View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Duisburg-Essen Publications Online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Universität Duisburg-Essen
Fakultät für Wirtschaftswissenschaften

Robust Risk Management in the Context of Solvency II Regulations

Dissertation

zur Erlangung des Doktorgrades Dr. rer. pol.

vorgelegt von
Dipl.-Math. oec. Robin Rühlicke
geb. am 29.7.1982 in Gifhorn

Essen, März 2013

Datum der mündlichen Prüfung: 21.5.2013

Gutachter

Prof. Dr. Rüdiger Kiesel
Prof. Dr. Antje Mahayni

Acknowledgments

First and foremost I wish to express my deepest gratitude to my first supervisor Prof. Rüdiger Kiesel. He has been an great scientific mentor and always assisted with good advice. At his chair, he created a friendly atmosphere that was very conducive to productive scientific work.

I also would like to express my thanks to Prof. Antje Mahayni for being my co-examiner.

My special thanks go to Prof. Gerhard Stahl of Talanx AG for extremely valuable input from a practical as well as scientific perspective and also for financial support of my position by his company.

Last but not least, I want to thank my parents, Heidemarie and Dieter, and my brother, Kristof, for their ongoing support and understanding.

Contents

List of Figures

List of Tables

Nomenclature

ANAV	Adjusted Net Asset Value
BSV	Black-Scholes-Vasicek model
cdf	cumulative distribution function
CEIOPS	Committee of European Insurance and Occupational Pensions Supervisors, now EIOPA
CoC	Cost of Capital
CVaR	conditional value at risk
EIOPA	European Insurance and Occupational Pensions Authority
EIOPC	European Insurance and Occupational Pensions Committee
ES	expected shortfall
HCIR	Heston-Cox-Ingersoll-Ross model
IASB	International Accounting Standards Board
IC	influence curve
iff	if and only if
IFRS	International Financial Reporting Standards
iid	independent and identically distributed
LSM	least-squares Monte-Carlo algorithm
LSMQ	least-squares Monte-Carlo with risk-neutral paths algorithm
MCEV	Market Consistent Embedded Value
MCR	Minimum Capital Requirement
NS	nested simulation algorithm
OIS	overnight indexed swap
ORSA	Own Risk and Solvency Assessment
pdf	probability density function
PH	policyholders’

PVFP	Present Value of Future Profits
QRT	Quantitative Reporting Template
RTS	Report to Supervisors
SCR	Solvency Capital Requirement
sect	section
SFCR	Solvency Financial Condition Report
SH	shareholders'
sms	separable metric space
SRP	Supervisory Review Process
TCE	tail conditional expectation
TCM	tail conditional median
thm	theorem
TVaR	tail value at risk
UGC	uniform Glivenko-Cantelli
VaR	value at risk
WLOG	without loss of generality

Chapter 1

Introduction

1.1 Motivation

Insurance is used to transfer losses from individuals to a group such that they can be shared among the group members (see Vaughan and Vaughan, 2008, p. 34). Consequently, measuring, assessing, and managing risks is of high importance to any insurance company. This is especially true in the life-insurance sector, where the premiums paid by the customers have to be invested over long time horizons to finance later payouts, thus adding market and credit risk on the asset side of the life-insurance undertaking's balance sheet. Risk management in this context is made even more complex by the nature of typical life-insurance policies that link assets and liabilities: the policies' payouts often depend on the performance of the investments made by the company, but with guaranteed minimum benefits or guaranteed yearly returns. Additionally, customers may surrender their contracts before maturity. That insurance undertakings should appreciate the value and the risks of these embedded options was already recognized by Brennan and Schwartz (1976). Today, this is rapidly becoming common industry practice (see CFO Forum, 2009, principles 6 & 7).

In earlier years, this was not necessarily the case. Risk management, as "a discipline for living with the possibility that future events may cause adverse effects" (Kloman, 1990), usually receives public attention when it fails to prevent risks from being realized that threaten the survival of a company. In the life-insurance industry infamous examples are two companies that failed to see the enormous value of guarantees embedded in the products they sold to their customers: Nissan Mutual Life in Japan and Equitable Life in the UK. Nissan Mutual Life was the first life insurance company to fail in Japan after

World War II, when it could not earn the guaranteed rate of return it promised to the holders of its pension products in 1997, leading to losses for policyholders. Equitable Life, then the world's oldest mutual life-insurance company, had to succumb to its financial difficulties in 2000, which again were the result of selling guaranteed annuity options with high guarantees. The Sharma-Report discusses other cases of (near) failures of European insurance undertakings and also identifies realized financial risks as important factors for critical situations at life-insurance companies (see Sharma, 2002). The report was an important first step towards the Solvency II regulations for the insurance industry that are to come in force in the European Union in the near future.

The Solvency II regulations imitate the structure of their counterpart in the banking world, Basel II: the regulations can be split into three pillars, covering quantitative, qualitative, and transparency requirements. The valuation of embedded options and guarantees and the quantification of risk associated with them is a main concern of the quantitative requirements that form the first pillar. It includes the prescription of the market consistent valuation of the insurance undertaking's balance sheet, therefore also including possible optionalites on the asset side. The result of this exercise has to be—at least if an insurance company opts to apply an internal model—not only the value of the balance sheet, but also a distribution of losses which leads to the regulatory capital requirements, in particular the Solvency Capital Requirement. It should not be a surprise that the necessary calculations are not simple in the majority of cases. The assets of a life-insurance undertaking are usually comprised of bonds by different issuers, sovereigns as well as corporates, shares, and alternative investments, possibly also denominated in foreign currency. Therefore, the problem is high dimensional. On the liability side, cliquet-style guarantees—the entitlements of the insured grow each year by a minimum guaranteed rate or a fraction of the return of the insurance undertaking's portfolio—and additional path-dependency introduced by surrender options and participation schemes complicate matters, often leaving simulation-based methods as the only choice.

As a principle-based framework, though, Solvency II does not specify exact methods for the calculations, but rather the properties the applied methods are required to have to gain the regulators' approval. This is done by demanding an internal model to pass several quality tests and by demanding adherence to modeling principles such as robustness. A robust model still yields stable results if model inputs such as data or the portfolio change only slightly and therefore facilitates interpretation and understanding of the results, especially by non-experts in stochastic models (see also the introduction to Stahl, Zheng, Kiesel and Rühlicke, 2012).

Non-robust risk models can have severe consequences, as highlighted by the recent financial crisis[1]. Hull (2009) identifies the "mechanistic" application of models as a source of problems, arguing that managerial judgement is always required as a complement. This is made easier by robust models. The CRO Forum (2009) is explicitly in favour of robustness in its report on the crisis and the consequences for the insurance industry, without going into details how this can be achieved.

1.2 Literature, Objectives, and Contribution

Insurance undertakings face the challenge of implementing internal models that allow for the timely computation and assessment of the financial risk of their balance sheets in order to fulfill the regulatory requirements of Solvency II, but also in order to equip themselves with instruments that allow management decisions that take risk into account in a comprehensive and consistent manner. We want to contribute in the following ways: first, by showing that robustness is an important property of methods employed for quantitative risk management, from a theoretical, practical, and regulatory point of view; second, by showing how robust methods can be used consistently in risk management; third, by discussing a variant of a simulation-based method for the computation of the loss distribution of a life-insurance company that has not shown up in the academic literature and comparing its properties—especially with regards to regulatory requirements and robustness—with more established methods.

Risk and risk assessment have been discussed by various authors in the literature (see, e.g., Kloman (1990); Klügel (2007)). In particular, Aven (2011*a*,*b*, 2010, 2007) devotes a lot of work to the topic and develops consistent definitions and frameworks, especially taking into account uncertainty about the problems at hand. However, he argues from a perspective that is suited rather for safety and reliability problems than for financial risks. In contrast, the financial risk management literature either just touches these conceptual topics, see Föllmer and Penner (2011), Cont (2006), or Cont, Deguest and Scandolo (2010), or essentially ignores them, see Hull (2011), McNeil, Frey and Embrechts (2005), and Vaughan and Vaughan (2008). In chapter 2 we contribute by interpreting the rigorous

[1]Interestingly, Kloman (1990) already predicted that "'diseconomies of risk' can overwhelm 'economies of scale' in the future." One could certainly argue that this was realized during the financial crisis.

framework by Aven in a financial context and conclude that robustness is required if we follow these ideas.

The Solvency II regulations are discussed in the literature by various authors, see Rittmann (2009) or Renz and Best (2005), or from a regulators perspective (see Financial Services Authority, 2008). While robustness and model risk show up in some works, notably in the collection by Bennemann, Oehlenberg and Stahl (2011), there is no thorough review of the regulatory body with this in mind. We provide this review in chapter 3.

Although several authors have investigated robustness of risk measures, the accounts are often quite specific, see, for example, Wozabal (2012), Föllmer (2012), Föllmer and Schied (2002*b*), and rather technical, see Krätschmer and Zähle (2011) and Krätschmer, Schied and Zähle (2012*a*). A more comprehensive approach is followed by Cont et al. (2010), who introduce a *risk measurement procedure* and discuss robust estimators. We extend this concept and introduce the *risk management procedure* which combines all the elements necessary for (robust) quantitative risk management. Our robustness approach is based on the concept of approximations using probability distances, which was introduced into the literature on statistics by Davies (1995). We recall the definitions of probability distances and probability metrics and show that, among the various examples, the Wasserstein metric has many desirable properties that make it a particularly suitable choice in a risk management context. In order to use probability distances in risk management and to relate them to risk measures, we adapt classical (see, e.g., Huber and Ronchetti (2011)) as well as recent results in statistical robustness (see Krätschmer and Zähle (2011)) for our applications to risk measures. These results allow us to to judge the robustness properties of risk measures, coming from a different direction than Cont, Deguest and He (2011).

The computation of the Solvency Capital Requirement is not a straightforward exercise in practice. Different suggestions have been made to speed up calculations: replicating portfolios (see Oechslin, Aubry, Aellig, Käppeli, Brönnimann, Tandonnet and Valois (2007)) increasing the speed of nested simulations (see Gordy and Juneja (2008), Broadie, Du and Moallemi (2010), Lan, Nelson and Staum (2007), Liu and Staum (2010)), and adapted simulation algorithms like the least-squares Monte-Carlo method (see Bauer, Bergmann and Reuss (2011*a*) and Bauer, Bergmann and Reuss (2011*b*)). We built on an idea from practice (Bergmann, Reuss, Siebert, Stahl and Zwiesler, 2009) and adapt this method in such a way, that a change of measure at the time horizon of the risk measure is not necessary. We show convergence of the algorithm and, using the asset-liability model by Zaglauer and Bauer (2008) with a Heston-Cox-Ingersoll-Ross asset model, we conduct an empirical study comparing this algorithm to alternatives. We focus on robustness and stability properties, using the previously discussed results.

1.3 Structure

Part I contains basic definitions and discusses the regulatory environment. We start by defining the general notions of "risk" and "uncertainty" and by discussing the risk management process, in particular in a financial and insurance context, in chapter 2. We see that robustness can be derived as a necessary property of risk management procedures from these definitions. In practice, however, regulatory requirements are of highest importance to insurance companies. Therefore, we discuss the upcoming Solvency II regulations for the European insurance industry in chapter 3. Again, we focus on their implications for the use of robust quantitative methods in financial risk management.

Part II is concerned with the ingredients that we need for a robust quantitative risk management process. The first element are probability distances. In chapter 4 we discuss definitions, properties, and examples, the main one being the Wasserstein metric. Probability distances are a prerequisite for obtaining many of the results in robust statistics. We streamline these results for our applications in chapter 5. In chapter 6, before applying the robustness results, we discuss axiomatic approaches for risk measures, on probability spaces as well as on data. Finally, we combine all ingredients into the risk management procedure, introduced in chapter 7.

Part III is devoted to the calculation of the Solvency Capital Requirement. We discuss several—in particular simulation-based—approaches and set up a mathematical framework for the upcoming analysis in chapter 8. Then in chapter 9, we review and extend the model on which our empirical study is based. The results of this study are presented and discussed.

Chapter 10 concludes.

Part I

Risk Management in Insurance

Chapter 2

Risk and Risk Management

2.1 Risk and Uncertainty

It is natural to begin a discussion of robustness in risk management with the definitions of the terms under consideration. Starting with the dictionary, the definition of *robust*—"capable of performing without failure under a wide range of conditions[1]"—is certainly sufficient as a starting point and is easily concretized for our use in chapters 3 and 5. In contrast, the definition of *risk*—"possibility of loss or injury[2]"—seems to be incomplete, as it does not consider the severity of a loss. In the risk management literature, the picture is even less clear. Some authors' definitions are close to the one presented, focusing on the possibility of a loss (e.g., Vaughan and Vaughan, 2008, ch. 1); some authors add the requirement that the possibility is quantifiable (e.g., McNeil et al., 2005, ch. 1), while others—those who emphasize the discussion of quantitative methods—define risk in terms of random variables (see Denuit, Dhaene, Goovaerts and Kaas, 2006, 1.4.3) or ignore the issue altogether (e.g., Pflug and Römisch, 2007).

Aven (2011*b*, ch. 8) provides a more thorough discussion of the topic, classifying definitions of risk into three groups: those that see risk "as a concept based on events, consequences and uncertainties;", those that see risk "as a modelled quantitative concept", and "risk descriptions" (Aven, 2011*b*, p. 138). He calls the latter two groups (A, C, P) definitions because they are based on events A, their consequences C and associated probabilities P. The probabilities are either frequentist probabilities—probabilities derived

[1]Merriam-Webster OnLine, s.v. "robust," accessed June 7, 2012, `http://www.merriam-webster.com/dictionary/robust`

[2]Merriam-Webster OnLine, s.v. "risk," accessed June 7, 2012, `http://www.merriam-webster.com/dictionary/risk`

using the assumption that we observe the first trials of an experiment potentially repeated an infinite number of times and interpreted as parameters of a probability model—or subjective probabilities—he also calls them *knowledge-based probabilities*—which are related to Bayesian statistics where background information and expert knowledge are used to obtain a prior distribution incorporating uncertainties about the parameters into the analysis (see Aven, 2011*a*). A representative for the (A, C, P) definitions belonging to the second of the three groups is given by Ale (2002, p. 113): "risk is defined as the combination of consequence(s) and the frequency(ies) that it arrives (they arrive)." Aven (2011*b*) classifies the definition into the third group if we replace "frequencies" with "subjective probabilities".[3] As Aven (2011*b*) states that "risk should exist as a concept without modelling and subjective probability assignments", he prefers approaches that belong to the first group. One such definition is provided by Aven and Renn (2009, abstract): "Risk refers to uncertainty about and severity of the events and consequences (or outcomes) of an activity with respect to something that humans value." Aven (2011*b*) refers to definitions of this type as (A, C, U) definitions, where U stands for uncertainty; he strongly prefers them over the (A, C, P) definitions, arguing that a definition of risk should exclude modeling assumptions. The term uncertainty used in these definitions encompasses everything that is probabilistic or otherwise unknown about the subject of interest. In particular, "uncertainties beyond the probabilities"(Aven, 2011*b*, p. 19) are included.

The usage of the term uncertainty in the literature on risk goes back to Knight (1971, p. 233), who introduced a distinction between "risk" and "uncertainty": "To preserve the distinction [. . .] between the measurable uncertainty and an unmeasurable one we may use the term 'risk' to designate the former and the term 'uncertainty' for the latter." In this context, *measurable* means that a true model with corresponding "objective" probabilities (see Knight, 1971, p. 233) is known—such as for the game of Roulette or urn experiments. Knight's (1971) definition of risk has not turned out to be useful in financial applications because the situation we encounter there is virtually never one in which a model and probabilities are known a priori (see Aven, 2011*b*, ch. 2.4). Consequently, there would be no risk and only uncertainty in this context. In contrast, the so-called *Knightian uncertainty* has become a widely used concept (see, e.g., Föllmer and Penner, 2011; Föllmer, 2008; Cont, 2006), being interpreted as model ambiguity or model uncertainty—a situation

[3] In fact, Aven (2011*b*, p. 140) misquotes Ale (2002): "Risk is the combination of probability and extend of consequences." He then places the definition into the second or third group by either interpreting the term "probability" as a frequentist or subjective probability.

where several models are possible to represent the data available.

Two related terms that also show up in the risk literature are *aleatory*[4] *variability* and *epistemic uncertainty.* Aleatory variability can be defined as "uncertainty inherent in a nondeterministic [...] phenomenon [which] is reflected by modelling the phenomenon in terms of a probabilistic model" (Klügel, 2007, p. 2198). Epistemic uncertainty, on the other hand, is "uncertainty attributable to incomplete knowledge about the phenomenon" (Klügel, 2007, p. 2199). These two notions therefore compare directly to Knightian risk and Knightian uncertainty, respectively, and are sometimes used in their place (see Aven, 2011*b*, sect. 2.4). This implies that they also carry with them the same problem in practical applications—"in the end all uncertainties are epistemic" (Klügel, 2007, p. 2199).

This discussion of risk definitions and uncertainty suggests that, even if the uncertainty concepts presented here are not directly applicable, any approach to dealing with risk should recognize uncertainties about models and their parameters. Whether a good approach to this problem is based on Bayesian probabilities and Bayesian statistics as suggested by Aven (2011*b*) or on different methods certainly depends on the problem at hand, as we will see shortly.

2.2 Risk Management

Having defined what we mean by risk, we have to think about what to do about it. The answer is risk management, which is "a discipline for living with the possibility that future events may cause adverse effects" (Kloman, 1990) or, according to Aven (2011*a*), "all coordinated activities to direct and control an organisation with regard to risk." The term *risk management* first appeared in the literature, if we believe Kloman (1990), in the article by Gallagher (1956). While the research on risk management has, of course, progressed significantly since the publication of his article, there are some noteworthy points. Gallagher (1956) sees changes in the legal environment concerning workplace safety as the main driver for the management of companies to consider a systematic treatment of risks. Today, legal reasons are still an important motivation for changes in the risk management of companies. Examples are the Basel regulations in the banking industry (see, e.g., Basel Committee on Banking Supervision, 2011) and the Solvency II project

[4]From the Latin "alea", fig. dice, see The Oxford Latin Dictionary s.v. "alea"

concerning insurance companies, which we discuss in detail in chapter 3. Additionally, the three main elements of risk management identified in this article are risk analysis, risk abatement, and risk coverage. These are still main parts of the risk management process and are recognized as such in the literature. Kloman (1990) calls the phases risk assessment, risk control, and risk financing, Vaughan and Vaughan (2008) call them risk identification and evaluation, risk control, and risk financing. Table 2.1 shows a summary. These elements enter into the risk management process which includes additional steps.

Element	Description
risk analysis, risk assessment, risk identification & evaluation	What are the risks? What is their (subjective) probability and severity?
risk abatement, risk control	How can a risk be avoided, its probability or severity reduced?
risk coverage, risk financing	How are the costs of risks financed? Are risks retained or transferred? If so, how?

Table 2.1: Elements of Risk Management, (see Gallagher, 1956; Kloman, 1990; Vaughan and Vaughan, 2008)

In a dynamic environment such as a company, risk management cannot be a one-time exercise. The measures taken and the detailed steps of the process have to be reviewed and possibly revised. This is the reason why the risk management process is often shown as a risk management cycle, such as in figure 2.1.

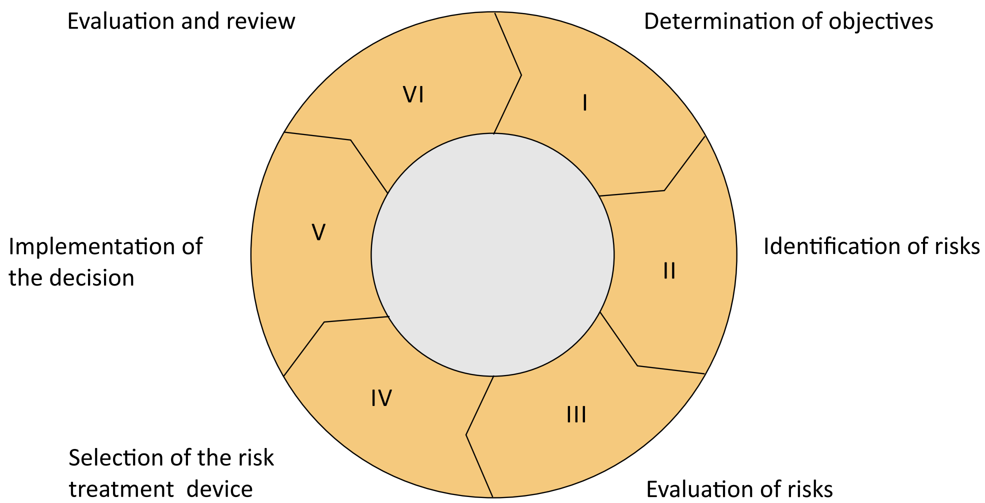

Figure 2.1: The risk management cycle, steps according to Vaughan and Vaughan (2008)

The first step is the determination of objectives. Doherty (1985, ch. 2) sees value maximization from the point of view the equity holders as the main objective of risk management (This is in line with general studies of corporate objectives, at least under some assumptions, see Hansen and Lott (1996) for a review of the literature). Although in theory, market mechanisms should lead to an incorporation of other stakeholders' objectives into the value of equity, this is not always the case, due to market imperfections and externalities. An example could be the events of the recent financial crisis, where the put options that owners of a limited liability company implicitly buy from society became very valuable for some banks. Doherty (1985, p. 23) puts the responsibility for such risks to the state, which has to introduce regulation. Due to the presence of bankruptcy costs—empirical studies show that they can be as high as 20% of firm value (see Bris, Welch and Ning, 2006)—decreasing the probability of default is often a consequence of value maximization. Vaughan and Vaughan (2008) start from this perspective. They see as the primary objective of risk management that it should allow the company to reach its goals by making sure that the company does not go out of business because of losses due to realized risks. Such a risk management philosophy calls for measures that, for example, reduce the likelihood of major accidents that lead to indemnifications the company would be unable to pay, and also protects debt holders. It is also in line with the goals of the regulations of the insurance industry in Europe, which are in our focus: the Solvency II Directive names "protection of policy holders and beneficiaries" (The European Parliament and the Council of the European Union, 2009, Preambel, (16)) as the main objective of the regulation and this implies that insurance companies should be solvent. The selection of the primary objective has to be augmented by a discussion of the details. These could include the specification of a survival probability (as in Solvency II) if guaranteed protection is not feasible. Once agreed upon, the objectives should be summarized in a risk management policy (see Vaughan and Vaughan, 2008, p. 26).

Risk identification, the second step in the process, can be achieved using a number of different means. A starting point can be any documentation of the company's structure and implemented processes, but also items like balance sheets, lists of physical assets and (future) products can be a valuable sources of information. Vaughan and Vaughan (2008, pp. 27f) list tools that can be used to extract information from these sources.

Risk assessment is the next step in the risk management cycle. The risks that have been identified have to be described in terms of their severity and their probability. Here, the different definitions of risk discussed in section 2.1 imply different methods of completing this task. Starting with one of Aven's (2011*b*) definitions of risk, either of (A, C, P) or (A, C, U)-type, we see that this is not sufficient to progress further. An (A, C, P) definition has to be extended to incorporate model ambiguity, and for tasks like risk description, risk assessment, or risk quantification—discussed in the following section—a probabilistic approach is lacking for (A, C, U) definitions. This view is shared by Aven (2011*b*), who introduces a risk description as

$$\text{Risk description}_{A,C,P} = (A, C, P_f, U(P_f), U, K)$$

in the former case and

$$\text{Risk description}_{A,C,U} = (A, C, U, P_s, K)$$

in the latter case, adding background knowledge K as well as uncertainties or probabilities, respectively, to the risk definitions. In the first case, P_f are frequentist probabilities, $U(P_f)$ refers to the uncertainty about them, and U stands for further uncertainty factors. In the second case, P_s relates to subjective, Bayesian probabilities. In order to obtain a risk description, Aven (2011*b*, sect. 8.2) adapts a framework for quantitative risk assessment from de Rocquigny, Devictor and Tarantola (2008), as shown in figure 2.2. The

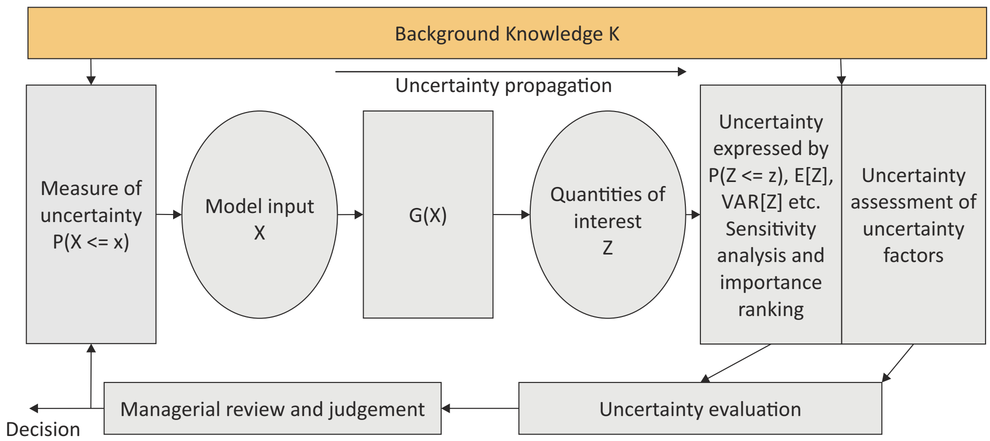

Figure 2.2: A framework for risk assessment (Aven, 2011*b*, Fig. 8.3)

uncertainty in this framework is in the input variables X and the quantities of interest Z. This uncertainty is measured using probabilities in this framework. Nevertheless, Aven (2011*b*, sect. 8.2.1) requires both X and Z to have values that can be objectively determined—today or in the future—in order to use them as a basis for scientific risk assessment. The uncertainty about X is then propagated by the model G—sometimes by means of simulations—and the result is a (subjective) probability distribution of Z. This distribution can be evaluated using probabilistic measures such as expectation, variance, or quantiles, but this analysis should also be augmented by qualitative assessment of the background knowledge K and the model assumptions implicit in G. The information gathered in this step enters into the decision-making step, which comes next in the risk management cycle.

