X_{m+1}).$$

It is often convenient to introduce the so-called *loss operator*[1]

$$l_{[m]} : \mathbb{R}^{d+1} \to \mathbb{R}, \quad l_{[m]}(x) = f(t_m, Z_m) - f(t_{m+1}, Z_m + x).$$

3.1.2 Interpretation

If m is the current time point and $m+1$ is the next time point of interest in the future, then the risk factors Z_m at time m are know, so $l_{[m]}$ is totally determined by the (random) risk factor changes X_{m+1}.

3.1.3 Linearization

The loss operator is often a hard to deal with as it is a complex function of the risk factors. One way out is an approximation:
Using Taylor expansion of order 1, one can write

$$f(t_{m+1}, Z_m + x) = f(t_m, Z_m) + f_t(t_m, Z_m)\Delta t + \sum_{i=1}^{d} f_i(t_m, Z_m)x_i + \text{ higher order terms.}$$

This motivates to consider the 1st order approximate operator

$$L^{\Delta}_{m+1} := -\left(f_t(t_m, Z_m)\Delta t + \sum_{i=1}^{d} f_i(t_m, Z_m)X_{m+1,i} \right)$$

with corresponding loss operator $l^{\Delta}_{[m]}$.

3.1.4 Example (for those who is familiar with option pricing)

Assume that the portfolio consists of a European call option on a stock S with maturity T, strike K, i.e. at time T you get the payoff

$$\max(S_T - K, 0).$$

[1]note that the loss operator is formally a random operator as it depends on Z_m

Assuming a Black-Scholes model, the fair price at time t can written in the form

$$C(t, S_t, r_t, \sigma_t),$$

where r_t is the interest rate at t and σ_t the volatility. Hence, using the risk factors

$$Z_t = (\log S_t, r_t, \sigma_t)$$

we have

$$X_{m+1} = (\log(S_{m+1}/S_m), r_{m+1} - r_m, \sigma_{m+1} - \sigma_m),$$
$$V_m = f(t_m, Z_m) = C(t_m, S_m, r_m, \sigma_m).$$

The function C here is a function involving the Gaussian cdf Φ. Appliying the linearization as described before, the corresponding coefficients are known as the *Greeks* of the option.

3.1.5 The statistical problem

Consider a portfolio with value $V_m = f(t_m, Z_m)$, where f is assumed to be a known function and the future losses

$$L_{m+1} = l_{[m]}(X_{m+1}).$$

We furthermore assume the availability of historical data

$$z_m, z_{m-1}, ..., z_{m-n}.$$

The statistical question we are faced with is: How do we use these data to find reliable estimators for VaR, ES,... .

3.2 Empirical distribution function and empirical VaR, ES

As it is difficult to specify a suitable parametric family of distributions in risk modeling, one often uses methods from nonparametric statistics.

Let $X_1, \cdots, X_n$ be iid real valued random variables and let F denote their distribution function.

1. The standard nonparametric estimator for the distribution function is the *empirical distribution function* F_n

$$F_n(x) = \frac{1}{n}|\{i \leq n : X_i \leq x\}|.$$

2.
$$q_p(F_n) := F_n^{\leftarrow}(p) = \inf\{t : F_n(t) \geq p\}$$
is called *empirical quantile function* and is the standard estimator for the quantile function.

3. A reason for using F_n and $F_n^{\leftarrow}$ as standard estimators is that they are consistent in an appropriate sense. Indeed, the Glivenko-Cantelli Theorem states that
$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 \text{ a.s. as } n \to \infty.$$

4. $VaR_p(F_n)$, also called *empirical VaR*, and $ES_p(F_n)$, the *empirical ES*, are estimators for the VaR and the ES. But these estimators have certain drawbacks. For example, both empirical VaR and ES are bounded above by $\max_i X_i$. So, you never expect higher losses than the highest loss in the past. The other estimators discussed below do not have this property.

5. The empirical VaR and ES can be expressed using *order statistics*. Considering observation $x_1 = X_1(\omega), \ldots, x_n = X_n(\omega)$ (where we assume for simplicity that $x_i \neq x_j$ for $i \neq j$) we can order them in increasing order:
$$x_{1:n} < \cdots < x_{n,n}.$$
Then $F_n(x_{k:n}) = \frac{k}{n}$ and
$$\begin{aligned} F_n^{\leftarrow}(x_{k:n}) &= \inf\{x : F_n(x) \geq p\} \\ &= \inf\{x_{k:n} : \frac{k}{n} \geq p\} \\ &= \inf\{x_{k:n} : k \geq pn\} \\ &= x_{\lceil pn \rceil : n}, \end{aligned}$$
where $\lceil y \rceil = \inf\{k \in \mathbb{Z} : k \geq y\}$. So
$$VaR_p(F_n) = x_{\lceil pn \rceil : n},$$
and similarly
$$ES_p(F_n) = \frac{\sum_{k=\lceil pn \rceil}^{n} x_{k:n}}{n - \lceil pn \rceil + 1}.$$

3.3 Historical simulation

Assume now that the present time point is t_m and we have the data of the risk factors in the past till t_{m-n}:
$$z_{m-n}, \ldots, z_m, \quad x_{m-n+1} = z_{m-n+1} - z_{m-n}, \ldots, x_m = z_m - z_{m-1}.$$

Using this data, we want to approximate the distribution of

$$L_{m+1} := l_{[m]}(X_{m+1}),$$

where $l_{[m]} = l_{[m]}(X_{m+1})$ is the loss operator depending on the unknown future risk factor changes X_{m+1}.
We now obtain the historical losses by applying the loss operator to the known historical risk factor changes

$$l_k = l_{[m]}(x_k), \quad k = m - n + 1, ..., m.$$

We use these data to get estimators for the empirical distribution function and corresponding

$$\widehat{VaR_p} = l_{\lceil pn \rceil:n}$$
$$\widehat{ES_p} = \frac{\sum_{k=\lceil pn \rceil}^{n} l_{k:n}}{n - \lceil pn \rceil + 1}$$

This method is called *historical simulation*. It is simple, but needs sufficient quantities of historical data for all risk factors.

3.4 Variance-Covariance method

The Variance-Covariance method is (among others) based on the assumption that the changes of the risk factors have a multivariate normal distribution. I will (perhaps) come back to this in Chapter 5.

3.5 Monte Carlo Simulation

The general idea can be described as follows: First, take a parametric model for the distribution W_θ of the risk factor change X_{m+1} depending on a certain unknown parameter θ. Use the historical data $x_{m-n+1}, ..., x_m$ to estimate θ by $\hat{\theta}$. In many cases, we will not be able to explicitly compute our risk measure with regard to the distribution $l_{[m]}(X_{m+1})$ (w.r.t $W_{\hat{\theta}}$). This problem is now circumvented using Monte Carlo simulation:
Use a computer to obtain realizations $\tilde{x}_1, ...\tilde{x}_N$ of a sequence $\tilde{X}_1, \tilde{X}_2, ...$ of iid random variables w.r.t. $W_{\hat{\theta}}$. This given values

$$\tilde{l}_1 = l_{[m]}(\tilde{x}_1), ..., \tilde{l}_N = l_{[m]}(\tilde{x}_N).$$

Then we can compute risk measures with respect to the empirical distribution for $\tilde{l}_1, .., \tilde{l}_N$; e.g.

$$\widehat{VaR_\alpha} = \tilde{l}_{\lceil \alpha n \rceil}$$

$$\widehat{ES}_\alpha = \frac{\sum_{k=\lceil \alpha n \rceil}^{n} \tilde{l}_{k:n}}{n - \lceil \alpha n \rceil + 1}$$

Chapter 4

Tools from Extreme Value Statistics

Now, we go deeper into the statistics for risk measures. But in this chapter, the distribution of extreme events plays a crucial role, as most risk measures used in practice are based on these.

4.1 Regular variation

When dealing with rare events based on the tails of the distribution, it is important to study distributions where such events are not too unlikely to occur in practice. This is often made precise by considering laws with *heavy tails*, which are laws that assign more mass to extreme events than any exponential distribution, i.e.

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{\mathrm{e}^{-\lambda x}} = \infty \text{ for all } \lambda > 0,$$

where $\overline{F}(x) = 1 - F(x) = P(L > x)$ denotes the so-called survival function of L. The heaviness of the tails can be quantified by a number α such that $\overline{F}(t) \approx t^{-\alpha}$ for t large. More precisely:

Definition 4.1. *A function $L : (0,\infty) \to (0,\infty)$ is called* slowly varying *($L \in RV_0$) if*

$$\lim_{t\to\infty} \frac{L(tx)}{L(t)} = 1 \textit{ for all } x > 0.$$

A random variable with cdf F is called regularly varying with index α *if*

$$\overline{F}(t) = L(t)t^{-\alpha}$$

for some slowly varying function L.

Remark 4.1. 1. *A natural example of a random variable that is regularly varying with index α is given by the distribution function $F(t) = 1 - t^{-\alpha}$ for $t > 1$ ($L(t) = 1$); see also Subsection 4.3.1.*
Another important example is the Student-t-distribution with parameter ν, i.e. with density of the form $f_\nu(t) = const_\nu(1 + \frac{t^2}{\nu})^{-\frac{\nu+1}{2}}$. It can be shown (using Karamata's theorem below) that the tail index of this distribution is ν.

2. *If X is regularly varying with index α then $E((X^+)^\beta) = \infty$ for $\beta > \alpha$ and $< \infty$ for $\beta < \alpha$.*

4.2 Hill estimator

One approach for to obtain estimators for VaR_p and ES_p is to assume that the loss variable L is regularly varying with (unknown) index α. If one has a good estimator $\hat{\alpha}$ for α, then (when p is large enough) one can assume that the tails of $\overline{F}$ nearly coincide $t^{-\hat{\alpha}}$. This then leads to good estimators for VaR_p and ES_p. One such method is based on *Karamata's Theorem* for slowly varying functions L, which states that for $\beta < -1$

$$\frac{\int_t^\infty L(x)x^\beta dx}{L(t)t^{\beta+1}} \to -\frac{1}{\beta+1} \text{ as } t \to \infty.$$

(You can immediately check this formula for constant L).

The important observation is that the RHS is independent of the function L. This is the main ingredient for constructing an estimator for α:

1. Using partial integration for the Stieltjes Integrals [1]

$$\begin{aligned}
&\frac{1}{\overline{F}(t)}\int_t^\infty (\log(u) - \log(t))dF(u) = \frac{1}{\overline{F}(t)}\int_t^\infty (\log(u) - \log(t))d(-\overline{F})(u) \\
=&\frac{1}{\overline{F}(t)}\left([(\log(u) - \log(t))(-\overline{F}(u))]_t^\infty + \int_t^\infty \frac{\overline{F}(u)}{u}du\right) \\
=&\frac{1}{L(t)t^{-\alpha}}\int_t^\infty \frac{L(u)u^{-\alpha-1}}{u}du \\
\to&\frac{1}{\alpha}.
\end{aligned}$$

[1]

$$\int_a^b G(x)dH(x) = \int_a^b H(x)dG(x) = [G(x)H(x)]_a^b,\ H(x) = -\overline{F}(x),\ G(x) = \log(x) - \log(t)$$