The decision-maker—guided by the risk management policy, if available—has the choice between different responses to the risks that have been identified and evaluated. The first decision is whether to do something about the risk in question or whether to retain it. The answer depends on the severity of the potential loss, its probability, the resources available to finance the loss, and the costs of risk mitigation measures (see Vaughan and Vaughan, 2008, p. 30). A common device to visualize the first two dimensions is a risk heat map such as shown in figure 2.3. Risks with low probability and low severity are

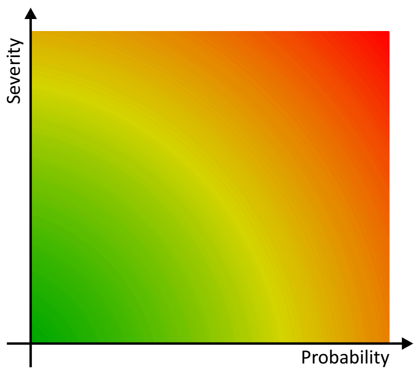

Figure 2.3: A risk heat map (see, for example, Weis, 2009, Abb. 5)

located in green areas of the heat map, indicating that usually no action is necessary, and risks with high risk and high severity are located in red areas, indicating the need for risk mitigation. For risks that are located in between the two areas, careful consideration of costs and benefits of possible actions is necessary (see Weis, 2009, pp. 62f). Depending on the area of application and the type of the risk, various measures can be taken. figure 2.4 provides a summary.

The main risk control techniques are risk avoidance and risk reduction. Risk avoidance

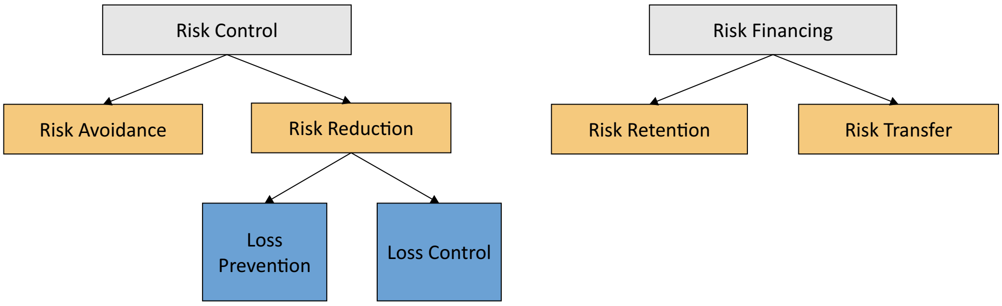

Figure 2.4: Risk treatment techniques, compiled from Vaughan and Vaughan (2008, sect. I.2) and Weis (2009, pp. 66f)

is often applied to risk in the red area of the heat map. Activities that lead to such risks are stopped or not taken up. As this might mean that business opportunities are not realized, Vaughan and Vaughan (2008, p. 18) call risk avoidance "the last resort in dealing with risk." Risk reduction techniques can again be separated into two groups, loss prevention and loss control. Loss prevention aims at lowering the probability of the risk. We can think, for example, of improved maintenance procedures or the installation of additional safety features. Loss control procedures aim to reduce the severity of a loss. Examples in a financial context are some hedging strategies or limits systems for the traders of a bank.

Having applied risk control techniques, risk financing strategies describe how losses are funded once they have occurred. One possibility is to basically do nothing, that is, retain the risk. This is the default handling for risk with low probability and low severity (see Weis, 2009, p. 67)—the green area on the risk heat map in figure 2.3—as well as for risk that remains after risk avoidance, reduction and transfer have been applied. In these cases, risk retention is voluntary and intentional. In contrast, risk retention is involuntary if the other strategies are not available and unintentional if risk is retained that has not been identified in the first place (see Vaughan and Vaughan, 2008, p. 19). The classical example of a risk transfer is the purchase of insurance, another the producer's forward sale of her products on commodity markets. There is a variety of possible contracts that transfer risk. The common characteristic is that the nature of the risk remains unchanged, only the party exposed to the risk changes.

The combination of measures that is selected in the previous step has to be implemented. The results are evaluated and suggestions for improvements enter again into the risk management cycle.

2.3 Financial Risk Management

We have discussed risk management in a general context and are now ready to focus on financial risk management, especially in the (life-) insurance industry. As we once again go through the risk management cycle, we concretize the ideas of the preceding section. We center our attention on the risk evaluation step, which is the point in the cycle where quantitative methods are most important.

The risk management objectives do not change in the financial context—equity value maximization or reduction of default probability are certainly the main objectives. McNeil et al. (2005) name other areas where risk management can be useful for companies. We could call these secondary objectives. They name decreasing the cash flow variability, which often leads to lower taxes, and cost benefits from a reduced need of external financing (see Hennessy and Whited (2007) for an empirical study on the size of the costs of external financing).

The task of risk identification is made easier if we know what types of risk we are likely to encounter and how to name them. The types of risk most frequently found in the financial industry are, according to McNeil et al. (2005), *market risk*, *credit risk*, *operational risk*, *model risk*, and *liquidity risk*. For insurance companies, in particular, *underwriting risk* is also important. Table 2.2 summarizes the definitions. An especially important problem in the risk identification of financial risk is to find options and guarantees embedded in contracts into which the company has entered. For example, life-insurance undertakings include guarantees and surrender options in their products (see, e.g., Grosen and Jørgensen, 2000).

Next in the risk management cycle is the risk evaluation step. As suggested by Aven's (2011*b*) framework (see figure 2.2), quantitative methods are used to deal with the tasks. We consider the financial risk associated with a portfolio. The portfolio under consideration can be the position of a single trader at a bank or—in the case of an insurance undertaking—the whole balance sheet (see section 3.2.1). We start with the model input X which consists of a number of risk factors driving the portfolio value. Common risk factors are stock prices, commodity prices, interest rates, and economic factors such as inflation. For these risk factors, we have to find a measure of uncertainty $P(X \leq x)$. In practice, this is much more difficult in our financial setting than in the example presented by Aven (2011*b*, sect. 8.2.2) for two reasons. First, the risk factors can take not only a

Risk Type	Definition
Market Risk	risk of loss or of adverse change in the financial situation resulting, directly or indirectly, from fluctuations in the level and in the volatility of market prices of assets, liabilities and financial instruments
Credit Risk	risk of loss or of adverse change in the financial situation, resulting from fluctuations in the credit standing of issuers of securities, counterparties and any debtors ... in the form of counterparty default risk, or spread risk, or market risk concentrations
Operational Risk	risk of loss arising from inadequate or failed internal processes, personnel or systems, or from external events
Liquidity Risk	risk that ... undertakings are unable to realise investments and other assets in order to settle their financial obligations when they fall due
Model Risk	[risk of loss from] using a misspecified (inappropriate) model for measuring risk
Underwriting Risk	risk of loss or of adverse change in the value of insurance liabilities, due to inadequate pricing and provisioning assumptions

Table 2.2: Types of risk, definitions from the Solvency II directive (The European Parliament and the Council of the European Union, 2009, Article 13)(all except for model risk) and (McNeil et al., 2005, p. 3)(model risk)

finite number of values, but any (positive) real number. Second, the quantity of interest Z, which is the portfolio value or portfolio loss at some time T in the future, often depends not only on the value of the risk factors at time T, but also on their history from today up to T. If this is the case, the risk factors have to be modeled as stochastic processes (X_t). The probability model for the risk factors is selected using the background knowledge K, in particular historical data, theoretical considerations, and expert opinion (Pflug and Römisch, 2007, p. 2). While this implies that the probabilities thus computed are subjective probabilities and in line with Aven's (2011*b*) interpretations, his Bayesian ansatz for the quantification of uncertainty is not promising in this case. Even for relatively simple problems, experts are notoriously bad at estimating probabilities (see, e.g., Kahneman, Slovic and Tversky (1982)). Additionally, the problem at hand calls at best for the estimation of prior distributions on the parameters, which might not be independent, or even for the estimation of prior probabilities for different classes of stochastic processes. As model uncertainty is an important part of risk and risk management, we do not ignore the issue and discuss our approach in section 2.4. The model M which translates the risk factors into the distribution of portfolio value or loss Z can take different forms. Examples are linear combinations of the risk factors, Delta-Gamma approximations of option values

(Hull, 2012, 21.5 & 21.6), and the model for participating life-insurance in chapter 8. Risk measures (see chapter 6) play an important part in the uncertainty assessment in financial risk management. The qualitative aspects are part of the Solvency II regulations, see section 3.2.1.

The general risk treatment devices discussed above can also be used in financial risk management. Particular to financial risk are hedging strategies and the use of derivatives. Implementation, evaluation, and review receive attention in the Solvency II regulations, see section 3.2.

2.4 Robustness and Uncertainty in Financial Risk Management

The need to deal with model uncertainty emerged from our discussion of risk definitions and descriptions, but it is also a problem we encounter in practice. Stahl et al. (2012, ex. 1) present an example from life-insurance that illustrates the potential impact of model uncertainty on risk assessment. The example also suggests that robust procedures can help with the issues that arise if we are not—and cannot—be sure about our probability model and its parameters[5].

The approach by Stahl et al. (2012), which we also use in this thesis, is based on the idea that all models are—we cannot know whether a true model exists and what this model might look like—approximations. The concept of approximations was introduced into the statistics literature by Davies (1995). The term *approximation* immediately provokes the questions “How good is the approximation? How ’far’ are we away?” To answer these questions, we need a distance or metric on the space of the objects of interest, in this case probability distributions (see chapter 4). The concept has the advantage that it can also be applied to other approximations that might be necessary in the quantitative risk assessment process. This could be a dimension reduction if the number of risk factors driving the portfolio value is too high to be tractable, as it is often the case for life-insurance companies, or a Monte-Carlo simulation that is necessary if closed-form solutions do not exists for the quantities of interest.

Having acknowledged that we use approximations throughout the risk assessment process, we have to adapt the risk measurement procedures accordingly: distributions of the

[5]Here and in the following, the term *model* refers to the probability model for the risk factors X in the risk assessment framework and, if necessary, the propagation model M.

portfolio value that are "close" (in terms of the chosen distance) should be assigned "similar" (in terms of some distance) risk values. This intuition is closely related to continuity and robustness concepts in the Statistics literature (see chapter 5). Therefore, the results from robust statistics can provide a guideline for the selection of risk measures that fit our framework among the various classes of such functionals (see chapter 6).

Chapter 3

Regulatory Framework: Solvency II

Risk management in an insurance company is strongly influenced by regulatory requirements. This is especially true with the upcoming Solvency II regulations in the European Union[1], which follow a risk-based approach (The European Parliament and the Council of the European Union, 2009, Preamble (15)). In this chapter we show why and how the regulatory framework is set up and give an overview of the regulations. We will see that quantitative risk management and robustness play an important role in the Solvency II framework.

3.1 The Road to Solvency II

An insurance company takes on and pools risks from its customers, sharing the risk among them (see Vaughan and Vaughan, 2008, p. 34). It is therefore clear that the company itself faces risks which stem directly from this business, the so-called underwriting risk (see CEA Groupe Consultatif, 2007, p. 55). Historically, the regulation of insurance companies was focused on underwriting risk. In Germany prior to 1994, for example, the supervisory authorities had far reaching control over the business of insurance undertakings, including its products (see Knauth, 2005, sect. 2). In life insurance, biometric actuarial assumptions and cost calculations were prescribed by the regulator, in essence leading to fixed premia across the life insurance market (see Faulhaber, 2004, sect. 2).

[1] In the discussion of institutions, legal texts, and other matter related to the European Union, we follow the style guide of the European Commission (European Commission Directorate-General for Translation, 2012)

The insurance regulation in the European Communities and now the European Union shares the focus on underwriting risk. The First Council Directive 73/239/EEC concerning non-life insurance (The Council of the European Communities, 1973) and the First Council Directive 79/267/EEC concerning life insurance (The Council of the European Communities, 1979) first aimed at coordinating the insurance regulations in the Member States. The Directives broadly state when insurance companies may take up and continue business and under which conditions the authorization may be withdrawn. The capital requirements for insurance companies, a solvency margin and a minimum guarantee fund, are derived from earned premiums, claims, and mathematical reserves, all of which are related to underwriting. Since the 1970s, there have been changes of and additions to these Directives, most recently by the Directives 2002/13/EC and 2002/83/EC (The European Parliament and the Council of the European Union, 2002*a*,*b*). The resulting body of regulation is known as "Solvency I" (see Wagner (n.d.)).

While drafting the Solvency I Directives, the Insurance Committee agreed that "a more fundamental and wider-ranging review of the overall financial position of an insurance undertaking, including investment risk, (Solvency II) should be commenced" (The European Commission, 1999, sect. 1.1). Further evidence for the necessity to rework the regulation was provided by the Sharma-Report (Sharma, 2002), which analyzes failures and "near-misses" of European insurance companies and conducts detailed case studies. The authors conclude that a regulatory regime should encompass risk-based capital and solvency levels, and a broad set of early-warning tools for the regulators that cover different types of risks. Also, it should include supervision of qualitative aspects such as management and risk management processes. They recommend the Lamfalussy process for the legislative procedure which was developed by the Committee of Wise Men on the Regulation of Security Markets (Lamfalussy, Herkströter, Rojo, Ryden, Spaventa, Walter and Wicks, 2001). Council, Commission and European Parliament followed this advice.

There are four levels under the Lamfalussy approach (The European Commission, 2007, 2004*b*). At the first level, the Commission combines, after a full consultation process, framework principles into a proposal to the Council and the European Parliament, which adopt the suggestions in a Directive or Regulation. For the Solvency II project, Council and European Parliament adopted Directive 2009/138/EC. The Commission has recently proposed an amendment for the Directive called "Omnibus II" (The European Commission, 2011). At the second level, the implementing measures are designed using

Level	Main Actors	Results
I	Council and European Parliament	Directive 2009/138/EC & Omnibus II Directive
II	Commission, consulting EIOPC, European Parliament, EIOPA	Implementing measures
III	EIOPA	guidance for national supervisors
IV	Commission	enforced Community Law

Table 3.1: The Lamfalussy process in context of the Solvency II project

the "comitology" procedure laid down in The Council of the European Communities (1999, 2006). In the case of the Solvency II project, the Commission asked the Level 3 Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS) (The European Commission, 2004*a*)[2] for technical advice. At the time of writing, CEIOPS has issued final versions of the Level 2 Advice[3] and the Commission has conducted a public consultation. The next step in the Lamfalussy process is the consultation of the Level 2 "comitology" committee (the European Insurance and Occupational Pensions Committee (EIOPC), (see The European Parliament and the Council of the European Union, 2005)) and the European Parliament. If both agree, the Commission will adopt the Level 2 implementing measures. In part, this advice is based on five "Quantitative Impact Studies" (QIS), which gauge the impact of the new rules (see The European Commission (2010) for further details on the fifth quantitative impact study (QIS5)). At Level 3, the responsible committee, EIOPA, facilitates the cooperation between national supervisors, aiming at consistent implementation of the Solvency II framework in the Member States. The Commission is the main actor at Level 4. It enforces Community Law in the Member States.

3.2 Structure

The Solvency II regulations can be grouped into three pillars, analogously to the Basel II framework in banking regulation. Figure 3.1 gives an overview. We examine the three

[2]CEIOPS was given additional authority and was renamed "European Insurance and Occupational Pensions Authority" (EIOPA) in 2010 (The European Parliament and the Council of the European Union, 2011). We cite documents published by CEIOPS under the original name, but refer to current and future work using EIOPA.

[3]Available at `https://eiopa.europa.eu/publications/sii-final-l2-advice/index.html`(accessed 18/10/2011)

pillars in the following sections. The methods for measuring risk which are at center of this thesis have to be used extensively in Pillar I, which is why we put a strong emphasis on this part in our discussion of the Solvency II framework.

In this section we refer repeatedly to articles in the Solvency II Directive (The European Parliament and the Council of the European Union, 2009). To make the text more readable, we use only the name and the article number, omitting the reference to the Directive.

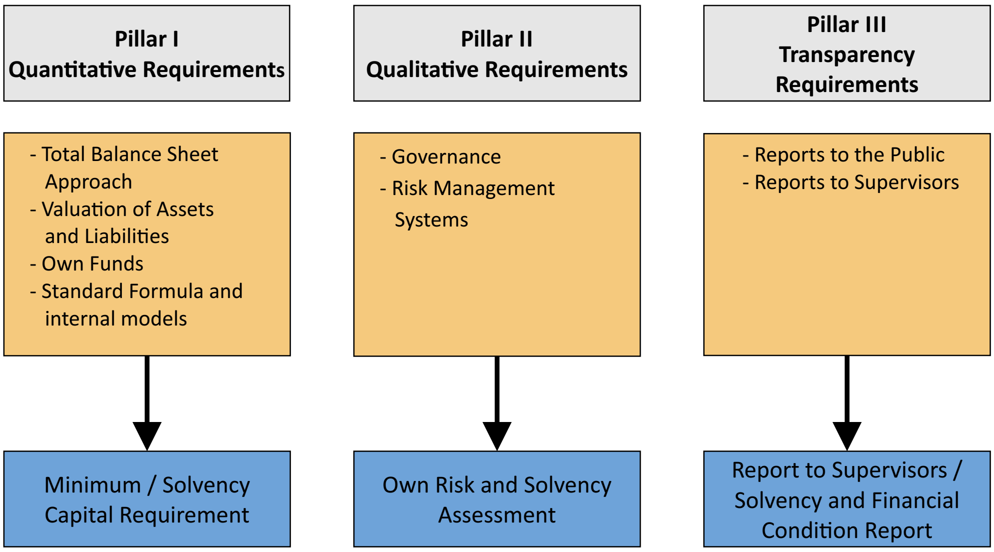

Figure 3.1: The three pillars of Solvency II

3.2.1 Pillar I: Quantitative Requirements

Pillar I covers the calculation of capital requirements in the form of the *Solvency Capital Requirement (SCR)* and the *Minimum Capital Requirement (MCR)* and the necessary steps leading up to them. The corresponding rules can be found in chapter VI of the Solvency II Directive.

Total Balance Sheet Approach and Capital Requirements

The calculation of own funds and the capital requirements are based on an economic balance sheet, meaning that it is obtained using market values of assets and liabilities where available. Figure 3.2 shows the main elements.

Under Solvency II, assets and liabilities should be valued such that they could be exchanged, or transferred or settled, respectively, "between knowledgeable willing parties in an arm's length transaction." (The European Parliament and the Council of the European Union, 2009, Article 75). EIOPA in the Level 2 implementing Advice (see CEIOPS (2009*n*)) and The European Commission (2010) recommend to follow the standards set by the International Accounting Standards Board (IASB) in their International Financial Reporting Standards (IFRS) for the economic valuation of assets and liabilities other than technical provisions. The main exception from this is that the value is not adjusted for the credit quality of the insurance undertaking itself (see Article 75 (1)). The European Commission (2010, V.1.4) provides detailed explanations of other deviations from IFRS. If possible the valuation should be based on a mark-to-market approach. Otherwise mark-to-model valuation is permitted. In this case the undertaking has to take into account all relevant market data and take additional measures to convince management and supervisors that the valuation is appropriate and reliable (see CEIOPS, 2009*n*, 3.36-3.38).

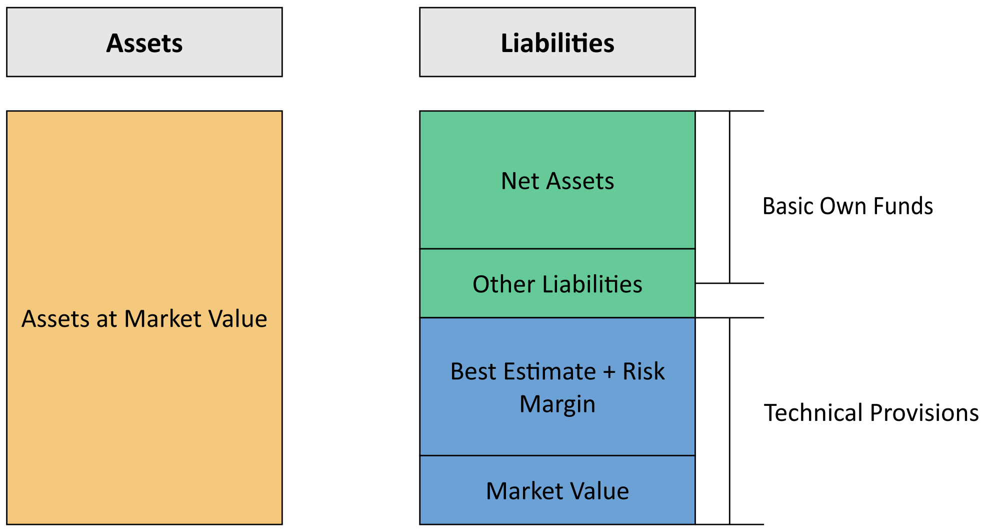

Figure 3.2: The Solvency II Balance Sheet

Concerning the valuation of technical provisions, the mark-to-market approach is a special case. If insurance obligations can be replicated reliably with financial instruments that are traded in an active market, their value can be derived directly from the value of those instruments (Article 77 (4)). The European Commission (2010, V.2.4) envisages this possibility mostly for insurance products that promise the delivery of a portfolio of such assets. When the cash-flow depends on other factors such as "the level, trend, or volatility of mortality, disability, sickness and morbidity rates" (The European Commission, 2010, TP.4.8) or is related to expenses, it cannot be reliably replicated (see The European Commission, 2010, TP.4.8). In this case, the value consists of a best estimate and a risk margin (see Article 77). The computation of the best estimate has to be segmented at least by line of business. The European Commission (2010, V.2.1) suggests how this can be done. The best estimate is a probability weighted average of discounted future cash flows and therefore dependent on the prediction of future cash flows and the discount rates.

The projections of future cash-flows should include cash inflows (such as future premiums and receivables but not investment returns) and outflows (such as benefits to policyholders, expenses, and tax payments) over the full lifetime of the insurance contracts (see The European Commission, 2010, V.2.2.1). Also included are payments resulting from bonuses to policy holders and expectations about inflation (see Article 78). Cash flows from reinsurance contracts or special-purpose vehicles are not included and are valued separately according to Article 81, taking credit risk into account.

The discount rates should be based on a risk free term structure of interest rates, either derived from government bonds or, if this is not possible for some currency, from other financial instruments such as swaps (see CEIOPS, 2009*l*). In the fifth quantitative impact study (see The European Commission, 2010), different curves were provided, some with an illiquidity adjustment.

Mostly following the final Level 2 Advice (CEIOPS, 2009*m*), the risk margin is calculated in QIS5 (see The European Commission, 2010, V.2.5) based on a scenario where the insurance obligations, including corresponding reinsurance contracts, of the undertaking are transferred to a new, empty, entity which is then equipped with sufficient capital to cover the Solvency Capital Requirement (SCR) (see CEIOPS (2009*m*, 3.1.3.1) and the discussion of the SCR later in this section) of the liabilities and corresponding as sets that minimize this SCR. This SCR includes (see CEIOPS, 2009*m*, 3.49) underwriting risk, counterparty credit risk from reinsurance, operational risk, and unavoidable market risk. It can be calculated with the same method as the underlying's SCR and has to be segmented at least by line of business. The Cost-of-Capital margin $CoCM$ is then calculated as (see CEIOPS, 2009m, 3.125)

$$CoCM = CoC \sum_{t \geq 0} \sum_{\text{line of business } i} \frac{SCR(t,i)}{(1+r_t)^t}$$

where r_t is the risk free rate for t, $SCR(t,i)$ is the Solvency Capital Requirement at t for line of business i, and CoC is the Cost-of-Capital rate. The Cost-of-Capital rate used in QIS5 was 6%, and this equals the lower bound set by CEIOPS (2009m, 3.1.3.2.2).

A balance sheet valuation as described in the previous paragraphs allows the insurance undertaking to calculate its own funds. Figure 3.3 provides an overview of own funds in Solvency II. They consist of so called *basic own funds* and *ancillary own funds* (see Article 87). Basic own funds are calculated as the difference between the value of assets and liabilities, valued as described above, plus subordinated liabilities (see Article 88). They are therefore on balance sheet items. In contrast, ancillary own funds, such as unpaid share capital, letters of credit and guarantees, are off balance sheet items and their use has to be approved by supervisors (see Articles 88 and 89).