2. From this, we can obtain an estimator for α: Replace $\overline{F}$ by the empirical distribution function $\hat{F}_n$ and t by some *large* observed value $x_{k:n}$ in the formula above. Then, to find the estimator, we use the (obviously somewhat unclear) notation $\sim_{hct}$ to denote steps where we hope that the expressions are close in some reasonable sense (htc: hopefully close to). This is used only to find a candidate for a good estimator, but is of course no proof for anything. This can than be carried out in a second step (but not in this course).

$$\begin{aligned}\frac{1}{\alpha} &\sim_{hct} \frac{1}{1-\hat{F}_n(x_{k:n})}\int_{x_{k:n}}^{\infty}(\log(u)-\log(x_{k:n}))d\hat{F}_n(u)\\ &= \frac{1}{1-k/n}\sum_{j=k+1}^{n}(\log(x_{j:n})-\log(x_{k:n}))\frac{1}{n}\\ &= \frac{1}{n-k}\sum_{j=k+1}^{n}(\log(x_{j:n})-\log(x_{k:n}))\end{aligned}$$

This yields the *Hill estimator* for α:

$$\hat{\alpha}_{k,n} = \left(\frac{1}{n-k}\sum_{j=k+1}^{n}(\log(x_{j:n})-\log(x_{k:n}))\right)^{-1}.$$

One can indeed prove that this estimator is consistent. More precisely, if $k = k(n)$ is such that

$$n-k(n)\to\infty,\quad \frac{n-k(n)}{n}\to 0, \tag{4.1}$$

then $\hat{\alpha}_{k(n),n}\to\alpha$ in probability, see ... .

Note that the Hill estimator does not make sense for negative values of x as one has to take the logarithm. However, note that we are only interested in the tail behavior, so that the Hill estimator is not needed at all for these values.

3. As we assume that our loss variable is regularly varying of some (unknown) parameter α, we have that

$$\frac{\overline{F}(ty)}{\overline{F}(t)} = t^{-\alpha}\frac{L(ty)}{L(t)} \approx t^{-\alpha}$$

as $t\to\infty$, we have

$$\begin{aligned}\overline{F}(t) = \overline{F}(\frac{t}{x_{k:n}}x_{k:n}) &\sim_{hct} \left(\frac{t}{x_{k:n}}\right)^{-\alpha}\overline{F}(x_{k:n})\\ &\sim_{hct} \left(\frac{t}{x_{k:n}}\right)^{-\hat{\alpha}_{k,n}}(1-\hat{F}_n(x_{k:n}))\\ &= \frac{n-k}{n}\left(\frac{t}{x_{k:n}}\right)^{-\hat{\alpha}_{k,n}}\end{aligned}$$

From this we obtain an *estimator for the VaR* as follows:

$$\begin{aligned} VaR_p(L) &= \inf\{t : \overline{F}(t) \le 1-p\} \\ &\sim_{hct} \inf\{t : \frac{n-k}{n}\left(\frac{t}{x_{k:n}}\right)^{-\hat{\alpha}_{k,n}} \le 1-p\} \\ &= \left[\frac{n}{n-k}(1-p)\right]^{-1/\hat{\alpha}_{k,n}} x_{k:n}. \end{aligned}$$

Note that this estimator can also be applied to levels p where no empirical data is available. Similarly, an estimator for ES can be found (see Excercises).

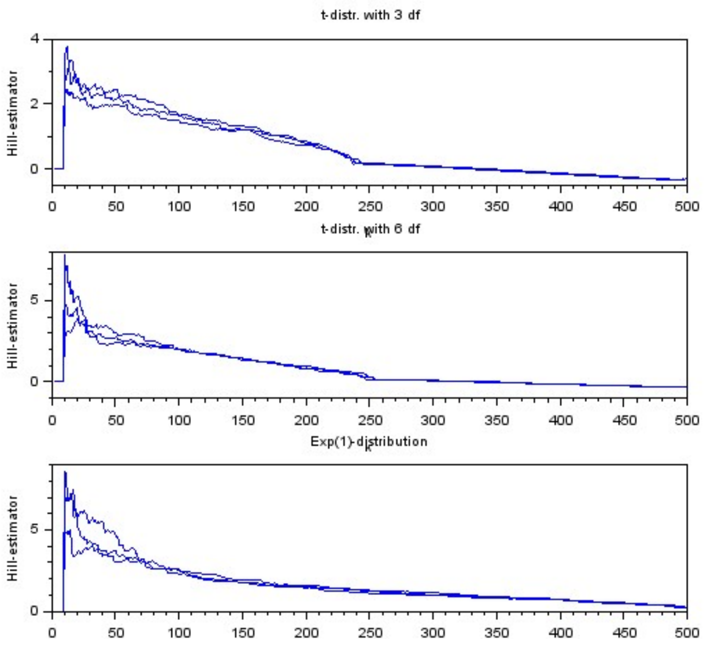
4. For using the Hill estimator, the choice of the parameter k is crucial. (4.1) gives some insight, but it does not really lead to a concrete choice. The general problem is that if k is too large we are left with too little observations, if k is too small we are using observations which are too far away from the tails.

   A practical answer is the following: Look at the *Hill plot*

   $$\{(k, \hat{\alpha}_{k,n}), k = 1, 2, \ldots, n-1\}$$

   and choose k such that the plot stabilizes.

Remark 4.2. *Although the method described in this subsection has a good theoretical foundation, one often has problems with it in practice: there are several examples of very unstable plots (→ Hill horror plot), see also the technical project.*



4.3 Peak-Over-Threshold (POT) and Pareto distributions

As before, we have data $x_1, ..., x_n$ from random variables $X_1, ..., X_n$ which are regularly varying with some (unknown) parameter α.