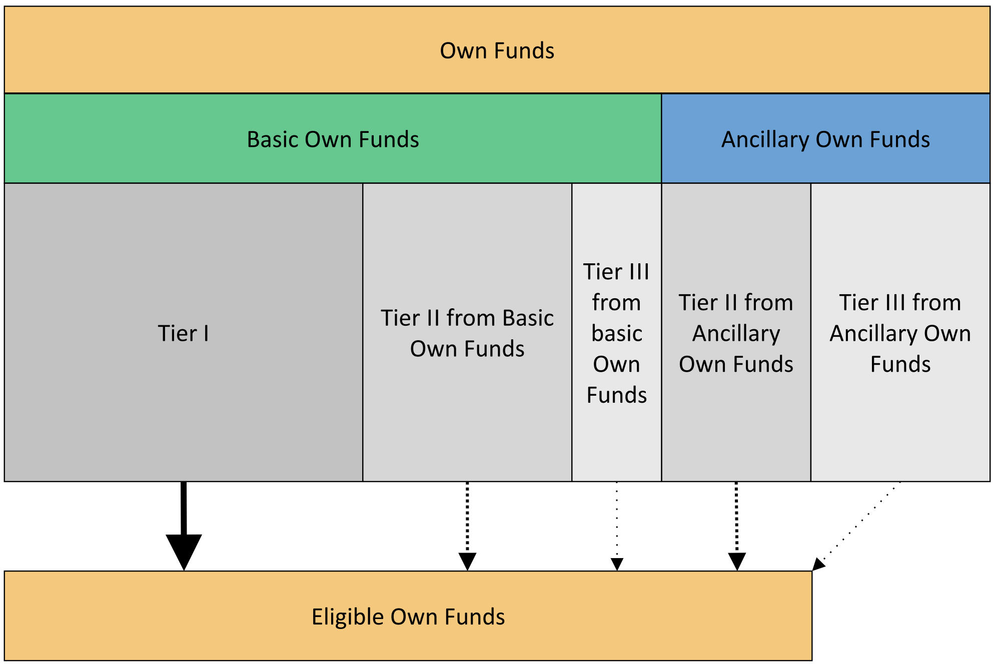

Figure 3.3: Own funds in Solvency II

Basic own funds and ancillary own funds are further classified into tiers according to their characteristics regarding loss absorption (see Articles 93-97). Only basic own fund items can be in the best class, Tier 1. Both basic own fund items and ancillary own

	Tier 1	Tier 2	Tier 3
Subordination	deeply subordinated	effectively subordinated	
Loss absorbency	fully paid in, first to take loss	can be called up, must absorb losses	
Sufficient duration	$\geq$ 10 years	$\geq$ 5 years	$\geq$ 3 years
Free from requirements or incentives to redeem	no incentives, redeemable only at option of undertaking	moderate incentives to redeem, redeemable only at option of undertaking	no coupon payments/redemption after breach of SCR
Free from mandatory fixed charges	coupons/dividends cancelable, full discretion over amount	coupons/dividends deferred after breach of SCR	
Absence of encumbrances	yes	yes	yes

Table 3.2: Requirements for classes of basic own funds in Solvency II, Source: CEIOPS (2009*c*, 3.208, 3.227, 3.231)

fund items can be in Tier 2 and Tier 3. Table 3.2 summarizes the classification criteria from the Level 2 final Advice issued by CEIOPS (2009*c*). The criteria for ancillary own funds are somewhat simpler: in general, ancillary own fund items “which, if called up and paid in, would be classified in Tier 1” CEIOPS (2009*c*, 3.235), are classified in Tier 2; ancillary own fund items that would be classified in Tiers 2 or 3, are classified in Tier 3. The classification of own funds is relevant when coverage of the capital requirements is concerned.

The two levels of capital requirements in Solvency II are the Solvency Capital Requirement and the Minimum Capital Requirement. The SCR has to be calculated at least once a year (see Article 102) on a “going concern” basis as the 99.5% value at risk (VaR, see chapter 6 for an in-depth discussion) of the basic own funds of the insurance undertaking (see Article 101). The SCR can be calculated in one of three ways, using the standard formula (see Articles 103-111), a full, or a partial internal model (see Articles 112-127). We summarize the main features of the standard formula and internal models in the following sections. Here, a main point of interest is that the SCR has to be covered by eligible own funds. Both basic own funds and ancillary own funds are eligible to cover the SCR, but there are restrictions concerning the tiers of the own funds. CEIOPS (2009*c*, 3.195) opines that at least 50% of the eligible funds should consist of Tier 1 own funds and that no more than 15% should consist of Tier 3 own funds. Article 138 details the consequences of non-compliance with the SCR. An undertaking that breaches the SCR immediately has to inform the supervisor about the event and develop a recovery plan—detailed in Article 142—within 2 months. Within 6 months the undertaking has to make sure that it complies with the SCR. The supervisor can extend this deadline or take additional measures such as restrictions of the use of assets in exceptional circumstances.

The MCR has to be calculated quarterly as a linear function of technical provisions, written premiums, capital-at-risk, deferred tax and administrative expenses (see Article 129(3)); it has to be calibrated such that it matches the one year, 85% value at risk of the basic own funds (see Article 129(1)c). Additionally, it should lie in a corridor between 25% and 45% of the undertaking's SCR (see Article 129(3)), but subject to a fixed floor (see Article 129(1)d), which takes precedence over the corridor requirement. Taking this together, CEIOPS (2009*b*) suggests the formula

$$\mathrm{MCR} = \max\left\{\min\left[\max\left(\mathrm{MCR}_l; 0.25\cdot\mathrm{SCR}\right); 0.45\cdot\mathrm{SCR}\right]; \mathrm{Floor}_B\right\}$$

where MCR_l is the linear MCR-function referred to in Article 129(3) and Floor_B is the MCR-floor, depending on the type of business of the undertaking. Only basic own funds of Tier 1 and Tier 2 are eligible to cover the MCR, and the proportion of Tier 1 own funds should be at least 80%. As losses to policyholders are more likely upon a breach of MCR compared to just a breach of the SCR, the consequences of non-compliance with the MCR, laid down in Article 139, are potentially more severe. Analogously to the SCR case, an undertaking that breaches the MCR immediately has to inform the supervisor about the event and develop a financial scheme, detailed in Article 142, that shows how MCR-compliance can be restored. But in contrast to the SCR-recovery plan, it has to be presented within one month, and compliance with the MCR has to be restored within three months. The free disposal of assets can be restricted. If the supervisory authority finds the financial scheme "manifestly inadequate" (Article 144(1)) or if the undertaking does not comply with it within three months, the supervisory authority may withdraw the authorization to do business (see Article 144(1)).

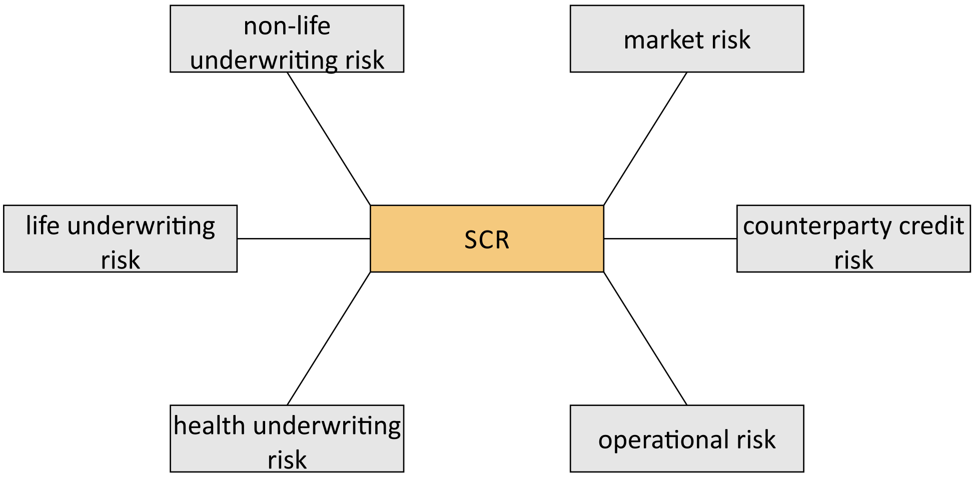

Figure 3.4: Risks that have to be included in the SCR according to Article 101

Standard Formula

The calculation of the SCR using the standard formula is detailed in Articles 103-112. It is the sum of the Basic Solvency Capital Requirement, a capital requirement for operational risk, and an adjustment for the loss-absorbing capacity of technical provisions and deferred taxes (see Article 103). According to Article 104, the calculation of the Basic Solvency Capital Requirement is further split up into risk modules for non-life underwriting risk, life underwriting risk, health underwriting risk, market risk, and counterparty credit risk. In this way, all types of risks shown in figure 3.4 are covered by the Standard formula. The result of each of the risk modules is a one-year value at risk at a 99.5% confidence level. Explicit formulas for the risk modules and their sub-modules derived from the requirements in Article 105 and correlations for the aggregation are provided in the final Level 2 Advice, see CEIOPS (2009*h*,*e*,*f*,*i*,*d*, 2010). Studies have shown that the standard formula has deficiencies in some cases (see, e.g., Pfeiffer and Strassburger (2008) and Sandström (2007)).

Internal Models

The Solvency II standard formula is, by definition, not adapted to the specifics of a single undertaking. Already at the beginning of the process leading up to the Solvency II regulations, the European Commission wanted to encourage the use of internal models, which better capture the risk profile of an undertaking, by opening up the possibility of reduced regulatory capital requirements (see The European Commission, 2002*a*, 207). An internal model, which produces a probability distribution forecast of relevant quantities, can also aid with identification, understanding, and measurement of risks, which facilitates calculation and allocation of economic capital (see International Association of Insurance Supervisors, 2007; The European Commission, 2002*b*). The following discussion of the requirements for internal models in Solvency II will suggest that the cost of implementation and maintenance of an internal model, both in terms of financial and human resources, can be high. Therefore, undertakings have to take positive and negative aspects into account when making a decision about the development of such a model (see International Association of Insurance Supervisors, 2007, 13). This is especially true in the Solvency II context because a reversion to the standard formula requires authorization of the supervisors (see Article 117).

The use of a full or partial internal model in Solvency II is governed by Articles 113-127. The calculation is divided just as for the standard formula. Insurance undertakings can use internal models for one or more of the three main parts of the SCR (Basic Solvency Capital Requirement, operational risk, adjustment), and for one or more of the risk (sub-) modules of the Basic Solvency Capital Requirement, given supervisory approval (see Article 113). In any case, supervisory approval is necessary for the introduction and major changes of an internal model. An insurance undertaking has to apply for authorization, providing documentation that the model fulfills the requirements of Articles 120-125. The final Level 2 Advice by CEIOPS (2009*g*) on the application process suggests the introduction of a pre-application phase in which the supervisory authorities understand basic facts of the model and communicate possible difficulties to the undertaking. The application process itself, which takes a maximum of six months, starts with the submission of the complete documentation of the internal model to the regulator. Being based on this documentation, it can also include on-site inspections and requests for additional information (see CEIOPS, 2009*g*). The supervisor may also request changes to the model, which can stop the process in severe cases.

Standards for Internal Models

The internal model of an insurance undertaking under Solvency II has to pass a series of quality checks in order to be approved by the supervisors. The requirements encompass the *use test* (Article 120), *statistical quality standards* (Article 121), *calibration standards* (Article 122), *profit and loss attribution* (Article 123), *validation standards* (Article 124), and *documentation standards* (Article 125).

The use test is a regulatory instrument that is used in many jurisdictions where internal models are used in financial market regulations, both in banking and insurance. Examples are the Basel II rules (see Basel Committee on Banking Supervision, 2006) and the Swiss Solvency Test (SST) (see Eidgenössische Finanzmarktaufsicht FINMA, 2008, X.B) for Swiss insurance undertakings. The rationale behind the use test is that an undertaking that only uses its internal model for the calculation of regulatory capital has an incentive to tweak the model in such a way that the capital requirement is reduced. In contrast, an undertaking that has to demonstrate the use of its internal model for internal purposes beyond the calculation of regulatory capital has an incentive to improve the quality of the model. CEIOPS (2009*a*, 3.3.7 & 3.105-3.127) derives from Article 120 the areas in which the internal model should be used (see figure 3.5) and suggests that a "sufficiently material" use of the internal model should be checked based on ten principles (see table 3.3).

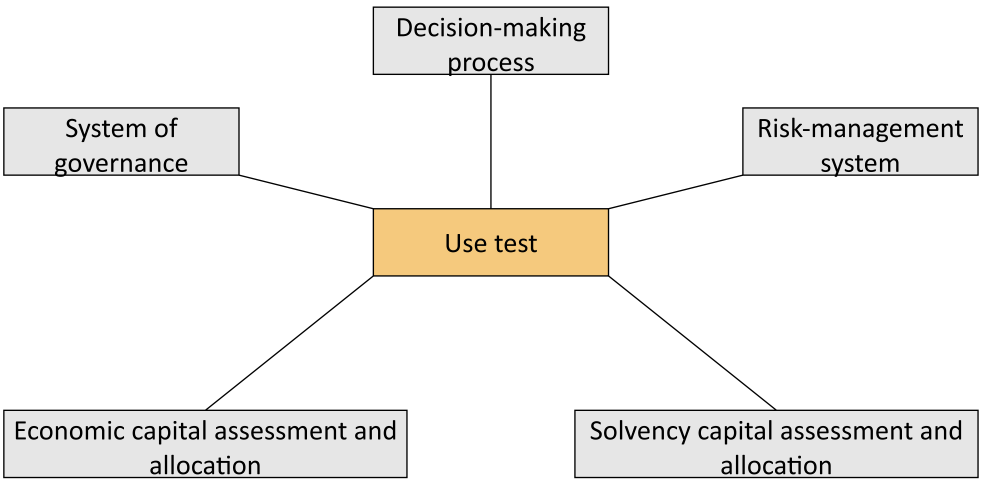

Figure 3.5: Areas of use of an internal model according to CEIOPS (2009*a*, 3.35).

Some of these principles have robustness implications for the internal model. CEIOPS (2009*a*, 3.103) mentions robustness as one aspect of the quality of the internal model which is central to the Foundation Principle. Additionally, Principle 1 states that senior management and the advisory board—both groups usually include non-experts in stochastic and actuarial models—have to understand the internal model, including the methodologies employed and its limitations. While non-robustness is a limitation itself, it is also much harder to understand a model whose output changes widely with small changes of data

Foundation Principle:	The undertaking's use of the internal model shall be sufficiently material to result in pressure to improve the quality of the internal model
Principle 1:	Senior management and the administrative, management or supervisory body, shall be able to demonstrate understanding of the internal model
Principle 2:	The internal model shall fit the business model
Principle 3:	The internal model shall be used to support and verify decision-making in the undertaking
Principle 4:	The internal model shall cover sufficient risks to make it useful for risk management and decision-making
Principle 5:	Undertakings shall design the internal model in such a way that it facilitates analysis of business decisions
Principle 6:	The internal model shall be widely integrated with the risk-management system
Principle 7:	The internal model shall be used to improve the undertaking's risk-management system
Principle 8:	The integration into the risk-management system shall be on a consistent basis for all uses
Principle 9:	The Solvency Capital Requirement shall be calculated at least annually from a full run of the internal model, and also when there is a significant change to the undertaking's risk profile, assumptions underlying the model and / or the methodology arising from decisions or business model changes, and whenever a recalculation is necessary to provide up to date information for decision making or any other use of the model, or to fulfil supervisory reporting requirements

Table 3.3: Principles for the use test as suggested by CEIOPS (2009*a*, pp. 41-47)

or assumptions than to understand a robust one. Furthermore, non-robustness facilitates manipulation of model results by small tweaks of data or assumptions, which CEIOPS (2009*a*, 3.107) explicitly wants to prevent. The main part of compliance with the use test is, according to the principles given by CEIOPS (2009*a*), that the internal model output has to be used in decision-making processes (see Principles 3-5) and risk management (see Principles 6-8). This makes robustness of the internal model important for the undertaking itself, as reliable inputs are of high importance in any decision making process (see Meyburg, 2006). Finally, Principle 9 suggests that in many cases the internal model will have to be run more often than the once-a-year frequency necessary for SCR calculations. Especially for larger undertakings or groups, such a model run typically needs significant computational time and resources. Therefore, it is desirable that the methodologies used in the internal model facilitate (relatively) fast computations. We address this issue in Part III of this thesis.

The statistical quality standards (see Article 121) set quality requirements for the methods used in the internal model and the data used as input to the internal model. Additionally, they clarify the scope of the internal model. We focus on two aspects of the statistical quality standards: the nature of the probability distribution forecast and the required properties of the methods applied in the model.

The probability distribution forecast is defined in Article 13(38) as "a mathematical function that assigns to an exhaustive set of mutually exclusive future events a probability of realisation". Additionally, the Level 1 text only requires that the probability distribution forecast is sufficient to rank risks and to be applied in the internal processes required by the use test (see Article 121(4)). The output of internal models could range from continuous distributions to discrete distributions which only assign probabilities to some key events (CEIOPS, 2009*a*, 5.52). While the former is certainly preferable, at least if the assumptions necessary to obtain a continuous distribution through interpolation, extrapolation, or fitting of a parametric model are justified, the latter might still be reasonable in some cases. Such a model can be approved by supervisors if it performs as well as models that are standard market practice, if it is sufficient considering the undertaking's risk profile and the proportionality principle, or if there are no better models available (CEIOPS, 2009*a*, 5.54). A risk type that can fall into the latter category is operational risk, where usually only few observations are available to build a model with (see Schäl, 2011, Part II, B).

Concerning the methods used in the internal model, the Level 1 text states that they have to be applicable, relevant, and adequate (see Article 121(2)). CEIOPS (2009*a*, 5.62) interprets the term "adequate" as appropriate, up to date, detailed and parsimonious, transparent, and robust and sensitive. We therefore have an explicit requirement for the use of robust statistical methods such as those introduced in chapter 5. Another item on the list of requirements for the methods of the internal model relevant in the context of this thesis is concerned with the assumption underlying the applied methods. The undertaking has to identify and document all assumptions that enter the model together with their implications for model performance and model risk and compare them to alternative assumptions (see CEIOPS, 2009*a*, 5.115-5.118).

The calibration standards (see Article 122) give insurance undertakings the possibility to deviate from the standard time period of one year or the standard risk measure value at risk for internal model purposes or for the calculation of the SCR. Also, undertakings may use approximations when they cannot derive the SCR directly from the probability forecast distribution output by the internal model. In both cases, it is necessary that the level of policyholder protection obtained in this way is equivalent to that of Article 101(3)—that is as provided by a one-year, 99.5% value at risk (see CEIOPS, 2009*a*, 6.50-6.57).

The aim of the validation processes is that the undertaking gains confidence in the functioning and the results of its internal model (see CEIOPS, 2009*a*, 8.15). Accordingly, many different aspects of the internal model have to be validated. CEIOPS (2009*a*, 8.18) requires at least validation of data, methods, assumptions, expert judgement, documentation, systems and IT, model governance, use test, and the validation tools and processes themselves. For the quantitative aspects of the validation procedure, CEIOPS (2009*a*, 8.54) prescribes a number of tools which have to be used by all undertakings, namely the testing of results against experience, testing the robustness of the internal model, stress and scenario testing, and profit and loss attribution.

Testing of results against experience is known as backtesting, although in an insurance context, additional methods that are not part of the backtesting approach in banking might be required, for example due to a lack of data (see CEIOPS, 2009*a*, 8.57). It consists of running the model with input data in such a way that the model predictions can be compared to historical realizations (see, e.g., Crouhy, Galai and Mark, 2001, 4.3.4). The problem of robust backtesting of value at risk models has been addressed by Escaniano and Olmo (2008).

For the test of robustness, CEIOPS (2009*a*, 8.3.3.1.2) suggests a sensitivity analysis and a stability check. The sensitivity analysis consists of changing parameter or model assumptions and re-running the model. Ideally, only significant changes in the assumptions should lead to significant changes in the results. Large differences in the output of the model which are the results of only small changes to the assumptions have to be investigated and the reasons explained to the supervisor. The rationale behind the sensitivity analysis is quite similar to that behind the robustness concepts in statistics (see chapter 5 in particular) that are the basis for our work in the following chapters. The second part of the test of robustness, the stability test, is concerned with reproducibility of model results when the input assumptions for model and parameters do not change, but the (pseudo)random numbers used in a stochastic model do. In this case, any changes in the output have to be "reasonable, explicable and comprehensible" (CEIOPS, 2009*a*, 8.87).

The third of the validation tools, stress and scenario testing, is mainly an extension of the sensitivity analysis to scenarios. It is concerned with the development of adverse scenarios and the investigation of the consequences for the company if these scenarios occur (see Aragonés, Blanco and Dowd, 2008). Its main area of application is in Pillar II, though, where stress and scenario tests are an important part of the own risk and solvency assessment (CEIOPS, 2009*a*, 8.90).

Finally, the profit and loss attribution is a tool for both validation and use test. The yearly exercise of attributing profits and losses to business units leads to a risk profile that should be matched by the internal model (see CEIOPS, 2009*a*, 7.3 for details). In terms of the use test, the profits and losses from this step should be used in the risk governance processes (see CEIOPS, 2009*a*, 3.48).

Details on the documentation of the internal model are provided by CEIOPS (2009*a*, chapter 9). It has to show that the model complies with Articles 120-124 in such a way that an "independent and knowledgeable third party [...] could understand the reasoning and the underlying design and operational details of the internal model" (CEIOPS, 2009*a*, 9.55). This includes details on the employed mathematical methodologies and their assumptions, the model implementation and validation, weaknesses of the model, and changes to it.

3.2.2 Pillar II: Qualitative Requirements

Pillar II of the Solvency II framework is concerned with the system of governance of insurance undertakings. The qualitative requirements are laid down in Articles 41-50 (see table 3.4) and the corresponding Level 2 Advice on implementing measures by CEIOPS (2009*k*).

The governance system has to provide the means to prudently manage and run the insurance company. This includes a transparent organizational structure, clear responsibilities, written documentation of company policies with respect to the areas named in Articles 44-49 of the Directive (see also Article 41) as well as the persons who effectively run the undertaking (at least two, the "four eyes principle", see CEIOPS (2009*k*, 3.3)): they have to be fit (meaning qualified and experienced) and proper (of good repute) (see Articles 42 and 43).

Article 41 of the Level 1 text establishes four functions which are the responsible parties for certain areas of the governance system—namely risk management, internal control, internal audit, and actuarial (see figure 3.6). The undertaking has significant freedom in assigning (groups of) people to these functions, from single persons to organizational units and outside experts (see Oehlenberg, Stahl and Bennemann, 2011, Ch. 3). In some cases the proportionality principle might even allow one person or group of persons to take on more than one function (see CEIOPS, 2008*a*, 45-46).

The internal control function has to make sure that the insurance company complies

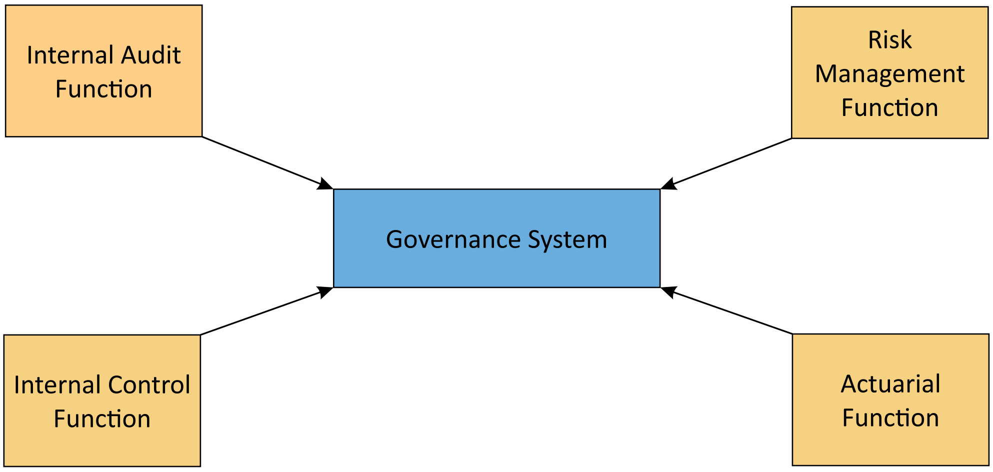

Figure 3.6: Functions in the Solvency II governance system

with "applicable laws, regulations and administrative provisions" (CEIOPS, 2009*k*, 3.254) and report problems to the administrative, management or supervisory body. The independent internal audit function has to evaluate the governance system and, especially, the internal control. The actuarial function is concerned with the calculation of technical provisions and assisting in risk management tasks and the building of internal models. In the context of this thesis, the most important function is risk management, and therefore we take a more detailed look.

The risk management system of an insurance company should encompass well documented risk strategies, processes and reporting procedures (see CEIOPS, 2009*k*, sect. 3.3). The risk categories covered are not exhaustively prescribed by those incorporated in the SCR, but also include other relevant risks (CEIOPS, 2009*k*). The Level 1 text names underwriting and reserving, asset liability management, investment, liquidity, concentration, and operational risk (see Article 44(2)). CEIOPS (2009*k*, sect. 3.3) adds credit, strategic and reputational risk, and also gives advice on risk mitigation techniques which have to be implemented.

The risk management function co-ordinates and monitors the risk management activities in the insurance company. It has to compile a view not only of the current risk situation in the company but also of emerging risks, and report the results to the administrative, management or supervisory body. While the proportionality principle applies to the risk management function, CEIOPS (2008*a*, 52) makes it clear that the end result of the risk management process, namely "getting a clear picture of the undertaking's risks"

(CEIOPS, 2008*a*), is not affected. The risk management function is also responsible for any internal model, its integration, documentation, performance evaluation, and validation (see CEIOPS, 2009*k*, 3.214-3.219).