4.3.1 Motivation

For a fixed $u \in \mathbb{R}$ and a random variable X the overshoot distribution for X over u is given as

$$F_u(x) = P(X - u \leq x | X > u),\ x \geq 0,$$

with a clear probabilistic interpretation. In the case that X is regularly varying with some parameter α we have for large u

$$\begin{aligned}\overline{F}_u(x) &= \frac{\overline{F}(u+x)}{\overline{F}(u)} = \frac{\overline{F}(u(1+x/u))}{\overline{F}(u)} \\ &= \frac{(u(1+x/u))^{-\alpha}}{u^{-\alpha}} \frac{L(u(1+x/u))}{L(u)} \\ &\sim (1+x/u)^{-\alpha}.\end{aligned}$$

More precisely, one can prove that

$$\lim_{u\to\infty} \sup_{x>0} |\overline{F}_u(x) - (1+x/u)^{-\alpha}| = 0.$$

That is, for large u the overshoot distribution of an arbitrary regularly varying random variable with parameter α is close to a distribution with cdf

$$1 - (1+x/u)^{-\alpha}.$$

This class of distributions is known in probability as the *(generalized) Pareto distribution.* More precisely, the Pareto distribution with parameters $\gamma,\ \beta > 0$ has distribution function

$$G_{\gamma,\beta}(x) = 1 - (1 + \frac{\gamma x}{\beta})^{-1/\gamma}$$

which coincides with the parametrization above for $\gamma = 1/\alpha,\ \beta = u/\alpha$.

As α is unknown in our setting, so are $\gamma,\ \beta$. But we are faced with a parametric problem that can be tackled using standard methods. The plan is now the following:

1. Find a method for choosing a sufficiently large threshold u (Subsection 4.3.2)

2. Use the *excesses over u* $Y_1, \ldots, Y_{N_u}$ given by $Y_i = X_i - u$ to estimate the parameters γ, β of the corresponding Pareto distribution (Subsection 4.3.3). Here and in the following

$$N_u = |\{i : X_i > u\}|$$

denotes the *number of exceedences* of u.

3. Use these to come to estimators for VaR and ES (Subsection 4.3.4).

4.3.2 Mean-Excess plot

Similar to the situation for the Hill estimator, one main problem in the program outlines above is a suitable choice of u:

- If u is chosen too large, one does not take enough observations into account,
- if u is too small, the approximation $\overline{F}_u \approx \overline{G}_{\gamma,\beta}$ is questionable.

Often, the following graphical method is useful. The basic quantity here is *mean excess function*, which is for a random variable X defined as

$$e(u) := E(X - u|X > u).$$

In the case that X has a Pareto distribution with parameters $\gamma < 1, \beta$, it is an easy excercise to see that

$$e(u) = \frac{\beta + \gamma u}{1 - \gamma}.$$

Here, the important observation is that this is a (affine) linear function. The Mean-Excess plot is now used as follows:

1. Use the natural estimator

$$\hat{e}_n(u) := \frac{1}{N_u} \sum_{k=1}^{n} (x_k - u)^+ \quad {}^2$$

for $e(u)$ to build up the *Mean-Excess plot*

$$\{(x_{k:n}, \hat{e}_n(x_{k:n})) : \; k = 1, ..., n-1\}.$$

2. Choose u in such a way that for $x_{k:n} > u$ the plot is approximately linear and there are enough data points left (if possible).

[2] $z^+ := \max(z, 0)$

4.3.3 Estimation of the parameters for the Pareto distribution

Let us assume that $X_1, ..., X_n$ are iid with cdf F such that

$$\overline{F}_u \approx \overline{G}_{\gamma,\beta}$$

for some unknown parameters α, β for u sufficiently large as it is the case for regularly varying variables with parameter > 0. Then the excesses $Y_1, ..., Y_{N_u}$ are also iid and their distribution is approximately $G_{\gamma,\beta}$. Taking the derivative one immediately obtains that the pdfs are equal

$$G'_{\gamma,\beta}(y) = \frac{1}{\beta}(1 + \frac{\gamma y}{\beta})^{-1/\gamma-1}.$$

Therefore, the joint likelihood function of $Y_1, ..., Y_{N_u}$ is (approximately) given by

$$L(\gamma, \beta; y_1, ..., y_{N_u}) = \prod_{i=1}^{N_u} \frac{1}{\beta}(1 + \frac{\gamma y_i}{\beta})^{-1/\gamma-1}.$$

(After taking logarithm) this quantity can be maximized numerically in γ, β, which yields the MLE $\hat{\gamma}, \hat{\beta}$.

4.3.4 Estimation of VaR and ES

Using $\hat{\gamma}, \hat{\beta}$ we can now define corresponding estimators for VaR and ES. We start with a estimator for $\overline{F}$. We have

$$\overline{F}(u) \sim \hat{F}_n(u) = \frac{N_u}{n}$$

and

$$\overline{F}_u(x) \sim_{hct} \overline{G}_{\gamma,\beta}(x) \sim_{hct} \overline{G}_{\hat{\gamma},\hat{\beta}}(x) = (1 + \hat{\gamma}x/\hat{\beta})^{-1/\hat{\gamma}}.$$

To estimate $\overline{F}(u + x)$ we use

$$\overline{F}(u + x) = \overline{F}(u)\overline{F}_u(x)$$

to obtain the estimator

$$\hat{\overline{F}}(u + x) := \frac{N_u}{n}(1 + \hat{\gamma}x/\hat{\beta})^{-1/\hat{\gamma}}.$$

Interpretation: From the threshold u on, we are confident that the data are represented by a Pareto distribution, so we use the estimators available for this distribution. This leads to a reasonable estimate for $\overline{F}(u+x)$ even if all observations are smaller than $u+x$, i.e. $\hat{\overline{F}}_n(u + x) = 0$.