The results of the activities of the risk management function also enter the Own Risk and Solvency Assessment (ORSA) (see Article 44), which requires the insurance company to “properly assess their own short and long term risk and the amount of funds necessary to cover them” (CEIOPS, 2008*b*, 8) with the purpose of ensuring “that undertakings have robust processes for assessing and monitoring their overall solvency needs”(CEIOPS, 2008*b*, 24). The process of compiling the ORSA should lead to a better understanding by the company of its own risk and solvency situation and enable it to prove this understanding to supervisors. It includes the identification of potential issues that could impact the solvency of the undertaking.

Article	Topic
41	General governance requirements
42	Fit and proper requirements
43	Proof of good repute
44	Risk Management
45	Own risk and solvency assessment
46	Internal control
47	Internal audit
48	Actuarial function
49	Outsourcing
50	Implementing measures

Table 3.4: Articles in Directive 2009/138/EC concerning Pillar II

3.2.3 Pillar III: Transparency Requirements

Pillar III of the Solvency II regulations covers disclosure requirements to supervisors (see Articles 35, and 254) and to the public (see Articles 51-56 and 256). CEIOPS (2009*j*) has issued final Level 2 Advice on the topic.

The Solvency Financial Condition Report (SFCR) is the report that addresses the public. CEIOPS (2009*j*, 3.62) envisages that the report is of interest not only for shareholders and policyholders, but also for rating agencies, financial analysts and competitors, among others. It has to be published once a year (14 weeks after the end of the undertaking’s financial year, at the latest) and in the case of predefined events such as breaches of the SCR and MCR levels (see Article 54 and CEIOPS (2009*j*, 3.522 & sects. 3.2 & 3.7.1)). The content of the SFCR is summarized in table 3.5. The Quantitative Reporting Templates (QRT) are used to summarize quantitative information like the balance sheet, MCR and SCR. They will be detailed in Level 3 documents (CEIOPS, 2009*j*, 3.32 & sect. 3.5). The

Executive Summary
Business and Performance
System of Governance
Risk Profile
Regulatory Balance Sheet
Capital Management
Undertakings with an approved internal model
Annex - Quantitative reporting templates

Table 3.5: Structure of the RTS and SFCR as suggested by CEIOPS (2009*j*)

Report to Supervisors (RTS) is similar to the SFCR, but for the use of the supervisory authorities only. While CEIOPS (2009*j*, sect. 3.3) suggests a very similar structure for both documents (see table 3.5), the RTS should contain additional information such as the ORSA (CEIOPS, 2009*j*, B.4) and non-public QRTs. Just as the SFCR, the RTS usually has to be published once a year and at predefined events. The supervisor may request additional information at any point of the Supervisory Review Process (SRP) (see Article 35(2)a(iii) and CEIOPS (2009*j*, sect. 3.7.2)).

Part II

Robust Risk Management Process

Chapter 4

Probability Distances

4.1 Metric Spaces and Probability Distances

The tool of choice for dealing with distances in a strict mathematical setup is the definition of a metric space. We start by recalling the general definition before moving on to probabilistic objects:

Definition 4.1.1 (see, e.g. Reed and Simon (2005, p. 4))
A *metric space* (M, d) consists of a set M and a function $d : M \times M \to \mathbb{R}$ such that for any $x, y, z \in M$

(i) $d(x, y) \geq 0$ (non-negativity).

(ii) $d(x, y) = 0$ if and only if $x = y$ (identity).

(iii) $d(x, y) = d(y, x)$ (symmetry).

(iv) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality/subadditivity).

The function d is called a *metric*. A function d that satisfies the relaxed condition

(ii)' $d(x, y) = 0$ if $x = y$

instead of (ii) is called semi-metric.

As we deal with random events, we need metrics defined on sets of objects from probability theory such as measures, distributions, and random variables. What makes our task easier is that our main application is the investigation of portfolio value or portfolio loss distributions which are related to real numbers. Assume that μ is a probability measure on the Borel sigma algebra $\mathcal{B}(\mathbb{R})$ on the real line. Then, setting

$$F(x) = \mu((-\infty, x])$$

or, in the other direction,

$$\mu((a, b]) = F(b) - F(a)$$

we have a direct relationship between a measure μ and a distribution function $F = F_\mu$. In some cases, we can get an additional relationship between measures and random variables. Let $\mathfrak{X} = \mathfrak{X}(\mathbb{R}, (\Omega, \mathcal{F}, \mathbb{P}))$ denote the set of real-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Rachev (1991, section 2.5) shows that the set of joint distributions generated by random variables $X, Y \in \mathfrak{X}$ equals the set of all two-dimensional probability measures $\mathcal{M}_2$ in this case,

$$\mathcal{L}\mathfrak{X}_2(\mathbb{R}, (\Omega, \mathcal{F}, \mathbb{P})) = \mathcal{M}_2(\mathbb{R}),$$

as long as $\mathbb{P}$ does not put any mass on atoms of $\mathcal{F}$. This fact gives us some flexibility when choosing the objects we want to consider.

Although we focus on distributions on the real line, we provide more general results where it does not complicate matters too much. Let $\mathcal{M}_k(M)$ denote the space of probability measures on the Borel σ-algebra $\mathcal{B} = \mathcal{B}(M^k)$ (M^k denotes the k-fold Cartesian product of M). Furthermore, let $\{\alpha, \beta, \ldots, \gamma\} \subseteq \{1, 2, \ldots, k\}$. Then $T_{\alpha,\beta,\ldots,\gamma}P$ denotes the marginals in the $\alpha, \beta, \ldots, \gamma$ components of some $P \in \mathcal{M}_k$.

Definition 4.1.2 (see Rachev (1991, pp. 10-11))
A mapping $d : \mathcal{M}_2 \to [0, \infty]$ is called a *probability distance (with parameter $K := K_\mu \geq 1$)* if the following holds:

(i) If $\mu \in \mathcal{M}_2$, then $\mu(\cup_{x \in M}\{(x, x)\}) = 1 \Leftrightarrow d(\mu) = 0$.

(ii) If $\mu \in \mathcal{M}_2$, then $d(\mu \circ B^{-1}) = d(\mu)$ where $B(x, y) := (y, x)$.

(iii) If $\mu_{12}, \mu_{13}, \mu_{23} \in \mathcal{M}_2$ and there exist $\nu \in \mathcal{M}_3$ such that $T_{12}\nu = \mu_{12}$, $T_{13}\nu = \mu_{13}$,

$T_{23}\nu = \mu_{23}$ then

$$d(\mu_{13}) \leq K\left(d(\mu_{12}) + d(\mu_{23})\right).$$

In case $K = 1$, d is called a *probability metric*. If d only satisfies the relaxed condition

(i)' If $\mu \in \mathcal{M}_2$, then $\mu(\cup_{x\in M}\{(x,x)\}) = 1 \Rightarrow d(\mu) = 0$

instead of (i), it is called *probability semi-distance* or *probability semi-metric*, respectively.

We can analogously define a probability distance on random variables. From now on, let (M,d) be a separable metric space (sms). Denote by $\mathfrak{X}(M) = \mathfrak{X}(M,(\Omega,\mathcal{F},\mathbb{P}))$ the set of random variables on $(\Omega,\mathcal{F},\mathbb{P})$ taking values in M and by $\mathscr{L}\mathfrak{X}_2$ the space of joint two-dimensional distributions $\mathbb{P}_{X,Y}$ generated by pairs of random variables $X, Y \in \mathfrak{X}$. We can define a probability distance on $\mathfrak{X}$ via

$$d(X,Y) = d(\mathbb{P}_{X,Y})$$

for any probability distance d.

Definition 4.1.3 (see (Rachev, 1991, pp. 10-11))
A mapping $d : \mathscr{L}\mathfrak{X}_2 \to [0,\infty]$ is called *probability semi-distance* (with parameter $K := K_\mu \geq 1$), if $d(X,Y) = d(\mathbb{P}_{X,Y})$ satisfies the following properties for all $X, Y, Z \in \mathfrak{X}$:

(i) $\mathbb{P}(X = Y) = 1 \Rightarrow d(X,Y) = 0$.

(ii) $d(X,Y) = d(Y,X)$.

(iii) $d(X,Z) \leq K(d(X,Y) + d(Y,Z))$.

d is a probability distance if equivalence holds in (i). For $K = 1$ we have a probability semi-metric or metric, respectively. A probability metric is called *s-ideal* if additionally

(iv) $\mu(cX, cY) = \|c\|^s \mu(X,Y)$ for $c, s \in \mathbb{R}$ (homogeneity).

(v) $\mu(X+Z, Y+Z) \leq \mu(X,Y)$ for any r.v. Z independent of X and Y (regularity).

In the following, we set $K = 1$.

Probability metrics can be classified into different groups by the type of "sameness" considered in (i) of definitions 4.1.2 and 4.1.3 (see, e.g., Rachev (1991) and Rachev, Stoyanov and Fabozzi (2008)). If the metric d depends on the joint distribution of the objects, it is called a *compound metric*. Since we have defined a probability metric in this way, a "compound [...] metric [...] is any probability [...] metric" (Rachev, 1991, p. 39). For random variables this means that a compound metric is zero if and only if the random variables considered are equal almost everywhere.

Example 4.1.4 (p-Average Metric, (see Rachev et al., 2008, p. 85))
Let $X, Y \in \mathfrak{X}^p(\mathbb{R})$ be two real-valued random variables with finite p-th moment on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the *p-average metri*c given by

$$L^p(X,Y) = \mathbf{E}\left[|X-Y|^p\right]^{1/p}$$

for some $p \in [1, \infty)$ is a compound metric. Note that $L^p(X,Y) = 0$ implies that $X = Y$ almost everywhere, but for a sequence $X_1, X_2, \ldots$ of random variables, each $X_i \in \mathfrak{X}^p$, $\mathbf{E}\left[|X_n - X|^p\right]^{1/p} \to 0$ does not imply that $X_n \to 0$ a.s.

We relax the condition in the definitions by ignoring the joint distribution and focusing on the marginal distributions. In this way, we call a probability semi-metric d a *simple probability metric* if for $\mu \in \mathcal{M}_2$ it holds that $T_1\mu = T_2\mu \Leftrightarrow d(\mu) = 0$, or, for random variables $X, Y \in \mathfrak{X}$, if $F_X(x) = F_Y(x) \Leftrightarrow d(X,Y) = 0$. This requirement turns out to be sufficiently strong for our applications and there are various examples of such metrics. We present those simple metrics here (see, e.g., Rachev et al. (2008) for the definitions), that are important in robust statistics and our applications.

Example 4.1.5
The *Kolmogorov metric* d_K, which is known from the central limit theorem, is defined as

$$d_K(F,G) = \sup_{x\in\mathbb{R}} |F(x) - G(x)| \tag{4.1}$$

or, for random variables,

$$d_K(X,Y) = \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)|\,. \tag{4.2}$$

The absolute distance between the two distribution function is often largest close to the means of the distributions, therefore the focus os this metric is on the middle part of the distributions while differences in the tails are less important (see Rachev et al. (2008, p. 77)).

The *Lévy metric* is given by

$$d_L(F,G) = \inf_{\epsilon>0} \{F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon, \forall x \in \mathbb{R})\} \tag{4.3}$$

and analogously for random variables. The Lévy metric can be interpreted as measuring the distance between the graphs of the cumulative distribution functions. Again as in the previous case of the Kolmogorov metric, there is some focus on the middle part of the distributions, because the c.d.f.s are close to zero or one, respectively, in the tails for any distribution by definition. The Lévy metric is a special case of the *Prohorov metric* defined by

$$d_{Pr}(\mu) = \inf_{\epsilon>0} \{T_1\mu(A) \leq T_2\mu(A^\epsilon) + \epsilon, T_2\mu(A) \leq T_1\mu(A^\epsilon) + \epsilon \forall A \subset \mathcal{B}\} \tag{4.4}$$

for $\mu \in \mathcal{M}_2$ and A^ϵ the ϵ-neighborhood of $A \in \mathcal{B} = \mathcal{B}(M)$ w.r.t. the distance of the measurable space (M,d) .

Another example of simple probability metrics are the *L^p-metrics between distribution functions*, defined by

$$\theta_p(F,G) = \left(\int_{-\infty}^{\infty} |F(x) - G(x)|^p \, dx \right)^{1/p} \tag{4.5}$$

for some $p \geq 1$. For larger value of p, the relevance of larger values of $|F(x) - G(x)|$ increases relative to that of smaller values. In the limit as $p \to \infty$, we obtain the Kolmogorov metric from equation (4.2).

In contrast to these examples, the *Wasserstein metric*—the L^p metric between inverse distribution functions—puts the focus on differences in the tails of the distributions. As risk management is concerned with tail-events, the Wasserstein metric turns out to be a natural choice in this context. Therefore, we discuss its properties in detail in section 4.3.

We obtain the class of *primary probability metrics* if we further relax the "sameness" requirement. Let $h : \mathcal{M} \to \mathbb{R}$. For a primary probability metric, condition (i) in definition 4.1.2 and is replaced by

$$h(T_1\mu) = h(T_2\mu) \Leftrightarrow d(\mu) = 0.$$

We can adapt definition 4.1.3 in a similar fashion. In this way, two distributions are equal if they have certain characteristics—for example moments or the sum of moments—in common. Such a characterization is too weak for risk management applications because risk measures often depend on the specific distribution function.

Example 4.1.6 (Engineer's Metric)
The *engineer's metric* is based on the expectation operator,

$$\mathrm{EN}(X,Y) = |\mathrm{E}X - \mathrm{E}Y|$$

for some random variables X, Y with finite first moments.

4.2 Weak Convergence

An important concept we need for robustness considerations is the notion of weak convergence. We discuss it here because it is related to the structure of the space of measures under consideration.

Let (M, d) be a separable complete metric space with Borel σ-field $\mathcal{B}$ and let $C(M)$ denote the space of bounded, continuous functions from M to $\mathbb{R}$. Furthermore, let $\mathcal{M}_R(M)$ denote the set of finitely additive Radon measures on $\mathcal{B}$. For the probability measures $\mathcal{M}$ on $\mathcal{B}$, we have $\mathcal{M} \subset \mathcal{M}_R$. Merkle (2000, Ch. 4) explains that the spaces relate as

$$\begin{array}{rr} \text{original space} & C(M) \\ \text{dual space} & \mathcal{M}_R(M) = C(M)^* \\ \text{double dual} & C(M)^{**} \supset C(M). \end{array}$$

Defining

$$\langle f, \mu \rangle = \int f(x) d\mu(x), \tag{4.6}$$

we have the choice of different topologies on $\mathcal{M}_R$ (and therefore $\mathcal{M}$), among them the weak topology defined by $\langle f, \mu \rangle$ with $f \in C(M)^{**}$ and the weak-* topology defined by $\langle f, \mu \rangle$ where $f \in C(M)$. We select the weak-* topology which is the weakest topology—the topology with the fewest open sets—such that $\langle f, \mu \rangle$ with $f \in C(M)$ is continuous. This is the standard approach in the relevant branch of literature: "The convergence in the weak star topology is usually called the weak convergence in the probabilistic literature. This does not lead to a confusion, since the true weak convergence is never studied." (Merkle, 2000, sect. 4.1).

Definition 4.2.1 (Weak Convergence of Probability Measures)
Let $(\mu_n)_{n\geq 0}$ be a sequence of probability measures in $\mathcal{M}(M)$ and let $\mu \in \mathcal{M}(M)$. We say that μ_n *converges weakly to* μ, $\mu_n \Rightarrow \mu$, if

$$\int \varphi(x)\mu_n(x) \to \int \varphi(x)d\mu(x)$$

for all bounded continuous functions $\varphi \in C(M)$.

Often we are not interested in the space of all probability measures but in a subspace $\mathcal{M}^p$ consisting of the probability measures with p-th finite moment $1 \leq p < \infty$. This is especially true if we want to use the p-Wasserstein distance as a metric, which is only defined for probability measures with this property. In this case, the class of functionals which have to converge in order to have weak convergence of measure is somewhat broader, see Villani (2009, Ch. 6).

Definition 4.2.2 (Weak Convergence of Probability Measures with finite p-th Moment)
Let $(\mu_n)_{n\geq 0}$ be a sequence of probability measures in $\mathcal{M}^p(M)$ and let $\mu \in \mathcal{M}^p(M)$. We say that μ_n *converges weakly to* μ, $\mu_n \Rightarrow \mu$, if

$$\int \varphi(x)\mu_n(x) \to \int \varphi(x)d\mu(x)$$

for all continuous functions φ with $\varphi(x) \leq c(1 + d(x_0, x))^p$ for some constant $c \in \mathbb{R}$ and some—and therefore any—$x_0 \in M$.

4.3 The Wasserstein Metric

We take a closer look at the Wasserstein metric, which was discovered independently by multiple researchers (among them Vasershtein (1969)). We use the more common German spelling "Wasserstein" instead of "Vasershtein". The metric is known under different names in different contexts, such as Mallow's metric, Kantorovich metric, or earth movers distance. Rüschendorf (2002) gives a short time line of its development. Among the various possible degrees of generality for the definition, we follow Bickel and Freedman (1981) and use notation from Shao and Tu (1996).

Definition 4.3.1

Let (M, d) be a separable metric space with norm. Let $1 \leq p < \infty$ and $\mathcal{M}^p = \mathcal{M}^p(M)$ the set of probability distributions γ on $\mathcal{B}(M)$ with $\int \|x\|^p \, \gamma(dx) < \infty$. The *Wasserstein metric* is defined as

$$d_p(F, G) = \inf_{\mathcal{L}((X,Y)) \in \tau(F,G)} \{E \left[d(X - Y)^p\right]^{1/p}\} \tag{4.7}$$

$$= \inf_{\gamma \in \tau(F,G)} \left(\int_{M \times M} d(x - y)^p d\gamma(x, y) \right)^{1/p} \tag{4.8}$$

where $\tau(F, G)$ denotes the set of joint distribution functions of pairs of M-valued random variables random variables X, Y with marginals $F \in \mathcal{M}^p$ and $G \in \mathcal{M}^p$, respectively.

We write $d_p(U, V) = d_p(F_U, F_V)$ for the distance between the distributions of two random variables U, V with distributions F_U and F_V, respectively, in $\mathcal{M}^p$.

The definition in equation (4.8) leads to interpretations in the context of transportation theory: the Wasserstein metric provides a measure for the cost of moving the (probability) mass distributed on M according to F such that it is distributed according to G.

Before using the Wasserstein metric and investigating its properties, we have to answer the question whether the functional in definition 4.3.1 is indeed a metric. We can obtain this result if the infimum in equation (4.8) is attained.

Proposition 4.3.2

The infimum in equation (4.8) is attained. We have the duality relationship

$$d_p(F, G) = \sup \left\{ \int f d(F - G) : |f(x_1) - f(x_2)| \leq \|x_1 - x_2\| \, \forall x_1, x_2 \in B \right\}.$$

Proof. Kellerer (1985, thm. 1) shows the first part for tight measures. By Ulam's theorem B.2.3, every probability measure in $\mathcal{M}^p$ is tight, so Kellerer's (1985) proof applies. He also shows the second part but tightness is not necessary in this case.

Dall'Aglio (1956) has shown the first part for M equal to the real line. □

Now that we have this result, the main idea is to find such measures for which the infimum in equation (4.8) is attained and then to apply Minkowski's inequality. This technique is also used in the proof of proposition 4.3.8. Here, we have the basic result that the Wasserstein metric is indeed a metric.

Proposition 4.3.3

The map $d_p(F, G)$ is a metric on $\mathcal{M}^p$.

Proof. See Bickel and Freedman (1981, Lemma 8.1) for the proof. □

As already noted above, we mainly work on the real line. In this case, proposition 4.3.4 gives us a useful representation in terms of the generalized inverses of cumulative distribution functions.

Proposition 4.3.4 (Bickel and Freedman (1981))

If the separable metric space (M, d) is the real line equipped with usual norm, then

$$d_p(F, G) = \left(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt \right)^{1/p}. \tag{4.9}$$

Proof. One way to prove the proposition is to show that the infimum in equation (4.8) is attained for the random variables $X = F^{-1}(U)$ and $Y = G^{-1}(U)$ where U has a continuous uniform distribution $U \sim \mathrm{U}(0, 1)$. (The rvs. X and Y obviously have laws F and G, respectively. See, e.g., Glasserman (2004, Ch. 2.2.1)). Vallender (1973) accomplishes this task for the case $p = 1$ using conditional expectations and basic inequalities. Mallows (1972) extends the result to $p = 2$ by showing the equality directly, first for measures F, G with bounded support (using the Hoeffding-Fréchet bounds known from copula theory (Hoeffding, 1940; Fréchet, 1951)), then for the general case.

The case $p > 0$ is proved by Dall'Aglio (1956, thm. IX, in Italian) using a similar technique as Mallows. Major (1978, thm. 8.1) proves a generalization to convex functions. He first shows that for measures F, G concentrated on a finite set, the random variables for which the infimum is attained can be characterized uniquely and agree with the random variables X and Y above. Taking limits, he extends the argument to bounded intervals and then to the real line. □

The representation shows that—at least for $p > 1$—the value of the Wasserstein metric between two distribution functions is influenced more by differences in their tails rather than differences in the center of the distributions. This desirable feature of the metric is in contrast to the other examples of simple probability metrics we have considered in section 4.1. Additionally, it suggests how the infimum in the definition is attained in this special case.

Corollary 4.3.5

The infimum in equation (4.8) for $M = \mathbb{R}$ is attained combining the marginal distributions F, G on the real line with the comonotonicity copula $M_2(u) = \min(u_1, u_2), u = (u_1, u_2) \in \mathbb{R}^2$ (see, e.g., Durante and Sempi (2010)).

Proof. Note from proposition 4.3.4 that the infimum in equation (4.8) is attained for the random variables $X = F^{-1}(U)$ and $Y = G^{-1}(U)$ where U has a continuous uniform distribution $U \sim \mathrm{U}(0,1)$. But these random variables are the result of coupling F and G using the comonotonicity copula (see, e.g., the review paper by Dhaene, Denuit, Goovaerts, Kaas and Vyncke (2002, thm. 2)). □

Convergence with respect to the Wasserstein metric is closely related to the concept of weak convergence of probability measures, as the following results show. In particular, convergence in the p-Wasserstein metric is equivalent to weak convergence in $\mathcal{M}^p$ and implies weak convergence in $\mathcal{M}$.

Proposition 4.3.6 (Bickel and Freedman (1981, Lemma 8.3))

Let $F_n, F \in \mathcal{M}^p$. The following statements are equivalent:

(i) $d_p(F_n, F) \to 0$ as $n \to \infty$

(ii) $F_n \to F$ weakly in $\mathcal{M}$ and $\int \|x\|^p \, F_n(dx) \to \int \|x\|^p \, F(dx)$.

(iii) $F_n \to F$ weakly in $\mathcal{M}$ and $\|x\|^p$ is uniformly F_n integrable.

(iv) $\int \varphi dF_n \to \int \varphi dF$ for every continuous φ such that $\varphi(x) = O(\|x\|^p)$ at infinity.

Proof. See Bickel and Freedman (1981, Lemma 8.3). □

Corollary 4.3.7

The Wasserstein distance metrizes the weak topology on the space of distributions with finite p-th moment, $1 \leq p < \infty$, that is

$$\mu_n \Rightarrow \mu \text{ in } \mathcal{M}^p$$

if and only if

$$d_p(\mu_n, \mu) \to 0$$

as $n \to \infty$.

Proof. We get the result directly from proposition 4.3.6 and the definition of weak convergence in $\mathcal{M}^p$. □

With regard to applications in finance outside of risk management, we can make the case that the Wasserstein metric is a sensible choice too.

Proposition 4.3.8 (see Bickel and Freedman (1981, ch. 8))
Let $X, Y \in \mathfrak{X}$, $F_1, F_2, G_1, G_2 \in \mathcal{M}^2$, let $U_i, i = 1, \dots, m$ and $V_j, j = 1, \dots, m$ be independent rvs with law in $\mathcal{M}^2$.

(i) Scaling properties:

$$d_p(aX, aY) = |a|\, d_p(X, Y) \quad \text{for any } a \in \mathbb{R}$$
$$d_p(LX, LY) \leq \|L\|\, d_p(X, Y) \quad \text{for any linear operator } L \text{ on } M.$$

(ii) Convexity: Let $\alpha \in (0, 1)$, $p = 2$. Then

$$d_p(\alpha F_1 + (1-\alpha)F_2, \alpha G_1 + (1-\alpha)G_2) \leq \alpha d_p(F_1, G_1) + (1-\alpha) d_p(F_2, G_2).$$

(iii) Subadditivity:

$$d_2\left(\sum_{i=1}^m U_i, \sum_{i=1}^m V_i\right) \leq \sum_{i=1}^m d_2(U_i, V_i).$$

(iv) The Wasserstein metric is 1-ideal.

Proof. (i) This follows directly from proposition 4.3.2, the scaling properties of the respective norms and linearity of the expectation.