This immediately leads to an estimator for the VaR:

$$\begin{aligned} VaR_p &= \inf\{y : \overline{F}(y) \leq 1-p\} \\ &= \inf\{u+x : \overline{F}(u+x) \leq 1-p\} \\ &\sim_{hct} \inf\{u+x : \hat{\overline{F}}(u+x) \leq 1-p\} \\ &= u + \inf\{x : \frac{N_u}{n}(1+\hat{\gamma}x/\hat{\beta})^{-1/\hat{\gamma}} \leq 1-p\} \end{aligned}$$

yielding

$$\widehat{VaR}_p = u + \frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\gamma}} - 1\right).$$

Similarly, the ES can be estimated as

$$\widehat{ES}_p = \widehat{VaR}_p + \frac{\hat{\beta} + \hat{\gamma}(\widehat{VaR}_p - u)}{1-\hat{\gamma}}.$$

Chapter 5

Short Introduction to Multivariate Distributions and Copulas*

Although the loss of a portfolio is a one-dimensional quantity, the underlying risk factors are typically multidimensional (including, e.g., interest rates, prices of different stocks, macroeconomic factors,...). So, we have to consider d-dimensional, $d \geq 2$, random variables. The distribution of such variables is often called *multivariate distribution*.

So, let $X = (X_1, \dots, X_d)^T$ be a d-dimensional random variable. In Risk Management, the variables $X_1, \dots, X_d$ are typically <u>not</u> independent as the different components of risk factor changes are interacting.

5.1 Reminder: Multivariate distributions

The *distribution function* F of X is given by

$$F(x) = F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

The corresponding marginal cdf of X_i can be calculated as

$$F_i(x_i) = P(X_i \leq x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty).$$

Using the fundamental theorem of calculus, one gets the corresponding identities for densities, whenever they exist. <u>If</u> the components X_i are independent, then $F(x) = F_1(x_1) \cdot ... \cdot F_d(x_d)$, but this is not the interesting case for us.

The *characteristic function* ϕ_X of X is given by

$$\phi_X : \mathbb{R}^d \to \mathbb{C},\ t \mapsto E(e^{it^T X}) = E(e^{i \sum_i t_i X_i})$$

and it also characterizes the distribution of X uniquely.

5.2 Some classes of distributions important in Risk Management

One well-known multivariate distribution is the d-dimensional normal distribution. But for many problems in risk management, more general distributions are needed. One reason is that the normal distribution does not have heavy tails. Elliptical distributions form a suitable class.

5.2.1 Reminder: d-dimensional normal distribution

X is a *multivariate normal distribution* with mean $\mu \in \mathbb{R}^d$ and (positive semidefinite) covariance matrix $\Sigma = (\sigma_{i,j})_{i,j=1,..,d}$, $X \sim N_d(\mu, \Sigma) = N(\mu, \Sigma)$, if the characteristic function is given by

$$\phi_X(t) = \exp\left(it^T\mu - \frac{1}{2}t^T\Sigma t\right), \quad t \in \mathbb{R}^d.$$

If Σ is non-singular, the density f_X is given by

$$f_X(t) = \frac{1}{\sqrt{(2\pi)^d |\det(\Sigma)|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

Some important properties are the following:

1. μ is the expectation of X and Σ is the covariance matrix.

2. If $X \sim N(\mu, \Sigma)$, $b \in \mathbb{R}^k$, $B \in \mathbb{R}^{k\times d}$, then $BX + b \sim N(B\mu + b, B\Sigma B^T)$.

3. If $\mu = 0$ and Σ is the unit matrix, i.e. X is standard normal, then $X_1, \ldots, X_d$ are independent (1-dim.) standard normals.

4. Properties 2 and 3 above lead to a method for simulating the $N(\mu, \Sigma)$-distribution: find $A \in \mathbb{R}^{d\times k}$ such that $A^T A = \Sigma$ (Cholesky decomposition), simulate k independent standard normal variables $Y_1, \ldots, Y_d$, and define

$$X := AY + \mu.$$

5. If X is standard normal and V is an orthogonal matrix ($V^T V = I$), then the rotated random variable VX is also standard normal.

5.2.2 Spherical distribution

We use observation 5 on rotation invariance of the standard normal distribution above for the first generalization. We call a general random variable X *spherical distributed* if for each orthogonal matrix V (i.e. $V^T V = I$) X has the same law as VX.

Each spherical distributed random variable X can be decomposed as

$$X = RS \text{ in distribution,}$$

where S has a uniform distribution on the sphere, $R \geq 0$, and R, S are independent. Therefore, simulating X is easy: Simulate S by simulating a multidimensional standard normal variable Y and define $S = Y/\|Y\|$ and simulate R independently.

One important example is the multivariate t-distribution, which is given as follows: Let Y be a multidimensional standard normal random variable and $S \sim \chi^2_\nu$ be independent[1]. Then $S := \sqrt{m}\frac{Z}{S}$ is called *standard t-distribution with m degrees of freedom*, denoted by $t_{d,\nu}(0, I_d)$. This is an example of a spherical distribution with heavy tails.

5.2.3 Elliptical distribution

As in the case of multivariate normal distributions, one can generalize spherical distributions as follows:

Start with a random variable Y with a spherical distribution on $\mathbb{R}^k$ and let $b \in \mathbb{R}^d$, A a $d \times k$-matrix, then the distribution of

$$X = b + AY$$

on R^d is called elliptical. If $Y \sim t_{k,\nu}(0, I_k)$, then the distribution of X is called *t-distribution with k degrees of freedom, mean b and dispersion matrix $\Sigma = AA^T$*, denoted by $t_{d,\nu}(b, \Sigma)$. In some literature, the parameter Σ is substituted by $\frac{\nu}{\nu-2}\Sigma$, as this matrix contains the covariances.

5.3 Copulas

In risk management, one often has better knowledge on the marginal distributions of the individual risk factors than on their dependence structure. The aim of this section is to

[1] recall that the χ^2_ν distribution is the distribution of $G_1^2 + ... + G_\nu^2$, where G_i are iid standard normal random variables

develop a tool that makes it possible to first model the marginal distributions and, in a second step, take care of the dependence structure. This approach plays a major role whenever the dependence of different variables is of importance (which is the case very often).

5.3.1 Definition and properties

The right notion for describing general dependence structures is the distribution function of a uniform variable on a cube:

Definition 5.1. *Let* $X = (X_1, \ldots, X_d)^T$ *be a random variable such that the components* X_i *are all uniformly distributed on* $[0,1]$*. The distribution function*

$$C : [0,1]^d \to [0,1]$$

is called a copula*.*

In more analytical terms, a copula is a distribution function with the additional property

$$C(1, ..., 1, u_i, 1, ..., 1) = u_i \text{ for all } i = 1, ..., d,\ u_i \in [0,1].$$

The basic result that links general multidimensional distributions and copulas is *Sklar's Theorem*:

Theorem 5.1. *Let* X *be a* d*-dimensional random vector with distribution function* F *and marginal distribution functions* F_i*,* $i = 1, ..., d$*. Then there exists a copula* C *such that*

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)) \text{ for all } x_1, ..., x_d. \tag{5.1}$$

If all F_i *are continuous, the* C *is unique and is then called the* copula of X*.*

Sklar's Theorem gives rise to a method to model the marginal distributions and, in a second step, take care of the dependence structure as follows:

Start with suitable marginal distribution functions F_i, $i = 1, ..., d$. Then specify a copula C describing the dependence structure and define F by (5.1). One can generate a random variable X with this distribution by first generating $(U_1, ..., U_d)$, U_i uniform, with distribution function C and then set

$$X = (F_1^{\leftarrow}(U_1), ..., F_d^{\leftarrow}(U_d)).$$

One important property is that the copula is invariant under increasing transformations in the coordinates. More precisely:

Proposition 5.2. *Let X have continuous marginal distributions and copula C. Let $T_1, ..., T_d : \mathbb{R} \to \mathbb{R}$ be strictly increasing functions. Then*

$$T(X) = (T_1(X_1), ..., T_d(X_d))$$

also has copula C.

Proof. The random variable $\tilde{X}_i := T_i(X_i)$ has distribution function

$$\tilde{F}_i(x) = P(T_i(X_i) \leq x) = P(X_i \leq T_i^{-1}(x)).$$

Therefore,

$$\begin{aligned} C(u_1, .., u_d) &= P(F_1(X_1) \leq u_1, ..., F_d(X_d) \leq u_d) \\ &= P(F_1(T_1^{-1}(T_1(X_1))) \leq u_1, ..., F_d(T_d^{-1}(T_d(X_d))) \leq u_d) \\ &= P(\tilde{F}_1(\tilde{X}_1) \leq u_1, ..., \tilde{F}_d(\tilde{X}_d) \leq u_d) \end{aligned}$$

and this is the copula associated with $T(X)$. □

5.3.2 Gaussian and t-copula

Let X be $N_d(\mu, \Sigma)$-distributed. By considering the strictly increasing transforms $T_i(X_i) = (X_i - \mu_i)/\sqrt{Var(X_i)} \sim N(0, 1)$, one obtains that

$$T(X) \sim N_d(0, R), \text{ where } R \text{ is the correlation matrix of } X.$$

By the previous proposition, the copulas of X and $T(X)$ coincide. This shows that the copula is fully described by the correlation matrix R. This copula is called *Gaussian copula C_R^G for R*:

$$C_R^G(u_1, ..., u_d) = \Phi_R(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)). \tag{5.2}$$

There is no simple explicit expression for C_R^G in general, but it is easy to simulate from C_R^G following the steps:

- Compute the Cholesky decomposition $R = AA^T$.
- Simulate $Z_1, \ldots, Z_d \sim N(0, 1)$, independent.
- Set $X = AZ$.
- Set $U_k = \Phi(X_k)$, then $U = (U_1, \ldots, U_d)$ has distribution C_R^G.

However, note that the dependence structure is fully determined by the correlations. In practice, this is often not appropriate as the real dependence structure is much more complex.

Roughly speaking all elliptical copulas share the structure of (5.2). In particular, the *t-copula* is given by

$$C^t_{R,\nu}(u_1, ..., u_d) = t_{\nu,R}(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d)),$$

where $t_{\nu,R}$ is the distribution function of the multivariate t-distribution with parameters ν, R.

There are further important classes of copulas that are defined in a different way. Here, Archimedean copulas are of major importance, but we do not go into the details here and refer to [MFE05, Section 5.4].

Appendix A

Toolbox

This is a list of facts and techniques from the elementary courses in mathematics which are needed during the lecture. If something is not known to you, please tell me.

- elementary calculus (differentiation and integration),
- random variables and their distributions,
- conditional probabilities,
- moments of random variables (expectation, variance...)
- distribution functions,
- densities,
- normal distribution, exponential distribution,
- basics about (maximum likelihood-)estimators and statistical tests,
- (characteristic function/Fourier transform),
- basic knowledge in matlab, R, scilab, or similar

Appendix B

On the Reading Projects

B.1 Projects

Here is a list of possible reading projects. They are divided into more mathematical topics and more practical, but this no clear distinction. You can set priorities as you like. You are also highly encouraged to find your own project.