(ii) By proposition 4.3.2, we can find measures $\overline{\mu}, \hat{\mu}$ on $M \times M$ such that $d_2(F_1, G_1) = \mathbf{E}^{\overline{\mu}} \|X_1, Y_1\|$ and $d_2(F_2, G_2) = \mathbf{E}^{\hat{\mu}} \|X_2, Y_2\|$ where X_1, X_2, Y_1, Y_2 have marginal distributions F_1, F_2, G_1, G_2, respectively, and $\mathbf{E}^{\cdot}$ denotes expectation w.r.t. the given measure. Choose a measure μ on $M \times M \times M \times M$ with the $\overline{\mu}$ and $\hat{\mu}$ as two-dimensional marginals (cf. Bickel and Freedman (1981, Proof of Lemma 8.1)) so that

$$
\begin{aligned}
& d_2^2(\alpha F_1 + (1-\alpha)F_2, \alpha G_1 + (1-\alpha)G_2) \\
\leq & \mathbf{E}^{\mu}\left[\|\alpha X_1 + (1-\alpha)X_2 - \alpha Y_1 - (1-\alpha)Y_2\|^2\right]^{1/2} \\
\leq & \mathbf{E}^{\mu}\left[\|\alpha X_1 - \alpha Y_1\|^2 + \|(1-\alpha)X_2 - (1-\alpha)Y_2\|^2\right]^{1/2} \\
\leq & \mathbf{E}^{\overline{\mu}}\left[\|\alpha X_1 - \alpha Y_1\|^2\right]^{1/2} + \mathbf{E}^{\hat{\mu}}\left[\|(1-\alpha)X_2 - (1-\alpha)Y_2\|^2\right]^{1/2} \\
= & \alpha \mathbf{E}^{\overline{\mu}}\left[\|X_1 - Y_1\|^2\right]^{1/2} + (1-\alpha)\mathbf{E}^{\hat{\mu}}\left[\|X_2 - Y_2\|^2\right]^{1/2} \\
= & \alpha d_2(F_1, G_1) + (1-\alpha) d_2(F_2, G_2)
\end{aligned}
$$

using Minkowski's inequality.

(iii) Use the same strategy as for part iii) or see Bickel and Freedman (1981, Lemma 8.6)

(iv) Regularity is a consequence of part ii), homogeneity is part i).

□

Suppose we want to measure the distance between two portfolio value distributions on a fixed time horizon T. The value is given in Euro, but our home currency are US dollars. The scaling property makes sure that conversion with today's forward FX-rate for T does not lead to unexpected changes of the metric.

The notion of an ideal metric is used in the context of (generalized) central limit theorems, see Rachev et al. (2008, Ch. 4), where we want to know how far the limit distribution obtained via such a theorem could be from the distribution of the sample. Sums of random variables occur in different contexts in finance applications. Examples are a portfolio consisting of (a "sum" of) different assets or a monthly log-returns which are the sums of daily log-returns.

The statistical robustness concepts we discuss in the next section are based on statistical functionals, that is, maps from a set of distribution functions to the real numbers. As we work with historical data or data obtained from simulations, we usually face the situation that we only have an empirical distribution function as an input to a statistical functional instead of the distribution function from which the data is sampled. Therefore, we are interested to control the distance between the two distribution functions. Bolley, Guillin and Villani (2007) provide the following probabilistic error bounds for the Wasserstein metric.

In the following, let $\hat{F}_n$ denote a (random) empirical measure associated with an i.i.d. sample $X_1, \ldots, X_n$ of size n of a random variable X with distribution F,

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

where δ_{X_i} denotes the Dirac measure. With a slight abuse of notation, we also denote by $\hat{F}_n$ the empirical distribution function associated with $X_1, \ldots, X_n$..

Theorem 4.3.9 (see Bolley et al. (2007, thms. 2.7 & 2.8))
Let $q \geq 1$ and let $F \in \mathcal{M}^q(\mathbb{R}^d)$. Then

(i) For any $p \in [1, q/2)]$, $\delta \in (0, q/p - 2)$ and $d' > d$ there exists some constant $N_0 = N_0(F, q, p)$ such that

$$\mathbb{P}\left[d_p^p(F, \hat{F}_N) > \epsilon\right] \leq \epsilon^{-q} N^{-\frac{q}{2p} + \frac{\delta}{2}}$$

for any $\epsilon > 0$ and any $N > N_0 \max\left(\epsilon^{-q\frac{2p+d'}{q-p}}, \epsilon^{d'-d}\right)$.

(ii) For any $p \in [q/2, q)]$, $\delta \in (0, q/p - 1)$ and $d' > d$ there exists some constant $N_0 = N_0(F, q, p)$ such that

$$\mathbb{P}\left[d_p^p(F, \hat{F}_N) > \epsilon\right] \leq \epsilon^{-q} N^{1 - \frac{q}{p} + \delta}$$

for any $\epsilon > 0$ and any $N > N_0 \max\left(\epsilon^{-q\frac{2p+d'}{q-p}}, \epsilon^{d'-d}\right)$.

(iii) Let $p \geq 1$ and assume that $\mathscr{E}_\alpha := \int_{\mathbb{R}^d} e^{\alpha|x|} dF$ is finite for some $\alpha > 0$. Then there exist, for all $d' > d$, constants K and N_0, depending only on d, α, and $\mathscr{E}_\alpha$ such that

$$\mathbb{P}\left[d_p^p(F, \hat{F}_N) > \epsilon\right] \leq e^{KN^{1/p} \min(\epsilon, \epsilon^2)}$$

for any $\epsilon > 0$ and $N > N_0 \max\left(\epsilon^{-(2p+d')}, 1\right)$.

It would be useful for robustness considerations to have a (uniform) Glivenko-Cantelli type of result for the Wasserstein distance $d_p, p \geq 1$. Unfortunately, the only results available are either restricted to the case $p = 1$ (see definition 5.1.11 and the following discussion) or by strong assumptions on the underlying distribution F:

Theorem 4.3.10 (see Boissard and Gouic (2011))
Let (M, d) be a measured Polish space and for $X \subset M$ define the covering number $N(X, \delta)$

by

$$N(X,\delta) = \min\left\{ n \middle| \exists x_1, \dots, x_n \text{ with } X \subset \bigcup_{i=1}^{n} B(x_i,\delta) \right\}$$

where $B(x,\delta)$ denotes the δ-Ball centered around x. Choose $t > 0$ and $F \in \mathcal{M}(M)$ with support in some X with $N(X,t) < \infty$. Then

$$\mathbf{E}\left[d_p(F,\hat{F}_n)\right] \leq c\left(t + n^{-1/(2p)} \int_t^{d/4} N(X,\delta)^{1/(2p)} d\delta\right).$$

Theorem 4.3.11 (see Horowitz and Karandikar (1994, thm. 1.1))
Consider the space $(\mathbb{R}^d, |.|)$ and set $c := \int |u|^{d+5}\, dF < \infty$ for $F \in \mathcal{M}(\mathbb{R}^d)$. Then there is a constant $C > 0$ depending only on c and d such that

$$\mathbf{E}\left[d_2^2(F,\hat{F}_n)\right] \leq Cn^{\frac{-2}{d+4}}.$$

4.4 The Wasserstein Metric and the L^p-Metrics

The p-Wasserstein metric d_p has similarities with both the L^p metric between distribution functions θ_p and the p-average metric L^p. In special cases it agrees with one or both of them. To avoid confusion, we examine the relationship more closely.

Recall the definition of the L^p metric between distribution functions.

Definition 4.4.1 (L^p metric between distribution functions)
Let $F, G \in \mathcal{M}^1$. Then

$$\theta_p(F,G) := \left(\int_{-\infty}^{\infty} |F(t) - G(t)|^p\, dt\right)^{\frac{1}{p}}, p \geq 1$$

defines the L^p metric between distribution functions. For X and Y having distribution functions F_X and F_Y respectively, we define, with a slight abuse of notation,

$$\theta_p(X,Y) := \theta_p(F_X, F_Y).$$

Note that θ_p is defined for distributions with finite first moment for any $p \geq 1$. It is not necessary that the p-th moment exists in order for θ_p to be defined ($p > 1$). Both d_p and θ_p are simple probability metrics and for $p = 1$ we have that $d_1(F,G) = \theta_1(F,G)$. In this case, the integral in either of the definitions calculates the area between the distribution functions, but in a different order of integration. This is illustrated by figure 4.1.

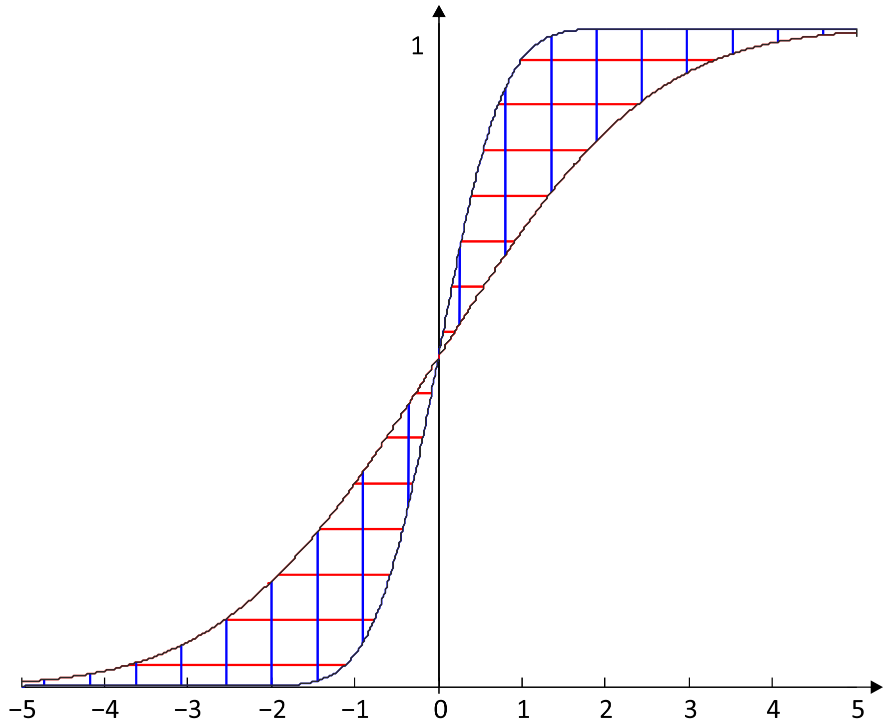

Figure 4.1: Integration directions for the Wasserstein metric (horizontal lines) and L^p metric between distribution functions (vertical lines)

For $p > 1$ the metrics differ, and the larger p the more the Wasserstein metric focuses on the tails and the more the L^p metric between distribution functions focuses on the center. We illustrate this with an example.

Example 4.4.2

We compare the distance of a standard normal distribution and a mixture distribution

$$G(x) = (1-\epsilon)\Phi(x) + \epsilon H(x;\vartheta)$$

where $\epsilon \in (0,1)$, Φ is the c.d.f. of the standard normal distribution, and H is a cdf of second distribution with parameter vector ϑ. We use the exponential distribution $Exp(\lambda)$ where λ is the mean of the distribution. This leads to differences in the right tail between the distributions. Table 4.1 shows the values of d_p and θ_p for some exemplary parameters. We can see that the distance as measured with the Wasserstein metric increases with p, while the distance decreases when measuring it with the L^p metric between distribution

	1	2	3	4
d_p	0.054	0.067	0.075	0.085
θ_p	0.054	0.039	0.038	0.038

Table 4.1: Values of the Wasserstein metric and the L^p metric between distribution functions for the standard normal distribution and a mixture distribution with $H = Exp(0.5)$ and $\epsilon = 0.1$, $p = 1, 2, 3, 4$.

functions.

Next, we consider the p-average metric. Recall that $\mathfrak{X}^s$ denotes the set of random variables with finite s-th moment.

Definition 4.4.3 (The L^p-metric)
Let $X, Y \in \mathfrak{X}^p$. Then

$$L_p(X, Y) := (\mathbf{E}[|X - Y|^p])^{\frac{1}{p}}, p \geq 1$$

defines the L^p-metric between random variables.

The p-average metric is a compound metric—it takes the joint distribution of the random variables or marginal distributions into account. Defining an analog p-average metric between distribution functions $F, G \in \mathcal{M}^p$ therefore only is possible if the joint distribution $H \in \Gamma_p^2$ is known. Then, we can write

$$L^p(H) := (\mathbf{E}[|X - Y|^p])^{\frac{1}{p}}, p \geq 1, (X, Y) \sim H.$$

Corollary 4.4.4
The Wasserstein metric d_p between the marginal distributions of two random variables is equal to the p-average metric between the random variables if their joint distribution is given by the comonotonicity copula.

Proof. This is a direct consequence of Corollary 4.3.5. □

Proposition 4.4.5
Let $X_n \in \mathfrak{X}^2$ be a sequence of random variables with distributions F_n, let $X \in \mathfrak{X}^2$ be a random variable with distribution F. Then $X_n \rightarrow_{L^2} X$ implies $F_n \rightarrow_{d_2} F$.

Proof. We have

$$\begin{aligned} 0 \leq d_2(F_n, F) &= \inf_{(Y_n,Y)\in\tau_{F_n F}} \{E\left[\|Y_n - Y\|^2\right]^{1/2}\} \\ &\leq E\left[\|X_n - X\|^2\right]^{1/2} \\ &= L^2(X_n, X) \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

□

Example 4.4.6

We consider two standard normally distributed random variables X, Y and specify their joint distribution via an Archimedean copula, namely the Gumbel copula, which is given by

$$C_\theta^{GH}(u) = \exp\left(-((-\log(u_1))^\theta + (-\log(u_2))^\theta)^{1/\theta}\right)$$

for some $\theta \geq 1$. The independence copula is a special case for $\theta = 1$, the comonotonicity copula is its limit as $\theta \to \infty$. The L^1 and L^2 distances between X and Y for $\theta \in [1, 50]$ are exhibited in figure 4.2. The limit as $\theta \to \infty$ is zero, equal to the Wasserstein distance between the marginals of X and Y.

The Wasserstein distance is a simple metric and so does not imply anything about the joint distribution between two marginals. We would therefore assume that convergence in the Wasserstein metric does not imply convergence in the p-average metric in general. This is indeed the case as the following example shows.

Example 4.4.7

We give a counterexample to show that the converse implication in proposition 4.4.5 does not hold. Take a sequence of distributions $F_n = B(1, 0.5 + \frac{1}{n})$[1] and a distribution $F = B(1, 0.5)$. Assume that the joint distribution is induced by the countermonotonicity

[1] $B(\alpha, \beta)$ denotes the beta distribution with mean $\frac{\alpha}{\alpha+\beta}$.

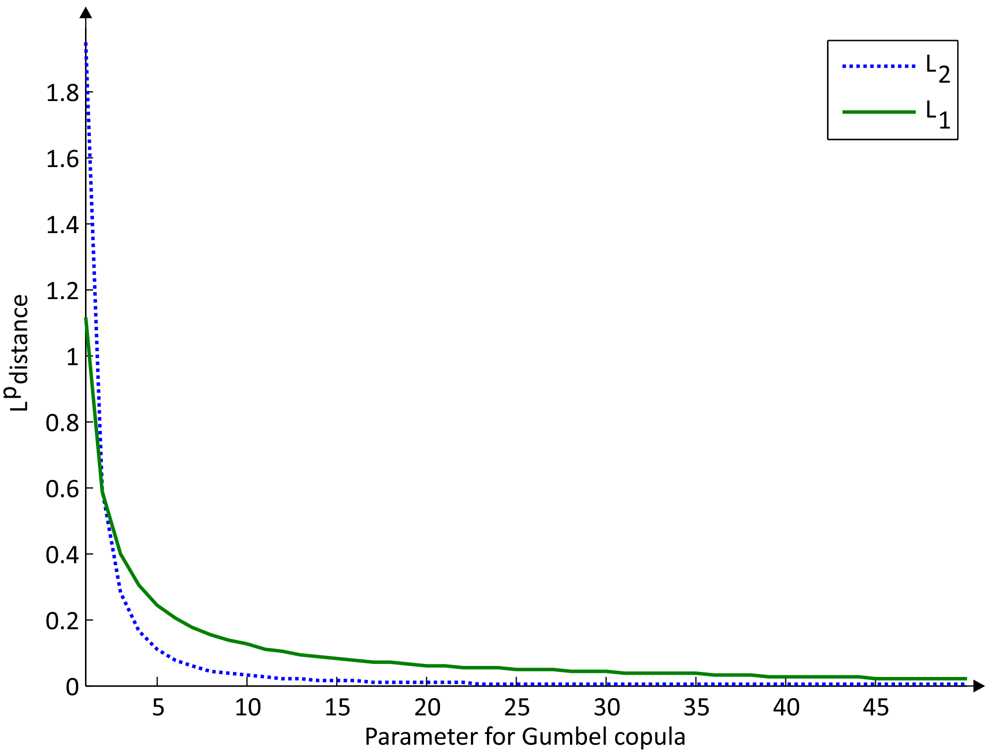

Figure 4.2: L^p distance for standard normal random variables with joint distribution given by the Gumbel copula with parameter $\theta \in [1, 50]$

copula. Then

$$\begin{aligned} d_2(F_n, F)^2 &= \int_0^1 \left| \mathbf{1}_{[0,0.5]}(x) - \mathbf{1}_{[0,0.5+\frac{1}{n}]} \right|^2 dx \\ &= \int_{0.5}^{0.5+\frac{1}{n}} dx \\ &= \frac{1}{n} \to 0 \end{aligned}$$

but

$$\begin{aligned} L^2(X_n, X)^2 &= \int_0^1 \left| \mathbf{1}_{[0,0.5]}(x) - \mathbf{1}_{[0.5-\frac{1}{n},1]}(x) \right|^2 dx \\ &= \int_0^1 1 - \mathbf{1}_{[0.5-\frac{1}{n},0.5]}(x) dx \\ &= \int_0^{0.5-\frac{1}{n}} dx + \int_{0.5}^1 dx \\ &= 1 - \frac{1}{n} \to 1. \end{aligned}$$

Chapter 5

Robustness in Statistics

A robust quantitative risk management process has to be based on robust statistical methods. Before we discuss such methods, we want to make the notion of robustness more precise. We quote Huber and Ronchetti (2011, p. 5), who answer the question of the properties of a robust statistical procedure as follows:

- "Efficiency: It should have a reasonably good [...] efficiency at the assumed model.
- Stability: It should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly[...].
- Breakdown: Somewhat larger deviations from the model should not cause a catastrophe."

Robustness in this sense is distinct from the concept of *resistance*. A statistic is called resistant, if changing a small part of the data—however big the change—does not change the statistic substantially, see Mosteller and Tukey (1977). Fortunately, it turns out that the two notions are equivalent in many practical applications (see theorem 5.1.7) so that we can talk about robustness in the following.

We need some notation for the discussion, which we borrow from Huber and Ronchetti (2011). In the following, we are interested in statistics that can be written as statistical functionals T. They depend on a sample $(x_1, \ldots, x_n)$ with corresponding empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i < x}. \tag{5.1}$$

With a slight abuse of notation we also denote the empirical measure by F_n, which is given by the set function

$$F_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Here, δ_{x_i} is the Dirac-measure that puts mass 1 on x_i. We write

$$T_n(x_1, \ldots, x_n) = T(F_n)$$

for a statistic that is defined at least on the space of empirical measures. $\mathcal{L}_F(T_n)$ denotes the distribution of T_n at F, that is, if the distribution underlying the observations $(x_1, \ldots, x_n)$ is F.

Definition 5.0.8 (Fisher Consistent Functionals, see Huber and Ronchetti (2011, p. 9)) Let F be the distribution underlying the observations $(x_1, \ldots, x_n)$ with empirical distribution function F_n. Let T be a functional defined for all empirical distribution functions. Then, T can be extended to the space of probability distributions $\mathcal{M}$ by setting

$$T(F) = \lim_{n \to \infty} T(F_n)$$

if the limit on the right-hand side exists in probability. Such a functional T is called *(Fisher) consistent* at F.

5.1 Qualitative Robustness

The intuition of robustness given in the introductory paragraph to this chapter refers to "small" and "somewhat larger" deviations from the model assumptions. This has to be made precise. The probability distances introduced in chapter 4 allow us to do this.

We use the setup outlined by Hampel (1971). Let $\mathcal{M}$ denote the set of probability measures on some measurable space $(\Omega, \mathcal{B}(\Omega))$ and $\mathcal{M}_n \subset \mathcal{M}$ the set of discrete probability measures with atoms equal to $\frac{m}{n}, m \in \{0, 1, \ldots, n\}$ for $n \in \mathbb{N}$. We observe the first n elements of a sequence $\omega_1, \omega_2, \ldots$ of independent and identically distributed realizations of a random variable with distribution F, which can be modeled on a product space $(\Omega^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}}, F^{\mathbb{N}})$. Observe the correspondence between observations $(\omega_1, \ldots, \omega_n) \in \Omega^n$ and measures F_n in $\mathcal{M}_n$ which is unique up to permutations of the observations. We consider sequences of measurable mappings $(T_n)_{n\geq 0}$ with $T_n : \mathcal{M}_n \to \mathbb{R}^k$ for some $k \geq 1$. Each random measure F_n corresponds to the first n elements of a sequence $\omega_1, \omega_2, \ldots$ of elements from Ω. Therefore, T_n induces a distribution on $\mathbb{R}^k$, mapping $F \mapsto \mathcal{L}_F(T_n)$. As we are interested in measuring risks of real valued random variable through real valued functionals—we discuss this in chapter 6—we use $\Omega = \mathbb{R}$ and $k = 1$ in the following.

Definition 5.1.1

Let T be a statistical functional and $F_0 \in \mathcal{M}$ some probability distribution. T is called *continuous* at F_0 w.r.t. some metric d on $\mathcal{M}$ that metrizes the weak topology if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for any $F \in \mathcal{M}$

$$d(F_0, F) < \delta \Rightarrow |T(F_0) - T(F)| < \epsilon. \tag{5.2}$$

Before we continue with the definition of qualitative robustness, we consider some examples.

Example 5.1.2 (Trimmed Mean)

Consider the α-trimmed mean with corresponding functional $T_\alpha(F) = \int_\alpha^{1-\alpha} F^{-1}(t)dt$ for some $\alpha \in (0,1)$. Fix $F_0 \in \mathcal{M}(\mathbb{R})$ and $\epsilon > 0$. Then for any $F \in \mathcal{M}(\mathbb{R})$ we have

$$\begin{aligned}
|T_\alpha(F_0) - T_\alpha(F)| &= \left| \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} F_0^{-1}(t)dt - \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} F^{-1}(t)dt \right| \\
&= \frac{1}{1-2\alpha} \left| \int_\alpha^{1-\alpha} F_0^{-1}(t) - F^{-1}(t)dt \right| \\
&\leq \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} \left| F_0^{-1}(t) - F^{-1}(t) \right| dt \\
&\leq \frac{1}{1-2\alpha} \int_0^{1} \left| F_0^{-1}(t) - F^{-1}(t) \right| dt \\
&= \frac{1}{1-2\alpha} d_1(F_0, F)
\end{aligned}$$

where $d_1(.,.)$ denotes the Wasserstein distance with $p = 1$. Choosing $\delta = (1-2\alpha)\epsilon$, this shows that the α-trimmed mean is continuous w.r.t. this distance at any $F_0 \in \mathcal{M}^1(\mathbb{R})$.

Example 5.1.3 (Mean)

Consider the mean with corresponding functional $T(F) = \int_\mathbb{R} x dF$. Fix $F \in \mathcal{M}^1(\mathbb{R})$ with $T(F) < \infty$ and set

$$F_n = \frac{n-1}{n} F + \frac{1}{n} \delta_{n^2}.$$

Then $F_n \in \mathcal{M} \forall n$ and $F_n \to F$ weakly: for any $\epsilon > 0$, set $n_0 = \lfloor \frac{1}{\epsilon} \rfloor$. We have:

$$\begin{aligned} |F(x) - F_n(x)| &= \Big| \underbrace{\frac{1}{n}F(x)}_{\in[0,1/n]} - \underbrace{\frac{1}{n}\mathbf{1}_{x \geq n^2}}_{\in[0,1/n]} \Big| \\ &\leq \frac{1}{n} \\ &< \epsilon \forall x \in \mathbb{R} \forall n \geq n_0. \end{aligned}$$

But for the mean, we have

$$\begin{aligned} |T(F) - T(F_n)| &= \left| \int_{\mathbb{R}} xd\left(\frac{1}{n}(F - \delta_{n^2})\right) \right| \\ &= \left| \frac{1}{n}T(F) - \frac{1}{n}T(\delta_{n^2}) \right| \\ &= \left| \frac{1}{n}T(F) - n \right| \\ &\to \infty \qquad \text{as } n \to \infty. \end{aligned}$$

Consequently, the mean is not continuous w.r.t. any metric metrizing the weak topology on all of $\mathcal{M}$.

Weak continuity is a property that can be checked easily for many functionals as in the examples above. For the class of L-statistics, which encompass many functionals that are relevant for risk measurement in practice, Huber and Ronchetti (2011) provide a result that further simplifies this task. An L-statistic is generally of the form

$$T_L(F) = \int F^{-1}(s)\mu(ds)$$

where μ is a signed measure on $(0, 1)$, $J : [0, 1] \to \mathbb{R}$. Sometimes the less general equation

$$T_L(F) = \underbrace{\int xJ(F(x))dF(x)}_{=:T_{L1}} + \underbrace{\sum_{i=1}^{m} a_i F^{-1}(\alpha_i)}_{=:T_{L2}} \tag{5.3}$$

is used as a definition. Here $m \in \mathbb{N}$ is fixed, and α_i, a_i with $0 < \alpha_i < 1$ are constants. L-statistics combine robustness derived from week continuity (see theorem 5.1.4) with efficiency of location estimates (see Bickel and Lehmann (1975)). L-statistics are important in our context because most of the distribution-based or law-invariant risk measures discussed in the literature are in fact in this class (see Cont et al. (2010, sect. 2.1) and our discussion in section 6.1).

Theorem 5.1.4 (see Huber and Ronchetti (2011, thm. 3.7))
Let $\mu = \mu^+ - \mu^-$ be a signed measure on $(0, 1)$ and let $T(F) = \int F^{-1}(s)\mu(ds)$. Define $\alpha = \max\{x | \text{supp}(\mu) \subset [x, 1-x]\}$. If $\alpha > 0$, then T is weakly continuous at F_0, provided that μ does not put any pointmass on a discontinuity point of F_0^{-1}. If $\alpha = 0$, this implies that T is discontinuous.

Example 5.1.5 (α-Quantiles)
We define the (lower) α-quantile by

$$q_{(\alpha)}(F) := \inf\{x | F(x) \geq \alpha\}$$

for some probability distribution $F \in \mathcal{M}(\mathbb{R})$. We can write $q_{(\alpha)}(F)$ as a functional

$$T_{q_{(\alpha)}}(F) = F^{-1}(\alpha) = \int_0^1 F^{-1}(u)\delta_\alpha(u)$$

where F^{-1} denotes the generalized inverse distribution function

$$F^{-1}(y) := \inf_{x \in \mathbb{R}} \{F(x) \geq y\}$$

and δ_α is the Dirac measure that put a point-mass of one on α. The α-quantile is an L-statistic with $\mu = \delta_\alpha$. If $\alpha > 0$, we are in the setup of theorem 5.1.4. Therefore, it is weakly continuous for all $F \in \mathcal{M}(\mathbb{R}))$ such that F^{-1} is continuous at $\alpha > 0$. As a further consequence of theorem 5.1.4, minimum and maximum are not weakly continuous.

Definition 5.1.6 (see Huber and Ronchetti (2011, p. 11))
Let d, d' be distances that metrize the weak topology on the set $\mathcal{M}$ of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. A sequence of functionals T_n is called *(qualitatively) robust* if for each $\epsilon > 0$ there exists a $\delta > 0$ and an $n_0 \in \mathbb{N}$ such that, for all $F \in \mathcal{M}$ and $n \geq n_0$,

$$d(F_0, F) \leq \delta \Rightarrow d'(\mathcal{L}_{F_0}(T_n), \mathcal{L}_F(T_n)) \leq \epsilon. \tag{5.4}$$

Hampel's theorem is a classical result that relates qualitative robustness to continuity of a functional, which is much easier to check. The only restriction on the metrics in the definition of robustness is the requirement that they metrize the weak topology. Originally, Hampel (1971) selected the Prohorov distance, $d = d' = d_{Pr}$. We show the result obtained by Huber and Ronchetti (2011, thm. 2.21).

Theorem 5.1.7 (Hampel's Theorem)
Set d to the Lévy metric and d' to the Prohorov metric. Assume that (T_n) derives from a functional T so that it is consistent in a neighborhood of F_0. Then T is continuous at F_0 if and only if (T_n) is qualitatively robust at F_0.

Proof. See the proof by Huber and Ronchetti (2011, thm. 2.21). □

Example 5.1.8
By Hampel's theorem, the (lower) α-quantile is qualitatively robust at any $\mu \in \mathcal{M}$ under the conditions stated in example 5.1.5.

Qualitative robustness "depends on the specific choice of metric" (Krätschmer, Schied and Zähle, 2012*b*), as also noted by Huber and Ronchetti (2011, p. 42). As we prefer to use the Wasserstein metric over the Lévy and Prohorov metrics, the usefulness of the version of Hampel's theorem stated above is limited for us. Consequently, we have to look for alternatives. The recent results by Krätschmer et al. (2012*b*) provide a Hampel-theorem for the Wasserstein metric with $p = 1$ for a modified robustness notion, which has been introduced by (Cont et al., 2010, def. 3.1) and Krätschmer et al. (2012*b*, def. 2.1):

Definition 5.1.9 (Qualitative $\mathcal{M}_0$-Robustness)
Let $\mathcal{M}_0 \subset \mathcal{M}$ and $F \in \mathcal{M}$. The sequence (T_n) is called *qualitatively* $\mathcal{M}_0$*-robust* w.r.t. (d, d') if for every $\epsilon > 0$ there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$d(F,G) \leq \delta \Rightarrow d'(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) \leq \epsilon \tag{5.5}$$

for all $G \in \mathcal{M}_0$ and $n \geq n_0$.

To prove the result, Krätschmer et al. (2012*b*) introduce the concept of $\mathcal{N}$-continuity and a uniform Glivenko-Cantelli property.

Definition 5.1.10 ($\mathcal{N}$-Continuity)
Let $F \in \mathcal{M}$ and $\mathcal{N} \subset \mathcal{M}$. A functional T is called $\mathcal{N}$-continuous at F w.r.t. d if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $G \in \mathcal{N}$

$$d(F,G) \leq \delta \Rightarrow |T(F), T(G)| \leq \epsilon. \tag{5.6}$$

Definition 5.1.11 (Uniform Glivenko-Cantelli (UGC) Property)
Let $\mathcal{M}_0 \subset \mathcal{M}$. A metric space $(\mathcal{M}, d)$ has the *UGC poperty* if for every $\epsilon > 0$ and $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$F^{\mathbb{N}}\left[\left\{\omega \in (\Omega)^{\mathbb{N}} \middle| d(F, F_n) \geq \delta\right\}\right] \leq \epsilon$$

for all $F \in \mathcal{M}_0$ and for all $n \geq n_0$.

Theorem 5.1.12 (Hampel's Theorem for M_0-Robustness)
Let $\mathcal{M}_0 \subset \mathcal{M}$ such that $(\mathcal{M}_0, d)$ has the UGC property, and let $F \in \mathcal{M}$. Set $\mathcal{E} = \bigcup_{n \geq 1} \mathcal{M}_n$. If T is $\mathcal{E}$-continuous at F w.r.t. d, then (T_n) is $\mathcal{M}_0$ robust at F w.r.t. (d, d_{Pr}). Conversely, if (T_n) is weakly consistent in some d-neighbourhood of F and (T_n) is qualitatively $\mathcal{M}_0$-robust a F w.r.t. (d, d_{Pr}), then T is $\mathcal{M}_0$-continuous w.r.t. d.

Proof. See Krätschmer et al. (2012*b*, Appendices A1 & A2). □

Proposition 5.1.13
For any $\kappa > 0$ and any $\lambda > 0$, denote by $(\mathcal{M}_{\kappa,\lambda}, d_1)$ the metric space consisting of the probability measures F on $\mathbb{R}$ with

$$\int (1 + |x|)^{\lambda} dF < \kappa$$

and the d_1-Wasserstein distance. $(\mathcal{M}_{\kappa,\lambda}, d_1)$ has the UGC-property. Consequently, if T is an $\mathcal{E}$-continuous functional at F w.r.t. d_1 on this space for some $\kappa > 0$, $\lambda > 0$, then (T_n) is qualitatively $\mathcal{M}_{\kappa,\lambda}$-robust at F w.r.t. (d, d_{Pr}).

Proof. See Krätschmer et al. (2012*b*, thm. 3.1 & Corr. 3.3) for the first part. The second part is a direct consequence of theorem 5.1.12. □

Example 5.1.14 (Qualitative Robustness of the Sample Mean)
Krätschmer et al. (2012*b*, ex. 4.2) show that the sample mean is $\mathcal{M}_{\kappa,\lambda}$-robust w.r.t. (d_1, d_{Proh}) at $F \in \mathcal{M}_{\kappa,\lambda}$ for any $\kappa \geq \int (1 + |x|)^{\lambda} dF$ and $\lambda > 1$.

5.2 Quantitative Robustness

We have seen that a qualitatively robust functional is well behaved for small to medium deviations from the model. Quantitative robustness is concerned with measuring how much a small deviation from the model assumption influences the functional. To make this more precise, we consider in this section a sequence (T_n) that derives from a consistent functional T. A small deviation from the model is characterized by a distribution that is in an ϵ-neighborhood $\mathcal{P}_\epsilon(F_0)$ of the model distribution F_0. The neighborhood depends on the distance of the metric space of probability measures under consideration. Popular choices, see Huber and Ronchetti (2011, sect. 1.4), are the *Lévy neighborhood*

$$P_{\epsilon,L}(F_0) = \{F|\forall t, F_0(t-\epsilon) - \epsilon \leq F \leq F_0(t+\epsilon) + \epsilon\} \tag{5.7}$$

$$= \{F|d_L(F_0, F) < \epsilon\} \tag{5.8}$$

and the *gross error model*

$$P_{\epsilon,GEM}(F_0) = \{F|F = (1-\epsilon)F_0 + \epsilon H, H \in \mathcal{M}\}\,. \tag{5.9}$$

Note that the Lévy neighborhood is a neighborhood in the weak topology sense on $(\mathcal{M}, d_L)$, but the gross error model—or contamination neighborhood—is not. Using a neighborhood such as those defined above, we obtain two measures of quantitative robustness.

Definition 5.2.1 (see Huber and Ronchetti (2011, p. 12))
Let (T_n) be a sequence derived from a statistical functional T, $F_0 \in \mathcal{M}$, and let $P_\epsilon = P_\epsilon(F_0)$ be an ϵ-neighborhood of F_0. Then the *maximum bias* is given by

$$b_1(\epsilon; T) = \sup_{F \in \mathcal{P}_\epsilon} |T(F) - F(F_0)|\,,$$

and the *maximum asymptotic bias* by

$$b(\epsilon) = \lim_{n\to\infty} \sup_{F \in \mathcal{P}_\epsilon} |\text{Median}(\mathcal{L}_F(T_n - T(F_0))|\,.$$

Huber and Ronchetti (2011) show that if b_1 is well-defined, $b(\epsilon) \geq b_1(\epsilon)$ and that equality holds in many cases, for example for the Lévy neighborhood.

Definition 5.2.2 (Asymptotic Breakdown Point)
[see Huber and Ronchetti (2011, p. 13)] The *asymptotic breakdown point* of a functional T at $F_0 \in \mathcal{M}$ is defined as

$$\epsilon^* = \sup\{\epsilon|b(\epsilon) < b(1)\}\,.$$

For the Lévy and contamination neighborhood, $b(1)$ gives the worst possible value of $b(.)$—corresponding to a change to any distribution in $\mathcal{M}$—and is usually infinite. Therefore, the breakdown point can be interpreted as the fraction of outlying observations the estimator can accept before becoming useless: the higher the breakdown point, the more robust the estimator to outliers.

Example 5.2.3

The arithmetic mean and the maximum have an (asymptotic) breakdown point of $\epsilon^* = 0$, since a single large observation can produce arbitrarily large values. The α-trimmed mean and the α-quantile have an (asymptotic) breakdown point of $\epsilon^* = \alpha$.

5.3 Infinitesimal Robustness

In the preceding sections, we have considered qualitative robustness, concerned with "small" and "somewhat larger" deviations from the model, and quantitative robustness, concerned with measuring the results of small deviations from the model for statistical functionals. It remains to ask the question, "What happens if we add one more observation [...] to a very large sample?" (Huber and Ronchetti, 2011, p. 14). The answer is provided by influence curves which are closely related to derivatives of statistical functionals. As differentiability of statistical functionals is helpful when considering statistical methods such as the bootstrap, we start with the general framework for derivatives in this context and obtain the influence curve as a special case.

The discussion of differentiable statistical functionals started with the work by von Mises (1947), who observed that, under some conditions,

$$T(F_n) - T(F) = T'(F_n - F) + R(F_n, F) \tag{5.10}$$

where $T'(F_n - F)$ is a derivative of T, and $R(F_n, F)$ is a remainder term from a Taylor series development of T. The existence and properties of the derivative, the asymptotic behavior of the remainder term, and the distribution of the left-hand side of equation (5.10) have to be investigated. We pursue this agenda in a slightly more general setup, replacing the empirical distribution function F_n by a more general c.d.f, say G.

Let $(\mathcal{M}, d)$ be a metric space of probability measures. A general definition of differentiability can be given assuming that the domain of T extends to the space given by the linear span of the set of probability measures, namely to finite signed measures.

Definition 5.3.1 (see Shao (1993, def. 2.1))

Let $\mathcal{S}$ be a class of subsets of the linear space $\mathcal{D}$ generated by $\mathcal{M}$. Then a functional $T : \mathcal{M} \to \mathbb{R}$ is $\mathcal{S}$-differentiable at $F \in \mathcal{M}$ if there exists a linear functional L_F on $\mathcal{D}$ such that for any $\mathcal{C} \in \mathcal{S}$

$$\lim_{t \to 0} \frac{|T(F + tG) - T(F) - L_F(tG)|}{t} = 0 \tag{5.11}$$

uniformly for all $G \in \mathcal{C}$ with $F + tG \in \mathcal{M}$. If $\mathcal{S}$ is the class of

(i) single-point subsets of $\mathcal{D}$, T is called *Gâteaux differentiable*;

(ii) compact subsets of $\mathcal{D}$, T is called *Hadamard* or *compactly differentiable*;

(iii) bounded subsets of $\mathcal{D}$, T is called *Fréchet differentiable*.

An immediate consequence of definition 5.3.1 is that Fréchet differentiability implies Hadamard differentiability which implies Gâteaux differentiability. It also shows that differentiability of statistical functionals depends on the topology we put on $\mathcal{M}$ and therefore on the distance we use in the metric space $(\mathcal{M}, d)$. Reeds (1976) criticizes the approach of defining differentiation with respect to a metric as "bad policy" (Reeds, 1976, p. 74) because different metrics can generate the same topology. While it is possible to put a norm on the space of probability measures (see Dudley, 1969), we prefer to do without a norm and just consider the metric space $(\mathcal{M}, d)$ with a metric d chosen to fit our applications.

There are equivalent differentiability definitions which do not require us to extend the statistical functional T beyond the class of probability measures. We add the requirement that $(\mathcal{M}, d)$ is a metric space such that d metrizes the weak topology on $\mathcal{M}$ and $d(F_t, F_s) = O(|t - s|)$ for $F_t = (1 - t)F_0 + tF_1$ for some $F_0, F_1 \in \mathcal{M}$. A statistical functional is then Fréchet differentiable at F w.r.t. d if there exists a linear functional $L = L_F$ with domain consisting of the finite signed measures such that

$$|T(G) - T(F) - L(G - F)| = o(d(F, G))$$

for all $G \in \mathcal{M}$. An additional advantage of this definition is that it makes the dependence of Fréchet differentiability on the metric more explicit. Similarly, a statistical functional T is Gâteaux differentiable at F w.r.t. d if for all $G \in \mathcal{M}$ there is a linear functional $L = L_F$

such that

$$\lim_{t \to 0} \frac{T(F_t) - T(F)}{t} = L(G - F)$$

where we set $F_t = (1 - t)F + tG$. We obtain a special case by setting G to the Dirac measure, $G = \delta_x$.

Definition 5.3.2 (see Huber and Ronchetti (2011, p. 39))
The *influence curve (IC)* (or influence function) of a functional T at F is given by

$$\mathrm{IC}(x; F, T) = \lim_{t \to 0} \frac{T(F_t) - T(F)}{t} \tag{5.12}$$

where $F_t = (1 - t)F + t\delta_x$.

The influence curve "describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of the contamination" (Hampel, Ronchetti, Rousseeuw and Stahel, 1986). It provides a possible answer to the question of the influence of a single observation on the value of the functional that we posed at the beginning of this section. From a robustness point of view, a smaller value of the influence curve is, of course, preferable to a larger value. This is captured by the *gross-error sensitivity*, given by (see Hampel et al. (1986, sect. 2.1c))

$$\gamma^* = \gamma^*(T, F) := \sup_{x|\mathrm{IC}(x) \text{ exists}} |\text{IC(x;F,T)}| .$$

Consequently, Rousseeuw (1981) calls a functional with finite gross-error sensitivity *B-robust*. As the influence curve can often be computed even if the functional T is not Gâteux differentiable in general, we can then use IC and γ^* to investigate the infinitesimal aspect of robustness.

Example 5.3.3 (Mean, Trimmed Mean, α-Quantiles)
The influence curve for the mean functional $T(F) = \int x dF(x)$, which is defined for all $F \in \mathcal{M}$ with existing first moment, is given by (see Hampel et al. (1986) for the special case of $F = N(0, \sigma)$)

$$IC_(x; F, T) = x - \mu_F$$

where μ_F denotes the first moment of F. It is unbounded, which implies $\gamma^*(T,F)=\infty$. In particular, the mean is not B-robust.

For the α-trimmed mean, we have (see Huber and Ronchetti (2011, p. 58))

$$IC(x;F,T_\alpha)=\begin{cases}\frac{1}{1-2\alpha}\left(F^{-1}(\alpha)-W_\alpha(F)\right) & \text{for } x<F^{-1}(\alpha)\\ \frac{1}{1-2\alpha}\left(x-W_\alpha(F)\right) & \text{for } F^{-1}(\alpha)<x<F^{-1}(1-\alpha)\\ \frac{1}{1-2\alpha}\left(F^{-1}(1-\alpha)-W_\alpha(F)\right) & \text{for } x>F^{-1}(1-\alpha)\end{cases}$$

where $W(F)$ is the Winsorized mean

$$W_\alpha(F)=(1-2\alpha)T_\alpha(F)+\alpha\left(F^{-1}(\alpha)+F^{-1}(1-\alpha)\right).$$

So the influence curve is bounded in this case and, as a consequence, the α-trimmed is B-robust.

If F has nonzero finite derivative f at $F^{-1}(\alpha)$, the influence curve of the α-quantile is given by (see Huber and Ronchetti (2011, sect. 3.3.1))

$$IC(x;F,T_{q_{(\alpha)}})=\begin{cases}\frac{\alpha-1}{f(F^{-1}(\alpha))} & \text{for } x<F^{-1}(\alpha)\\ \frac{\alpha}{f(F^{-1}(\alpha))} & \text{for } x>F^{-1}(\alpha)\end{cases}$$

which is a bounded function. The α-quantile is B robust.

Example 5.3.4 (L-Statistics)

We consider L-statistics as defined in equation (5.3). There, T_{L2} is a sum of quantiles. We can reuse the results from example 5.3.3 and obtain

$$IC(x;,F,T_{L2})=\sum_{i=1}^{m}a_i\frac{\alpha_i-\mathbf{1}_{\{x<F^{-1}(\alpha_i)\}}(x)}{f(F^{-1}(\alpha))}.$$

$IC(x;,F,T_{L2})$ has m step discontinuiuties. Serfling (1980, sect. 8.1) calculates the influence curve for T_{L1} as

$$IC(x;,F,T_{L1})=-\int_{-\infty}^{\infty}\left(\mathbf{1}_{\{x<y\}}-F(y)\right)J(F(y))dy. \tag{5.13}$$

$IC(x;,F,T_{L1})$ is continuous, but potentially unbounded, unless J vanishes outside some bounded interval. The influence curve of T_L is then given by the sum of $IC(x;,F,T_{L1})$

and $IC(x;,F,T_{L2})$.

Apart from the easy calculation of the influence functions and the weak continuity results from theorem 5.1.4, we can find results on the L-statistics's differentiability properties.

Theorem 5.3.5 (see Shao (1989, thm. 4.1))
Consider T_{L1} as defined in equation (5.3) with $J(t) = 0$ for $t \in [0, a) \cup (b, 1]$ for some $0 < a < b < 1$. If J is bounded and continuous a.e. w.r.t. the Lebesgue measure on $[0, 1]$ and a.e. w.r.t. F^{-1}, then T is Fréchet differentiable with differential given in equation (5.13) and locally Lipschitz continuous at F w.r.t. $\|.\|_\infty$.

Proof. See the proof by Boos (1979, thm. 1). Continuity w.r.t. F^{-1} means that the set $\{x|J \text{ is discontinuous at } F(x)\}$ is a Lebesgue null-set. □

Chapter 6

Risk Measures

As a wealth of risk measures and axiomatic approaches to classify them, both on a probability space or a sample space, have been proposed in the literature, the question arises which risk measure or set of risk measures is appropriate for robust risk management. The discussion of the merits and weaknesses of axiomatic systems and single risk measures has been ongoing since the famous contribution by Artzner, Delbaen, Eber and Heath (1999), who first introduced so-called coherent risk measures. More recently, Heyde, Kou and Peng (2007) suggested that the required properties of risk measures depend on the targeted audience, differentiating between external risk measures intended for regulators and internal risk measures directed at the undertaking's management. In this section, we review some popular classes of risk measures, give examples and analyze their properties with regard to the intended audience.

6.1 Axiomatic Approaches on Probability Spaces

The risk measures we discuss in this section are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathfrak{X}$ denote the set of real-valued random variables on this probability space. Adapting the definition by Artzner et al. (1999) slightly we have:

Definition 6.1.1
A measure of risk is a mapping ρ from (a subset of) $\mathfrak{X}$ into $\mathbb{R}$.

There are different possibilities for the interpretation of the set $\mathfrak{X}$ and $\rho(X), X \in \mathfrak{X}$. In their seminal paper on an axiomatic approach to risk measures, Artzner et al. (1999) interpret each $X \in \mathfrak{X}$ as a final net worth of a position, which is acceptable to the regulator a clearing firm or management if the chosen risk measure evaluated for the position is negative, $\rho(X) < 0$. The risk measure then gives the amount of cash, or, depending on the context, equity, that has to be added to the position to make it acceptable. In the Solvency II context, we think it is reasonably to follow Dhaene, Laeven, Vanduffel, Darkiewicz and Goovaerts (2008) and interpret $\mathfrak{X}$ as a set of losses of an insurance company or portfolio. The company is insolvent when its aggregated loss over a reference period is $X > 0, X \in \mathfrak{X}$. Dhaene et al. (2008) then think of $\rho(X)$ as a capital requirement a regulator imposes on the company to protect its policyholders with a certain probability from the event $X - \rho(X) > 0$ of default. This setup can easily be adapted to different industries such as the banking or energy sectors and also to internal risk management, using the risk measure in the calculation of the economic capital that shareholders of a company would require (for an application in this area, see, e.g., Prokopczuk, Rachev, Schindlmayr and Trück, 2007).

Risk measures can have a number of desirable properties, depending on the interpretation and intended usage. In the literature, sets of axioms have been put together that result in different classes of risk measures that fulfill these axioms. We consider coherent, convex and distortion risk measures in the following. Each of these class can be characterized by a subset of the following axioms, for which we also provide short interpretations:

RM1 Monotonicity: if $X(\omega) \leq Y(\omega) \forall \omega \in \Omega$ then $\rho(X) \leq \rho(Y)$. A portfolio with bigger losses in any state of the world should lead to a higher capital requirement.

RM2 Subadditivity: $\rho(X+Y) \leq \rho(X) + \rho(Y)$. "A merger does not create extra risk" (Artzner et al., 1999, p.209)

RM3 Positive homogeneity: for $\lambda > 0$ we have $\rho(\lambda X) = \lambda \rho(X)$. Increasing or decreasing the amount invested in a portfolio should increase or decrease, respectively, the risk by the same factor.

RM4 Translation invariance: for any constant function a we have $\rho(a + X) = \rho(X) + a$. Adding a sure loss to a portfolio should increase the capital requirement by the amount of the sure loss.

RM5 Convexity: for any $\lambda \in [0, 1]$ we have $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda)\rho(Y)$. "Convexity means that diversification does not increase risk" (Föllmer and Schied, 2002*a*). Note that convexity is implied by RM2 and RM3, but it does not imply RM2 and RM3. It is therefore a weaker axiom than RM2 and RM3 together.

RM6 Law invariance/conditional state independence: For a given market condition, $\rho(X)$ depends only on the distribution of X.

RM7 Comonotonic additivity: for any comonotonic[1] r.v.s X, Y, we have $\rho(X) + \rho(Y) = \rho(X + Y)$. Merging two portfolios that cannot be used as hedges for each other should not change the risk position.

RM8 Continuity: for any X, $d \geq 0$, we have

$$\lim_{d \to 0+} \rho(X - d)_+ = \rho(X)$$

and

$$\lim_{d \to \infty} \rho(\min(X, d)) = \rho(X).$$

Small truncation of the random loss should only lead to small changes of the risk measure, which can still be calculated if the loss X is approximated by bounded random variables (see Wang, Young and Panjer, 1997).

From a robustness point of view, axioms RM1 - RM8 do not have much to offer. Monotonicity (RM1) can be seen as a minimum requirement if a risk measure is to be used as a basis for consistent decisions. The continuity requirement (RM8) is a step in the right direction, but seems not to be sufficient: a change in the losses of a portfolio or an insurance undertaking will usually be stochastic—that is, occur by adding or removing a risky asset or an insurance contract—but RM8 considers only deterministic changes.

Definition 6.1.2 (Monetary Risk Measures)
A mapping[2] $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \cup \{\infty\}$ is called a *monetary risk measure* if it satisfies axioms RM1 and RM4 (see Föllmer and Schied, 2011, def. 4.1).