Part of the task is to find suitable literature. Even if articles or books are mentioned, there typically exists better and more recent literature.

B.1.1 With a practical orientation

P1 More on the turkey example
[Tal10]

P2 History of financial risk
L.P. Bernstein. Against the gods. Wiley, New York (1996)
A. Steinherr. Derivatives, the wild beast of finance. Wiley, New York (1998)

P3 Long-Term Capital Management (LTCM)

P4 Finance and ethics

P5 ENRON

P6 Should banks be allowed to go into bankrupcy?

P7 Model Risk
R. Gibson (ed.) Model risk, concepts, calibration and pricing. Risk Publications London (2000)

P8 The Basel Committee and Basel III

P9 Solvency II

P10 The Swedish 1993 banking crisis

P11 The Worldcom bankruptcy

P12 The subprime crisis

P13 Jerome Kerviel, and the 5 bilion euro loss he caused Societe Generale

P14 The Madoff fraud

P15 The Lehman brothers bankruptcy

P16 The AIG near bankruptcy

P17 Were big bonuses one of the causes of the recent financial crisis?

P18 Risk connected to high frequency trading

P19 Kweku Adoboli and the $2bn UBS loss

P20 The Libor fixing scandal

P21 Insider trading

B.1.2 With a mathematical orientation

M1 More on the axiomatic theory of risk measures
H.Föllmer and A.Schied, Convex and coherent risk measures, Working paper, www.math.hu-berlin.de/ foellmer/papers/CCRM.pdf

M2 Markowitz portfolio theory
H. Markowitz. Portfolio selection. Journal of Finance , 7 , 77 – 91 (1952)

M3 Normal mixture distributions and copulas
[MFE05]

M4 Fitting copulas to data
[MFE05]

M5 Further standard methods for market risk
[MFE05]

M6 The greeks in Risk management
J.C. Hull. Options, futures, and other derivatives. Prentice - Hall, Upper Saddle River, N.J. (1989)
Y.K. Kwok. Mathematical models of financial derivatives. Springer, New York (1998)

M7 Dynamic credit risk
[MFE05]

M8 More on the model by Li

M9 On the (approximate) subadditivity of VaR.

B.2 Project report

Each group writes an approximately 5 - page summary of the project. The aim should be to teach the other course participant about the area. The summaries will be an important part of the course literature, and the final version should incorporate important comments from the discussion, if any. They should contain

(a) some of the main facts, problems and results on the subject,

(b) a carefully selected, short, list of references,

(c) a reading guide for those who want to learn more about the subject,

(d) the groups personal conclusions about what could be done to make such risks smaller in the future.

A draft version of the summary should be made available to the discussants at least 3 days before the presentation. Please, try to contact each other before that.

B.3 Oral presentation

Each group gives a 15-20 min oral presentation of their reading project, which will be followed by a 5 – 10 min discussion, so that we can see 2 presentations within 45 minutes. The dates for the presentations are listed at the course homepage.

B.4 Discussion

Each group acts as discussants for one other project. However, everyone is encouraged to participate in all the discussions. The main part of the discussion should concern substantive issues. Only brief comments, which are of general interest, about the form of the presentation or the written report should be made. Further such comments are also useful, but should be given in private.

B.5 Individual comments

Every group should hand in individual comments on 4 of the remaining reading projects (i.e, not ones own project and not the project one has been discussant on, but the other projects). This could contain new facts, questions, criticisms, ...

B.6 Format for handins

The project report should be sent to me electronically. The file name should include the group number, the names of the group members, and the title of the project, abbreviated as much as possible. The front page of the report should include the full title of the project, full names of the group members, and a brief summary of who has done what. Pdf files are much preferred to other formats. A signed paper copy of the front page should be given to me. The individual comments should be sent to me as one file (if you send comments in separate files, they will not be graded). The filename should be your name. The front page should also include your name. Again, Pdf files are much preferred to other formats and you should also give me a signed paper copy of the front page.

Appendix C

Technical Project 1

Exercise 1 (C), for grades 3,4,5,G and VG

Denote by S_n the price of a stock at day t_n, $n \in \mathbb{N}$, and by $X_n = \log\left(\frac{S_n}{S_{n-1}}\right)$ the log return of the stock. Assume that the conditional distribution of X_{n+1}, given the stock prices up to time t_n is a $N(\mu_n, \sigma_n^2)$-distribution with

$$\mu_n = \frac{1}{251}\sum_{k=n-250}^{n} X_k \quad \text{and} \quad \sigma_n^2 = \frac{1}{251-1}\sum_{k=n-250}^{n} (X_k - \mu_n)^2,$$

i.e. the conditional distribution of the log return at time t_{n+1} is normal with empirical mean and empirical variance of the log returns from the past trading year. (We ignore the days of the first trading year.)

(a) Assume that the DAX time series data on the webpage of this course follow this model. Compute for each day after the first 252 days the VaR_p, $p = 98\%$, of the DAX time series and visualize the violations, i.e. the days when the actual loss lies above the computed VaR.

(b) How many violations do you expect theoretically, how many do you observe?

(c) Compute for each trading day $n = 253, ..., 5814$, the expected shortfall $ES_{0.98}(L_n)$ of a DAX-portfolio. For the days $k \in \{253, ..., 5814\}$ where the realized loss l_k is greater than $VaR_{0.98}(L_k)$, calculate the difference between realized loss and expected shortfall:

$$l_k - ES_{0.98}(L_k).$$

What is the expected value of this quantity if the model assumption would be correct, what is the realized empirical mean?