Example 6.1.3 (Standard Deviation)
The definition of a risk measure encompasses a wealth of functions, among the measures of

[1] Two random variables X, Y are comonotonic if $X(\omega_1) < X(\omega_2) \Leftrightarrow Y(\omega_1) < Y(\omega_2)$ for any $\omega_1, \omega_2 \in \Omega$, see Denneberg (1994, Prop.4.5). Note that comonotonicity is a distribution free property.

[2] We mostly use the classical definition of risk measures as maps on L^∞ in this chapter. The definitions can be translated to L^p spaces, see Cheridito and Li (2009).

dispersion like the standard deviation. The standard deviation of asset returns—it is often called volatility in this context—has been used, for example, to measure the riskiness of assets in Markowitz portfolio theory (Markowitz, 1952), but could also be applied to loss distributions, as long as the second moment exists. It obviously satisfies axioms RM1 and RM7, but it is not even a monetary risk measure. Consequently, we look for alternatives.

Example 6.1.4 (Value at Risk)
The value at risk (VaR) at level α of a loss $X \in \mathfrak{X}$ is defined as

$$\mathrm{VaR}_\alpha(X) = \inf\left\{x | \mathbb{P}(X \leq x) \geq \alpha\right\} \tag{6.1}$$

where we take the definition from Dhaene et al. (2008). Note that the α-VaR equals by definition the lower α-quantile $q_{(\alpha)}(X)$ of the loss X. Additionally, VaR is a monetary risk measure.

Remark 6.1.5
In our setting, interpreting a positive value of the random variable X as a loss defining the α-VaR by the lower α-quantile, we follow Heyde et al. (2007), Dhaene et al. (2008), and many others and adopt the convention to talk about a 95%- or 99%-VaR rather than a 5%- or 1%-VaR as in other works on the topic.

VaR was helped to widespread use by practitioners at J.P. Morgan (Hull, 2011, Business Snapshot 8.1), and is now the most widely used risk measure (Denuit et al., 2006, sect. 2.3). Among the reasons is that the computation of VaR as a quantile is relatively easy. Apart from that, it seems to be easy to understand for non-experts, as it is stated in units of money and can be loosely defined saying that "VAR summarizes the worst loss over a target horizon that will not be exceeded with a given level of confidence"(Jorion, 2009, p. viii). While this intuition is restated in many works covering the topic (see, among many others, Hull, 2011; Crouhy et al., 2001), critics argue that it gives non-expert users a false sense of security. Rowe (2010) suggests to call the 99% VaR a "minimum twice-a-year loss" instead.

We show in section 3 that the VaR is the risk measure of choice in the Solvency II framework. It also plays an important role in internal models in the Basel II regulations for banks (The European Parliament and the Council of the European Union, 2006*a*,*b*, Annex VII and Annex V, respectively). Dhaene et al. (2008) provide a theoretical justification for the application of VaR as an external risk measure in this area. They consider a measure

φ of shortfall or insolvency risk

$$\varphi[(X - \rho(X))_+]$$

where

$$\rho_1(X) \leq \rho_2(X) \Rightarrow \varphi((X - \rho_1(X))^+) \geq \varphi((X - \rho_2(X))^+)$$

holds and with an associated cost function

$$C(X, \rho(X)) = \varphi[(X - \rho(X))_+] + \rho(X)\epsilon \tag{6.2}$$

where $\epsilon \in (0, 1)$ determines how much capital costs enter into the cost function.

Proposition 6.1.6

The value at risk at level $1 - \epsilon$ minimizes equation (6.2).

Proof. See Dhaene et al. (2008, thm. 1). □

In this sense, value at risk is the most efficient capital requirement. Dhaene et al. (2008) also show that it satisfies their "regulator's condition" which states that for a risk measure used as a capital requirement, diversification should not increase shortfall risk as defined above. We revisit the application of VaR as a capital requirement when we discuss natural risk statistics in section 6.2.

But value at risk also has critics. It is clear that VaR at level α does not provide information on losses beyond this level (Dowd and Blake, 2006). This fact can be used to game VaR measurements if no precautions are taken. An example for a portfolio with low VaR and apparently high risk is a position in short out-of-the-money put option, which pay an up-front premium and only lead to payments in "bad" scenarios (Dowd and Blake, 2006).

As one of the main points of criticism of VaR in the academic literature concerns the axioms that it (does not) satisfy, we postpone further discussion of the properties of VaR until we consider the work by Heyde et al. (2007) in section 6.2 to focus on the axioms for now.

Proposition 6.1.7

The value at risk satisfies axioms RM1, RM3, RM4 and RM6. Consequently it is a monetary risk measure.

Proof. The properties are immediate consequences of the definition. □

In particular, VaR does not satisfy subadditivity RM2, and has been widely criticized for this fact—for example, by Artzner et al. (1999) and Dowd and Blake (2006), who also provide counter-examples. Daníelsson, Jørgensen, Mandira, Samorodnitsky and de Vries (2005) and Daníelsson, Jørgensen, Mandira, Samorodnitsky and de Vries (2011) show, however, that VaR is subadditive in many situations encountered in practice.

Whether subadditivity is in fact a desirable property for a risk measure is subject to debate in the literature. Föllmer and Schied (2002*a*) observed that subadditive risk measures ignore liquidity risk if portfolios with the same loss characteristics are merged (Already Artzner et al. (1999) noted that the same is true for positive homogeneous risk measures). Note that scaling a position does not mean that the associated loss random variable X is scaled by the same factor if liquidity risk is involved. Loebnitz and Roorda (2011) suggest a solution to this problem and also review different approaches in the literature.

But this is not the only criticism of subadditivity. Heyde et al. (2007) have compiled arguments against subadditive risk measures as external risk measures. They recall an observation by Dhaene, Vanduffel, Tang, Goovaerts, Kaas and Vyncke (2004) and Dhaene et al. (2008) that a merger can in fact increase (shortfall) risk because limited liability applies only to the whole company after a merger instead of to the separate parts before the merger. Additionally, they find evidence from prospect theory, developed by Kahneman and Tversky (1979), and utility theory that risk perception violates subadditivity in some cases.

6.1.1 Coherent Risk Measures

Monetary risk measures only satisfy the most basic axioms. We now move to risk measures characterized by larger sets of axioms, starting with the coherent risk measures introduced by Artzner et al. (1999).

Definition 6.1.8 (Artzner et al. (1999); Delbaen (2002))
A mapping $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \cup \{\infty\}$ is called a *coherent risk measure* if it satisfies axioms RM1-RM4.

Axioms RM2 and RM3 are central to coherent risk measures. Despite the criticism of those axioms we discuss above, coherent risk measures are used widely in the literature and as internal risk measures. Delbaen (2002) generalizes them to general probability spaces and relates them to distorted measures. Jaschke and Küchler (2001) discuss coherent risk measures in the context of arbitrage bounds for derivatives pricing. An interesting property in our context is the following:

Proposition 6.1.9

A coherent risk measure $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ can be written as

$$\rho(X) = \sup_{P \in \mathcal{P}} E_P[X] \; \forall X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}).$$

where $\mathcal{P}$ is a set of probability measures equivalent to $\mathbb{P}$ on $(\Omega, \mathcal{F})$ if and only if it fulfills the Fatou property

$$\rho(X) \leq \liminf_{n \to \infty} \rho(X_n)$$

for any bounded sequence $X_n \to X a.s.$.

Proof. See Föllmer and Schied (2011, Corollary 4.37). □

Following Artzner et al. (1999) and interpreting the measures $P \in \mathcal{P}$ as *generalized scenarios* the proposition justifies the use of scenario analysis in the context of coherent risk measures (see Heyde et al., 2007, sect. 2.1.1).

Remark 6.1.10

As has been noted by academics and practitioners alike (see, e.g., Dhaene et al., 2008; Holton, 2009), the name "coherent" risk measure can be deceiving. The dictionary lists "reasonable" and "valid" as synonyms of "coherent"[3]. Therefore, the name "coherent risk measure" can lead to the false conclusion that any risk measure that is not coherent is not reasonable and not valid.

Example 6.1.11 (Expected Shortfall)

The Expected Shortfall (ES) at level α of a loss X is given by

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha} \left(\mathbf{E}\left[X \mathbf{1}_{(X \geq q^\alpha(X))} \right] + q^\alpha(X) \left(1 - \alpha - \mathbb{P}\left[X \geq q^\alpha(X)\right]\right) \right) \tag{6.3}$$

[3]Merriam-Webster OnLine, s.v. "coherent," accessed June 7, 2012, `http://www.merriam-webster.com/dictionary/coherent`

where

$$q^{\alpha}(X) = \inf\{x|\mathbb{P}(X \leq x) > \alpha\} \tag{6.4}$$

is the upper α-quantile of the distribution of X (we adapt the definition by Acerbi and Tasche (2002)). With this definition, Acerbi and Tasche (2002) also show that ES is equal to the Conditional value at risk (CVaR). Other related risk measures are the tail value at risk (TVaR) and the tail conditional expectation (TCE). We focus on the ES in this example because Acerbi and Tasche (2002) provide a consistent reference[4].

Proposition 6.1.12
The expected shortfall of $X \in L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ can be represented as

$$ES_{\alpha}(X) = \frac{1}{1-\alpha}\int_{1-\alpha}^{1} VaR_u(X)du.$$

Proof. See the remark after proposition 3.2 in (Acerbi and Tasche, 2002). □

Proposition 6.1.13 (Expected Shortfall is a coherent risk measure)
The Expected Shortfall satisfies axioms RM1-RM4 and is therefore a coherent risk measure.

Proof. See Acerbi and Tasche (2002, Prop. 3.1). □

The coherence property distinguishes ES from VaR and it is the main point why its proponents prefer it to the latter, see Acerbi, Nordio and Sirtori (2001).

Another property that distinguishes the ES from the VaR is that it takes into account information on the losses beyond the level α, which are disregarded by VaR. The usefulness of this property in applications is, in our view, doubtful, due to the lack of robustness of ES. As a mean type statistic, it is sensitive to outliers in the data, as indicated by an unbounded influence curve, which Heyde et al. (2007) show to be given by

$$IC(x, \mathrm{ES}_{\alpha}, X) = \begin{cases} \mathrm{VaR}_{\alpha}(X) - \mathbf{E}\left[X|X \geq \mathrm{VaR}_{\alpha}(X)\right], & \text{if } x \leq \mathrm{VaR}_{\alpha}(X), \\ \frac{x}{1-\alpha} - \mathbf{E}\left[X|X \geq \mathrm{VaR}_{\alpha}(X)\right] - \frac{\alpha}{1-\alpha}\mathrm{VaR}_{\alpha}(X), & \text{if } x > \mathrm{VaR}_{\alpha}(X) \end{cases}$$

[4]We cannot take this for granted in the literature on risk measures. As Acerbi and Tasche (2002) note about TCE: "Meanwhile, several authors [...] proposed modifications to TCE, this way increasing confusion[...]". The same is true for other risk measures.

if X has continuous density which is positive at $\text{VaR}_\alpha(X)$. Additionally, the extra information on the losses beyond VaR might not be in the data. To show this, we extend an example by Heyde and Kou (2004). They consider the problem of distinguishing between a normalized Laplace distribution with exponential-type tails with density

$$f_{LP}(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, x \in \mathbb{R}$$

and a normalized t-distribution with power-type tails and density

$$f_{t_\nu}(x) = \frac{\Gamma((\nu + 1/2))}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}} \cdot \left(1 + \frac{x^2}{v-2}\right)^{\frac{\nu+1}{2}}, x \in \mathbb{R}$$

based on a sample. A sample size of 5000 observations, corresponding to about 20 years of daily data in a financial time series, is not enough to distinguish between Laplace distribution and t-distribution with 5.4899 degrees of freedom based on the tails of the distributions at 95% confidence, because the asymptotic confidence intervals of the sample quantiles overlap (see Walker, 1968). While the difference in 99.9% quantiles, and therefore VaR, between the two distributions is less than 0.001%, the difference in 99.9% ES in contrast is about 8%, see table 6.1.

	VaR 99.9%	ES 99.9%
Laplace distribution	-4.3944	-5.1015
Normalized t-distribution (5.4899 dg. of f.)	-4.3944	-5.4988

Table 6.1: VaR and ES of the Laplace distribution and the normalized t-distribution with 5.4899 degrees of freedom

6.1.2 Convex Risk Measures

Definition 6.1.14 (Föllmer and Schied (2002*a*))
A mapping $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \cup \{\infty\}$ is called a convex risk measure if it adheres to axioms RM1, RM4 and RM5.

As noted above, convexity (RM5) is a weaker requirement than positive homogeneity (RM3) combined with subadditivity (RM2). Therefore, all coherent risk measures are also convex risk measures.

The main attraction of convex risk measures is already implied by their name: The convexity allows the use of results from convex optimization theory in applications such as portfolio allocation. A lot of work has been published in the area, among others by Filipović and Svindland (2008), who proof the existence of solutions for a wide range of problems, and by Lüthi and Doege (2005), who focus on efficient numerical computations. The criticism of the coherent risk measures concerning axioms RM2 and RM3 does not apply to convex risk measures.

Just as for coherent risk measures, there is a dual representation for convex risk measures that links them to scenario analysis.

Proposition 6.1.15

A convex risk measure $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ can be written as

$$\rho(X) = \sup_{P \in \mathcal{P}} \left(E_P[X] - \alpha^{min}(P) \right) \ \forall X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}).$$

where $\mathcal{P}$ is the set of all probability measures equivalent to $\mathbb{P}$ on $(\Omega, \mathcal{F})$ and $\alpha^{min}(P)$ is the minimal penalty function (Föllmer and Schied, 2011, sect. 4.2) if and only if it fulfills the Fatou property

$$\rho(X) \leq \liminf_{n \to \infty} \rho(X_n)$$

for any bounded sequence $X_n \to X a.s.$.

Proof. See Föllmer and Schied (2011, thm. 4.33). □

We consider an example of a risk measure that is convex, but not coherent.

Example 6.1.16 (Entropic Risk Measure)

The expected exponential loss is a special case of entropy-based risk measures, which we obtain by setting the penalty function to the entropy function (see Lüthi and Doege, 2005)

$$\overline{\alpha}(P) = \begin{cases} \frac{P(\omega)}{\mathbb{P}(\omega)} \log \left(\frac{\mathbb{P}(\omega)}{P(\omega)} \right), \text{ if } P \ll Q, \\ +\infty, \text{ otherwise.} \end{cases}$$

This gives us the expected exponential loss

$$\rho_1^E(X) = \log \mathbf{E}_P[e^X],$$

while using $\gamma\overline{\alpha}(P)$ for some $\gamma > 0$ as the penalty function gives the entropic risk measure (see Barrieu and Karoui, 2009)

$$\rho_\gamma^E(X) = \gamma \log \mathbf{E}_P \left[\exp\left(\frac{X}{\gamma}\right)\right].$$

Föllmer and Schied (2011, chapter 4.2) consider risk measures from the class of bounded functions $\mathcal{X}$ on a measurable spaces $(\Omega, \mathcal{F})$. They see this as a situation with Knightian uncertainty, as no reference probability measure is fixed in advance. In this setting, they show the following representation for convex risk measures analogously to proposition 6.1.15 that they call "robust representation".

Proposition 6.1.17 (Robust Representation of Convex Risk Measures)
A convex risk measure ρ on the space $\mathcal{X}$ of bounded measurable functions on some measurable space $(\Omega, \mathcal{F})$ is of the form

$$\rho(X) = \max_{Q\in\mathcal{Q}}(\mathbf{E}_Q[X] - \alpha^{\min}(Q)), X \in \mathcal{X}, \mathcal{Q} \subset \mathcal{M}_{1,f} \tag{6.5}$$

where $\alpha^{\min}$ is the minimal penalty function and $\mathcal{M}_{1,f}$ is the set of finitely additive set functions $Q : \mathcal{F} \to [0, 1]$ with $Q(\Omega) = 1$.

The reason they give and that is reiterated by Föllmer (2012) is that the risk measure does not depend on a model, but equals an expectation under that model that results in the highest risk measurement, therefore greatly reducing model risk in theory.

We consider two examples that show that this kind of robustness unfortunately does not translate directly to improvements in practical applications.

Example 6.1.18 (Worst-Case Risk Measure)
The worst-case risk measure on $(\Omega, \mathcal{F})$ is defined as (see Föllmer and Schied, 2002*b*, ex.3.4)

$$\rho_{\max}(X) := \sup_{\omega\in\Omega} X(\omega) \tag{6.6}$$

can be represented as

$$\rho_{\max}(X) = \sup_{Q\in\mathcal{M}_1} \mathbf{E}_Q[X] \tag{6.7}$$

where $\mathcal{M}_1 = \mathcal{M}_1(\Omega, \mathcal{F})$ denotes the space of probability measures on $(\Omega, \mathcal{F})$. Note that the supremum is not attained in general because we restrict the set of finitely additive set functions $\mathcal{M}_{1,f}$ to the set $\mathcal{M}_1$ of probability measures, which can be strictly smaller (see Föllmer and Schied (2011, ex. A.53)). Here, $\mathcal{M}_1$ contains the Dirac measure for each $\omega \in \Omega$, so the supremum in equation (6.7) is attained if the supremum in equation (6.6) is attained, see Föllmer and Schied (2002*b*, ex. 3.4). Note that on L^∞, the worst-case risk measure is defined as

$$\rho_{max}(X) := \operatorname{ess\,sup} X.$$

The worst-case risk measure is the most conservative coherent (or convex) risk measure we can come up with (see Artzner et al., 1999) and there is, in the setup above, no model risk as we always chose the most conservative model. The first problem is that we often have unbounded loss profiles, in which case the measure is $\rho_{\max} = \infty$. If we have bounded losses and are therefore in the setup above, but have to estimate the supremum from data, we have seen that the corresponding estimator is not statistically robust at all. The measure is only useful if the losses are bounded with a known bound, but then model risk is a far less serious problem anyway.

Example 6.1.19 (Expected Shortfall Revisited)
We revisit the expected shortfall introduced in example 6.1.11. As a risk measure on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, it has the robust representation

$$ES_\lambda(X) = \sup_{Q \in \mathcal{Q}_\lambda} E_Q[X] \tag{6.8}$$

where

$$\mathcal{Q}_\lambda = \left\{ Q \sim \mathbb{P} \middle| \frac{dQ}{d\mathbb{P}} < \frac{1}{\lambda} \right\}.$$

Compared to the worst-case risk measure, the set of measures over which the supremum is taken, is smaller. It is restricted to probability measures equivalent to a reference measure $\mathbb{P}$ with bounded Radon-Nikodym density with respect to this measure. In fact, the definition given in equation (6.3) uses only this reference measure. We have to ask why the existence of representation equation (6.8) should have positive implications for robustness, if the reference measure alone is what matters. Additionally, the expected shortfall can be seen as a mean-type statistic when we are concerned with estimation, and so its standard estimators are not robust.

Very recently, Pflug, Pichler and Wozabal (2012) and Wozabal (2012) have taken a different approach to improve robustness of convex risk measures in a portfolio allocation context. They start with a reference probability measure on a fixed probability space which they see as a best estimate of the unknown true distribution of the loss and decide on a convex risk measure to use. In order to obtain a robust version of this risk measure, they calculate the "worst" value the risk measure can take for a loss with distribution in a neighbourhood around the reference distributions, determined via the Wasserstein metric. More precisely, we have the following proposition.

Proposition 6.1.20

Let ρ be a convex, law-invariant risk measure, $\rho : L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}, 1 \leq p < \infty$ with dual representation (see Appendix C)

$$\rho(X) = \max \{\mathbf{E}[XZ] - R(Z) | Z \in L^q\} \tag{6.9}$$

where p and q are Hölder conjugates, $\frac{1}{p} + \frac{1}{q} = 1$, and $R : L^q(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ can be the convex conjugate of ρ, (a representation where R is a convex function is sufficient here, see Pflug and Römisch (2007)). Let X^Q denote a random variable with image measure Q on $\mathbb{R}$. Additionally, let $\partial\rho(X) \subset L^q$ denote the set of maximizers of equation (6.9). Then

- for arbitrary measures Q_1 and Q_2 on $\mathbb{R}$ we have

$$\left|\rho(X^{Q_1}) - \rho(X^{Q_2})\right| \leq \sup_{Z, R(Z) < \infty} d_p(Q_1, Q_2)$$

  where d_p denotes the Wasserstein distance;

- for a probability measure Q_1 on $\mathbb{R}$, if

$$\|Z\|_{L^q} = C \; \forall \; Z \in \bigcup_{X \in L^P} \partial\rho(X) \qquad \text{with } R(Z) < \infty,$$

  it holds for any $\kappa > 0$ that there exists a measure Q_2 on $\mathbb{R}$ with $d_p(Q_1, Q_2) = \kappa$ such that

$$\left|\rho\left(X^{Q_1}\right) - \rho(X^{Q_2})\right| = C\kappa;$$

- for $\rho : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ with

$$\|Z\|_{L^\infty} = C \qquad \text{and } |Z| = C \qquad \text{or } |Z| = 0 \; a.e. \tag{6.10}$$

  for all subgradients Z of ρ. Fix a probability measure Q_1 on $\mathbb{R}$ and $\kappa > 0$. Then there exists a probability measure Q_2 on $\mathbb{R}$ with $d_1(Q_1, Q_2) = \kappa$ such that

$$\left|\rho(X^{Q_1}) - \rho(X^{Q_2})\right| = C\kappa.$$

Proof. See the proofs by Pflug et al. (2012, Lemma 1, Prop. 1, Prop. 2). □

Pflug et al. (2012) show that the requirements of the proposition, especially equation (6.10), hold for ES. We easily[5] get from equation (6.8) that

$$ES_\alpha(X) = \sup\left\{\mathbf{E}[XZ]\middle|\mathbf{E}[Z] = 1, 0 \leq Z \leq \frac{1}{1-\alpha}\right\}. \tag{6.11}$$

Following Pflug et al. (2012) and choosing $A \subset \Omega$ such that $\mathbb{P}(A) = 1 - \alpha$ and $X(\omega) \geq F_X^{-1}(\alpha)$ $\forall \omega \in A$, we have

$$Z(\omega) = \frac{1}{1-\alpha}\mathbf{1}_A(\omega) \in \partial ES_\alpha(X).$$

The fact that Z is a maximizer of equation (6.11) is shown by Föllmer and Schied (2011, thm. 4.52). So for any loss X^Q with Wasserstein distance $d_1(X^{\mathbb{P}}, X^Q) = \kappa$ from $X^{\mathbb{P}}$, the difference in ES is

$$\left|ES_\alpha(X^{\mathbb{P}}) - ES_\alpha(X^Q)\right| \leq \frac{\kappa}{1-\alpha}$$

and there is a such random variable for which equality holds. The robustified version of the ES is therefore

$$ES_\alpha^\kappa(X) = ES_\alpha(X) + \frac{\kappa}{1-\alpha}, \tag{6.12}$$

see Wozabal (2012), who also provide robustified versions of other risk measures, especially from a subclass of distortion risk measures.

[5]For the required measure theoretical facts, see Shreve (2004, thm 1.6.1 & def. 1.6.5).

6.1.3 Distortion Risk Measures

We take a look at one more class of risk measures, the distortion or Wang risk measures.

Definition 6.1.21 (see Wang et al. (1997))
A mapping $\rho : \{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \mid X \geq 0\} \to \mathbb{R} \cup \{\infty\}$ is called a distortion risk measure if it adheres to axioms RM1, RM6, RM7 and RM8

The axiomatic approach originates from the characterization of insurance prices by Wang et al. (1997). The representation via distorted expectations that gives them the name is based on ideas by Yaari (1987), who developed a modified expected utility theory. The conceptual difference to classical expected utility theory is that "distortion functions modify the probability, and keep the wealth function unchanged, whereas utility functions modify the wealth and keep the probability unchanged" (Wirch and Hardy, 1999, sect. 2.1).

Proposition 6.1.22 (Choquet Integral Representation of Distortion Risk Measures)
$\rho(X)$ is a distortion risk measure with $\rho(1) = 1$ if and only if ρ has the Choquet integral representation

$$\rho(X) = \int X d(g \circ \mathbb{P}) = \int_0^\infty g(1 - F_X(x))dx$$

where $g : [0,1] \to [0,1]$ is an increasing distortion function with $g(0) = 0$ and $g(1) = 1$ and F_X is the cdf of X.

Proof. See Wang et al. (1997, thm 3) who refer to Wang (1996) and Denneberg (1994) for details. □

An attractive property of distortion risk measures is that properties of the distortion function g, which is a relatively simple object as a function from the real numbers to the real numbers, translate to properties of the associated risk measure.

Proposition 6.1.23 (Properties of Distortion Risk Measures)
A distortion risk measure ρ with distortion function g as in proposition 6.1.22 is no-ripoff, that is, $\rho(X) \leq \max(X)$, positively homogeneous (RM3) and translation invariant (RM4). ρ preserves first order stochastic dominance. Additionally,

(i) if and only if g is concave, $\rho(X)$ is subadditive and therefore coherent;

(ii) if g is concave, ρ preserves second order stochastic dominance;

(iii) if g is strictly concave, ρ strictly preserves second order stochastic dominance.

Proof. For the proof and further references, see Wang (1996), Wirch and Hardy (2003) and Sereda, Bronshtein, Rachev, Fabozzi, Sun and Stoyanov (2010). □

Applications of distortion risk measures in the literature include portfolio optimization (Sereda et al., 2010; Balbás, Garrido and Mayoral, 2009) and capital allocation (Tsanakas, 2004).