Hint: Note that the VaR of the loss L_{n+1} at time t_{n+1} is given by

$$VaR_p(L_{n+1}) = S_n(1 - \exp(\mu_n - \sigma_n z_p)),$$

where $z_p = \Phi^{\leftarrow}(p)$ is the p-quantile of $N(0,1)$. (Do you see why?)
You may furthermore use (without a proof)

$$\mathrm{ES}_p(L_{n+1}) = S_n\left(1 - \frac{1}{1-p} e^{\mu_n + \frac{1}{2}\sigma_n^2} \Phi(-z_p - \sigma_n)\right),$$

where $z_p = \Phi^{\leftarrow}(p)$ is as above.

Exercise 2 (C), For grades 4,5 and VG
We go on with the DAX-data and the model from Excercise 1.

(a) Investigate the findings obtained in (b) further using a statistical test.
You can invent a test yourself or you can base your test on the following observation: you can consider the indicator functions

$$I_k = 1_{\{L_k > VaR_p(L_k)\}},$$

i.e. $I_k = 1$ if $L_k > VaR_p(L_k)$ and 0 else. One can prove that they are i.i.d. under our model assumptions. (You do not have to do that). So, under the model assumptions, the number of violations follows a $Bin(m, 1-p)$-distribution, where m is the number of considered trading days. Now, you can run a binomial test.

(b) Do you think the model is a good basis for risk management for the DAX? Give reasons for your opinion based on the results obtained in both exercises above.

Exercise 3 (C), For grades 3,4,5,G and VG
We go on with the DAX-data from Exercise 1 and 2, but do not use the model considered there anymore. Compute for each trading day of DAX-data the logarithmic returns $x_2, ..., x_{5294}$, which we use as risk factor changes. Compute for each trading day $m = 254, ..., 5294$ the estimates for value at risk and expected shortfall for $p = 0.98$ for the DAX, using the method of historical simulation based on the last $n = 252$ risk factor changes $x_m, x_{m+1}, ..., x_{m-n+1}$. Plot your results.

Exercise 4 (T), For grades 4,5 and VG
Prove the second equality in Lemma 2.4 (in the lecture notes).

Hint: You can substitute $y = F_L(x)$, i.e. $x = F_L^{\leftarrow}(y)$ in the integral and use that $\frac{dy}{dx} = f_L(x)$, $F_L(\infty) = 1$.

Exercise 5 (C), For grades 4,5 and VG

(a) Generate $n = 500$ simulations for

- the t-distribution with $\nu = 3$ degrees of freedom,
- the t-distribution with $\nu = 6$ degrees of freedom,
- the exponential distribution with parameter $\lambda = 1$,

and draw the corresponding Hill plot.

(b) On the webpage of this course you can find three data sets with iid simulations of three different distributions. Examine whether these data sets have a regularly varying distribution and, if appropriate, estimate the index.

Exercise 6 (T), For grades 5 and VG;
Find an estimator for ES_p based on the Hill-estimator for continuous underlying distributions analogously to the derivation of the estimator for VaR_p in the lecture notes.

Hint: Exercise 4 may be useful.

Exercise 7 (C), For grades 4,5 and VG
On the webpage of this course you find a data set with $n = 500$ iid observations of a random variable X with cdf F. Draw the corresponding mean excess plot and find a preferably small $u > 0$, such that the excess cdf F_u of X is approximately $G_{\gamma,\beta}$-distributed.

Exercise 8* (C), Bonus points!
Consider the following two (exchangeable threshold) models:
We have $m = 1,000$ (homogenous, interdependent) companies with some critical values modeled by $X^1 = (X_1^1, ..., X_m^1)$ in model 1 and $X^2 = (X_1^2, ..., X_m^2)$ in model 2. Our model is build up so that company i bankrupts if the corresponding critical value X_i is below a threshold d. The aim of this exercise is to study the effect of the choice the copula on the total number of bankruptcies.

Our model assumptions are $X^1 \sim N_m(0, R)$ and $X^2 \sim t_{m,\nu}(0, \frac{\nu-2}{\nu}R)$, where $R = (r_{ij})$ is the correlation matrix with $r_{ii} = 1$ and $r_{i,j} = 0.04$ for $i \neq j$, $m = 1,000$ is the number of companies. Furthermore, we assume that company i defaults if the corresponding critical value X_i^j falls below a threshold d^j, $j = 1, 2$. These thresholds are chosen such that the marginal default probability is 0.05. (So, in both models, the correlations and the marginal probabilities of default are the same). Furthermore, in model 2, let $\nu = 10$.

(a) Find d^j, $j = 1, 2$.

(b) Use a simulation study to compare the $VaR_p(L)$ for suitable values of p in these models, where L denotes the number of defaults. Also compare the maximal number of defaults in your simulations.

(c) Explain your findings in your own words for someone who is not familiar with risk management (but who has some knowledge in basic statistics).

Hint: You can, of course, use standard procedures in your software-packages to simulate the multivariate distributions.

Bibliography

[ADEH99] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[HL07] H. Hult and F. Lindskog. Mathematical modeling and statistical methods for risk management. https://people.kth.se/ lindskog/papers/RMlecturenotes07B.pdf, 2007.

[Kal14] J. Kallsen. Risk management. http://www.math.uni-kiel.de/numerik/kallsen/personen/kallsen/pub/skripte/risk1314.pdf, 2014.

[MFE05] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ, 2005. Concepts, techniques and tools.

[Tal10] N. N. Taleb. *The black swan: The impact of the highly improbable fragility*, volume 2. Random House, 2010.