Example 6.1.24 (Distortion Risk Measures)
We consider some examples. Plots of the distortion functions are exhibited in Figure 6.1.

- The VaR_α is a distortion risk measure with distortion function

$$g_\alpha(1 - F_X(x)) = \begin{cases} 0; x > \mathrm{VaR}_\alpha \Leftrightarrow 0 \leq 1 - F_X(x) < 1 - \alpha, \\ 1; x \geq \mathrm{VaR}_\alpha \Leftrightarrow 1 - \alpha \leq 1 - F_X(x). \end{cases}$$

  As the step function is not concave, we can see immediately that the value at risk is not a coherent risk measure.

- Using the representation in proposition 6.1.12, the expected shortfall at level α, ES_α, can be represented via the distortion function

$$g_\alpha(x) = \min\left\{\frac{x}{1-\alpha}, 1\right\}.$$

  The distortion function is concave, so we can confirm that the ES is coherent.

In analogy to convex risk measures, there is a robustified version for certain distortion risk measures developed by Wozabal (2012, ex. 4): If a distortion risk measure ρ_H : $L^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}, 1 < p < \infty$ has a representation as

$$\rho_H(X) = \int_0^1 F_X^{-1}(p) dH(p)$$

for a convex function $H : [0,1] \to \mathbb{R}$ with $H(p) = \int_0^p H(t)dt$, it can be written as

$$\rho_H(X) = \sup\left\{\mathbf{E}[XZ] | Z = h(U), U \sim U[0,1]\right\}.$$

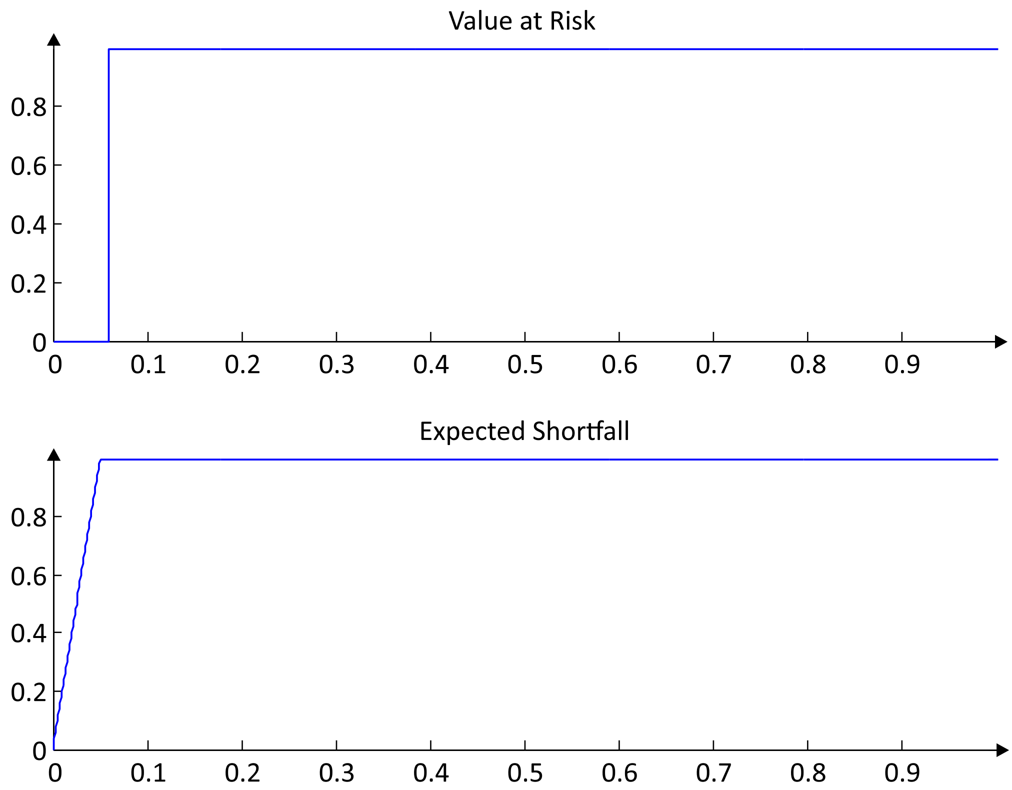

Figure 6.1: Distortion functions for VaR (upper panel) and ES (lower panel)

The robustified version of this risk measure, corresponding to equation (6.12) for the ES, is then given by

$$\rho_H^\kappa(X) = \rho_H(X) + \kappa \, \|h(U)\|_{L^q} \tag{6.13}$$

where $\kappa > 0$ and q is the Hölder conjugate of p.

6.1.4 Risk Measures as Statistical Functionals and Risk Estimators

In order to apply the notions of qualitative robustness we exhibit in chapter 5, we have to consider risk measures as statistical functional on a space of probability distributions.

Remark 6.1.25 (Statistical Functional and Risk Statistics)
Let F denote the cdf of a random variable $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

- The statistical functional for standard deviation is

$$T_{SD}(F) = \sqrt{\int x^2 dF - \left(\int x dF\right)^2}. \tag{6.14}$$

- The statistical functional for the VaR is given by

$$\mathrm{VaR}_\alpha(X) = T_{\mathrm{VaR}}(F) = F^{-1}(\alpha) = \inf\{x | F(x) \geq \alpha\} \tag{6.15}$$

for $0 < \alpha < 1$. It can be written as

$$T_{\mathrm{VaR}}(F) = \int_0^1 F^{-1}(\alpha) d\delta_\alpha$$

where δ_α puts mass one to $\alpha \in (0,1)$.

- The statistical functional for the ES is given by

$$ES_\alpha = T_{ES}(F) = \mathbb{E}(X \mid X > VaR_\alpha) = (1-\alpha)^{-1} \int_\alpha^1 F^{-1}(x) dx \tag{6.16}$$

as shown in proposition 6.1.12.

Example 6.1.26 (Robustness of VaR and ES)
The statistical functional of the VaR is the same as for the quantile, so we can collect the robustness properties from the examples in chapter 5. The VaR at level $\alpha \in (0, 0.5)$ is weakly continuous at $F \in \mathcal{M}$ if F^{-1} is continuous at $1-\alpha$. At such a distribution F, it is also qualitatively robust if we use the Lévy and Prohorov metrics (see theorem 5.1.7). It has an asymptotic breakdown point of α and a finite influence curve given by

$$IC(x; F, T_{q(\alpha)}) = \begin{cases} \frac{\alpha-1}{f(F^{-1}(\alpha))} & \text{for } x < F^{-1}(\alpha), \\ \frac{\alpha}{f(F^{-1}(\alpha))} & \text{for } x > F^{-1}(\alpha). \end{cases}$$

The robustness picture for the expected shortfall looks different. While we have continuity with respect to the Wasserstein distance with $p \geq 1$, the ES is not weakly continuous w.r.t. other metrics that do not require the convergence of moments. Additionally, it has an asymptotic breakdown point of 0 and an unbounded influence curve, which we can derive from the result for the trimmed mean, see example 5.3.3:

$$IC(x; F, T_\alpha) = \begin{cases} \frac{1}{1-\alpha}\left(x - W_\alpha(F)\right) & \text{for } x > F^{-1}(1-\alpha), \\ \frac{1}{1-\alpha}\left(F^{-1}(\alpha) - W_\alpha(F)\right) & \text{for } x \leq F^{-1}(\alpha). \end{cases}$$

However, the sample mean is $\mathcal{M}_{\kappa,\lambda}$-robust w.r.t. (d_1, d_{Proh}), see example 5.1.14.

In practice, we usually do not have a fixed distribution function F to which we can apply the statistical functionals, but we have to estimate the distribution from data. We then use a risk measure to obtain an estimator for the risk. Cont et al. (2010) formalize this procedure in two steps:

Definition 6.1.27 (Risk Measurement Procedure and Risk Estimator, (see Cont et al., 2010, def. 2.3))
A risk measurement procedure is a couple (M, ρ) where

$$\rho : \mathcal{M}^p \to \mathbb{R}$$

is a risk measure and

$$M : \mathcal{X} = \bigcup_{n \geq 1} \mathbb{R}^n \to \mathcal{M}^p,$$

a map from the set of possible data sets, is an estimator of the loss distribution. The composite function

$$\hat{\rho} : \mathcal{X} \to \mathbb{R}, x \mapsto \rho(M(x))$$

is called a risk estimator.

Cont et al. (2010) give examples of risk estimators, in particular the historical risk estimators—obtained by using $M = \hat{F}_n$, the empirical cdf—and maximum likelihood estimators. Using the classical results on robust statistics we collected in chapter 5, they show that the robustness properties for value at risk and expected shortfall that we derived in example 6.1.26 carry over to the historical risk estimators (also see examples 6.2.4 and 6.2.5).

6.2 Data-based Approach: Natural Risk Statistics

The classes of risk measures we consider in sections 6.1.1-6.1.3 have in common that they are usually defined as real-valued maps from some set of random variables on a probability space. The risk statistics introduced by Heyde et al. (2007) that we discuss here take a different approach. As risk measurements are, in practice, derived from data, usually either directly or via a model and simulation, risk statistics take this into account. Consequently, they are defined for data $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Definition 6.2.1

A *risk statistic* is a mapping $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$.

The following axioms can be used to characterize desirable properties of risk statistics, analogously to those for risk measure in section 6.1.

RS1 Monotonicity: $\hat{\rho}(x) \leq \hat{\rho}(y)$ if $x \leq y$, that is, $x_i \leq y_i, i = 1, \dots, n$

RS2 Subadditivity: $\hat{\rho}(x + y) \leq \hat{\rho}(x) + \hat{\rho}(y) x, y \in \mathbb{R}^n$

RS3 Positive homogeneity: $\hat{\rho}(\lambda x) = \lambda \hat{\rho}(x)$ for $\lambda \geq 0$, $x \in \mathbb{R}^n$

RS4 Translation invariance: $\hat{\rho}(x + b\mathbf{1}) = \hat{\rho}(x) + b$ for $b \in \mathbb{R}$ and $x \in \mathbb{R}^n$

RS5 Comonotonic subadditivity: $\hat{\rho}(x+y) \leq \hat{\rho}(x) + \hat{\rho}(y)$ for any comonotonic $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. x and y are comonotonic if $(x_i - y_i)(x_j - y_j) \geq 0$ for any $i \neq j$

RS6 Permutation invariance: $\hat{\rho}((x_1, \dots, x_n)) = \hat{\rho}((x_{\pi(1)}, \dots, x_{\pi(n)}))$ for any permutation $(\pi(1), \dots, \pi(n))$

RS7 Comonotonic additivity: $\hat{\rho}(x + y) = \hat{\rho}(x) + \hat{\rho}(y)$ for any comonotonic $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$

RS8 Scale normalization: $\hat{\rho}(\mathbf{1}) = 1$

Definition 6.2.2 (see Heyde et al. (2007, def. 1,2,3))

A risk statistic that satisfies axioms RS1, RS3, RS4, RS5, and RS6 is called a *natural risk statistic.*

A risk statistic that satisfies axioms RS1, RS2, RS4, and RS3 is called a *coherent risk statistic.* If, in addition, it satisfies RS6, it is called a law-invariant *coherent risk statistic.*

A risk statistic that satisfies axioms RS1, RS6, RS7, and RS8 is called an *insurance risk statistic.*

While the coherent and insurance risk statistics are data-based counterparts of coherent and insurance risk measures, there is no analog for the natural risk statistics. The introduction of the natural risk statistics by Heyde et al. (2007) is motivated by their criticism of subadditivity—see the discussion of the axioms in section 6.1—combined with the robustness requirements they derive for external risk measures in particular. While we have discussed practical arguments against subadditivity and for robustness, Heyde et al. (2007) also provide a philosophical basis for both aspects. The robustness requirements are based on their observation (see Heyde et al., 2007, sect. 3) that the concept of “legal realism” requires robustness of legal rules to further their consistent implementation. Their criticism of subadditivity is based on the concept of “legal positivism” which postulates that laws depend on existing social standards (see Green, 2009) and that “there must be a sufficient number [of members of the society] who accept it voluntarily” (Hart, 1997, p. 201). They argue that subadditivity is contrary to the risk perception of most people as suggested by experiments from prospect theory (see Heyde et al., 2007, sect. 3.2). The following results show how natural risk statistics perform in the area of robustness.

Proposition 6.2.3 (see Heyde et al. (2007, thms. 1, 4, 5))
Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a risk statistic and let $x_{(1)}, \ldots, x_{(n)}$ denote the order statistics of $x \in \mathbb{R}^n$.

- ρ is a natural risk statistic if and only if there exists a set of weights $\mathcal{W} \subset \mathbb{R}^n$ with each $w = (w_1, \ldots, w_n) \in \mathcal{W}$ satisfying $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0, 1 \leq i \leq n$ such that

$$\rho(x) = \sup_{w \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\} \forall x \in \mathbb{R}^n.$$

- ρ is a law-invariant coherent risk statistic if and only if there exists a set of weights $\mathcal{W} \subset \mathbb{R}^n$ with each $w = (w_1, \ldots, w_n) \in \mathcal{W}$ satisfying $\sum_{i=1}^n w_i = 1$, $w_i \geq 0, 1 \leq i \leq n$, and $w_1 \leq, \ldots, \leq w_n$ such that

$$\rho(x) = \sup_{w \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\} \forall x \in \mathbb{R}^n.$$

- ρ is an insurance risk statistic if and only if there exists a weight $w = (w_1, \ldots, w_n) \in \mathbb{R}$ satisfying $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0, 1 \leq i \leq n$ such that

$$\rho(x) = \sum_{i=1}^n w_i x_{(i)} \forall x \in \mathbb{R}.$$

Proof. See the proof for theorems 1, 4, and 5 in the online supplement to Heyde et al. (2007). □

The proposition shows that natural risk statistics can be represented by a supremum of L-statistics, law-invariant coherent risk statistics as a supremum of L-statistics with increasing weights for increasing observations, and insurance risk statistics as a single L-statistic. The weights over which the supremum is taken for the former two kinds of risk statistics can be interpreted as scenarios. So just as for the coherent and convex risk measures in section 6.1, we can say that natural and law-invariant coherent risk statistics incorporate scenario analysis. From a robustness point of view, we have seen that certain L-statistics have nice properties (see example 5.3.4). For the law-invariant coherent risk statistics, however, this is offset by their assignment of increasing weights to increasing observations, a property that increases the influence of outliers on the statistic.

Consequently, Heyde et al. (2007) note that the natural risk statistics combine the properties they consider important in a risk measure: consistency with empirical results of decisions under risk, incorporation of scenario analysis, and robustness. We have to be careful, though. While there are certainly natural risk statistics with these properties—as we see in the following examples—it is an immediate consequence of proposition 6.2.3 that every law-invariant coherent risk statistic is also a natural risk statistic.

Example 6.2.4 (Data-based VaR and Tail Conditional Median)
The data-based value at risk at level α can be derived from the definition in example 6.1.4. Assume we have observations $x = (x_1, \dots, x_n)$ of the random variable $X \in \mathbb{R}$ and let $F_n(\cdot; x)$ denote the corresponding empirical distribution function. Then we have

$$\mathrm{VaR}_\alpha^{data}(x) = \inf \{y | F_n(y; x) \geq \alpha\} = x_{(\lceil n\alpha \rceil)}, x \in \mathbb{R}^n.$$

If we set $\mathcal{W}$ to be the single element set $\mathcal{W} = \{w^{\mathrm{Var}_\alpha}\}$ with

$$w_i^{\mathrm{Var}_\alpha} = \begin{cases} 1, & \text{if } i = \lceil n\alpha \rceil, \\ 0, & \text{else,} \end{cases}$$

we get from proposition 6.2.3 that VaR is a natural risk statistic. The VaR has been faulted by Acerbi (2004) and Dowd and Blake (2006), among others, for not belonging to an axiomatic system of risk measures. The classification as a natural risk measure removes the reason for this line of criticism. As the VaR is a natural, but not a coherent risk measure, it satisfies the robustness requirements set up by Heyde et al. (2007). Consequently, they suggest the *tail conditional median* (TCM) at level α, defined as

$$\mathrm{TCM}_\alpha(X) := \mathrm{Med}\left[X|X \geq \mathrm{VaR}_\alpha(X)\right], \tag{6.17}$$

as a risk statistics which incorporates tail information and is a robust alternative to ES. Note that for continuous random variables we have

$$\mathrm{TCM}_\alpha(X) = \mathrm{VaR}_{\frac{1+\alpha}{2}}(X).$$

As a quantile, the TCM has, in contrast to the ES, a bounded influence curve.

Example 6.2.5 (Data-based Expected Shortfall)
The data-based expected shortfall at level $\alpha \in (0, 1)$ can be derived from equation (6.3) to be

$$\mathrm{ES}_\alpha^{data}(X) = \frac{1}{n - \lfloor n\alpha \rfloor}\left((n\alpha - \lfloor n\alpha \rfloor)x_{(\lfloor n\alpha \rfloor+1)} + \sum_{i=\lfloor n\alpha \rfloor+2}^{n} x_{(i)}\right).$$

As $0 \leq n\alpha - \lfloor n\alpha \rfloor < 1$ by definition, we have

$$0 = w_1 = \ldots = w_{\lfloor n\alpha \rfloor} \leq \frac{n\alpha - \lfloor n\alpha \rfloor}{n - \lfloor n\alpha \rfloor} = w_{\lfloor n\alpha \rfloor+1} < \frac{1}{n - \lfloor n\alpha \rfloor} = w_{\lfloor n\alpha \rfloor+2} = \ldots = w_n,$$

and therefore, we can conclude from proposition 6.2.3 the ES is a law-invariant coherent risk statistic.

Chapter 7

Risk Management Procedure

Our goal is to come up with a concept that facilitates robust risk management. The risk measurement procedure introduced by Cont et al. (2010) that we discuss in section 6.1.4 is a step in this direction. But the risk management process includes more than just risk measurement (see section 2.2). Therefore, we need a broader approach which we obtain by extending the risk measurement procedure.

Definition 7.0.6 (Risk Management Procedure)
A *risk management procedure* is a tuple $(G, d, M, \rho, \preceq)$ where G is a (class of) model(s) for the underlying risk factors of the portfolio, d is a probability distance, M is a (possibly simulation-based) estimator of the loss distribution, ρ is a law-invariant risk measure, and $\preceq$ a decision rule.

The ingredients of the risk management procedure are derived from our study of the risk management process—this requires a model for the risk factors, means to deal with unnty, that is, the distance for approximations, the risk measurement procedure M, ρ for risk evaluation, and a decision rule. Our robustness requirements can be incorporated by selecting the distance and the risk measure appropriately.

While the risk management procedure forces us to think coherently about its ingredients, it is still flexible enough to incorporate adjustments to different situations that turn up in practice. In particular, background knowledge about the portfolio and its risk drivers can—and has to—enter in the procedure.

This starts with the selection of possible models for the underlying risk factors. Is a distributional assumption for the risk factor at a single point in time sufficient, or is a time-series model or a model based on continuous time stochastic processes necessary?

Are risk factors likely to exhibit jumps in their price processes? Do we have to model volatility as a stochastic process? Answering these questions requires expert knowledge on the portfolio and the market.

The distance d can be chosen, for example, among the distances presented in chapter 4. Under robustness considerations, the Wasserstein metric is a good choice, as it allows us to derive some theoretical results, focuses on differences in the tails of the distributions, and is easily interpretable.

Concerning the estimator of the loss distribution, there is again a wide variety of possible choices, depending on the portfolio on the one hand and the model on the other hand. For a portfolio consisting of a single stock that is modeled as having normally distributed log-returns, we could simply use a properly scaled log-normal distribution as the loss distribution. Probably closer to practical problems is the asset-liability model for the balance sheet of a life-insurance company by Bauer, Bergmann and Kiesel (2008) that we use in chapter 8. It consists of a set of rules that determine the portfolio value based on the evolution of the underlying asset and has to be used in conjunction with Monte-Carlo simulation of asset paths. The result, however, is a loss distribution, and consequently models like this are a valid choice for M.

Chapter 6 has many examples of risk measures among which we can select one for ρ. Of the two most popular risk measures—value at risk and expected shortfall—the VaR turned out to be the better choice from a robustness point of view. However, we could also use the ES if the model G and the estimator of the loss distribution M are such that the resulting loss distribution allows for a limited robustness result in the sense of Krätschmer et al. (2012*a*).

It remains to discuss the decision rule. We do not fix the character of $\preceq$ in definition 7.0.6 because there is a again a wide variety of possibilities and because different decisions have to be made based on the results, for example to fulfill the requirements of the use test in Solvency II (see section 3.2.1). Simple decision rules can be based on the risk measure. A limit system—positions with a risk as measured by ρ of more than some limit $L \in \mathbb{R}$ are not taken up—constitutes a decision rule:

$$\preceq : \mathbb{R} \to \{0, 1\}, \preceq (\rho(.)) \mapsto \mathbf{1}_{\rho \leq L}.$$

Another possibility would be to use the ordering on the real numbers to rank portfolios

Mean	-0.0147	Skewness	-1.0356
Stand. dev.	0.1611	Kurtosis	9.4282

Table 7.1: Descriptive statistics of stock log-returns

based on the risk (or performance) measurement (see again section 3.2.1).

We present two examples of how the risk management procedure can be used.

Example 7.0.7

We observe 500 historic log-returns of a stock S (obtained by simulation) exhibited in figure 7.1 and want to decide whether or not to invest 10.000 into it. We will not invest if the 97.5% value at risk is higher than 3500. To facilitate later extensions of our portfolio

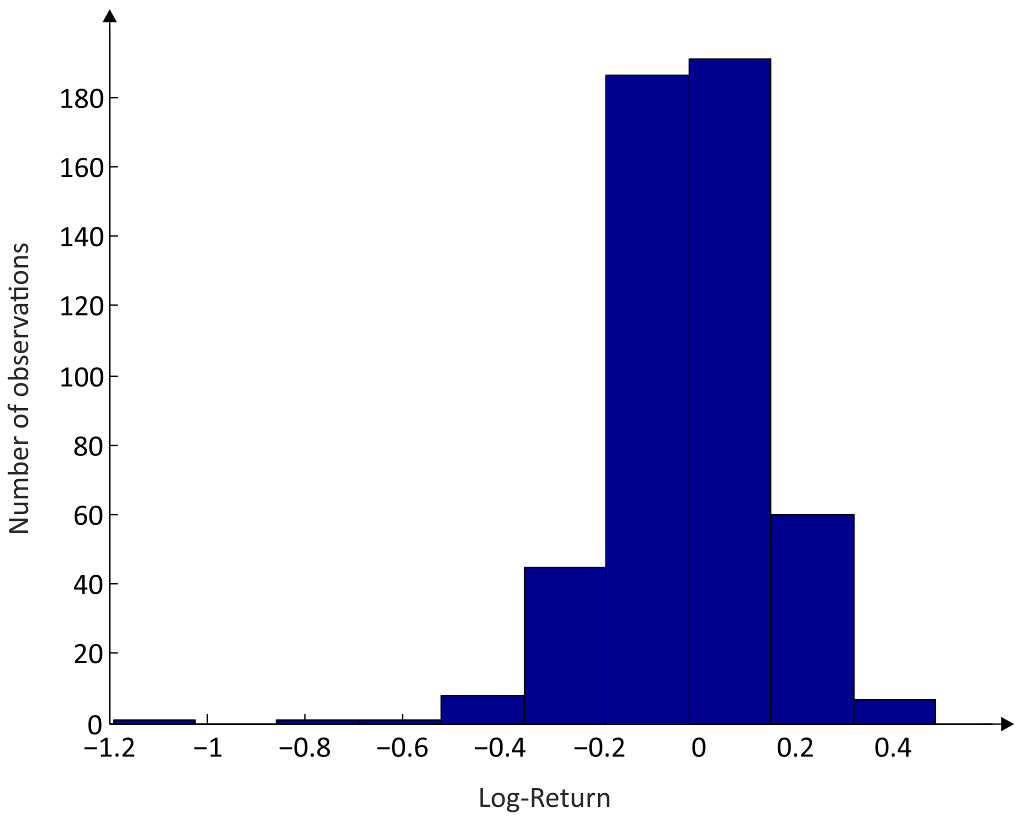

Figure 7.1: Histogram of stock log-returns (simulated data)

to options on this underlying, we want to use a parametric model. Inspection of the histogram and the descriptive statistics in table 7.1 indicate a heavier left tail than in a normal distribution. Based on this, we decide to try to fit a normal distribution for its simplicity and a t-location-scale distribution[1] which might provide a better fit to the left

[1] A r.v. X has a t-location-scale distribution with parameters μ, σ, and ν if $(X-\mu)/\sigma$ has a t distribution with ν degrees of freedom.

	Normal	T-Location-Scale
μ	-1.96%	-2.07%
σ	15.71%	12.64%
ν	N/A	5.83
d_2	0.037	0.028

Table 7.2: Fitting results

tail. We select one of the two distributions as model M based on the Wasserstein metric. In order to do this, we use numerical optimization to find the parameters that minimize the Wasserstein distance (set $d = d_2$) between the empirical distribution and the model distribution. The results are shown in table 7.2. As expected, the distance between the empirical distribution and the model is slightly smaller for the t-location-scale distribution. The difference is also noticeable in a plot of the left tails in figure 7.2.

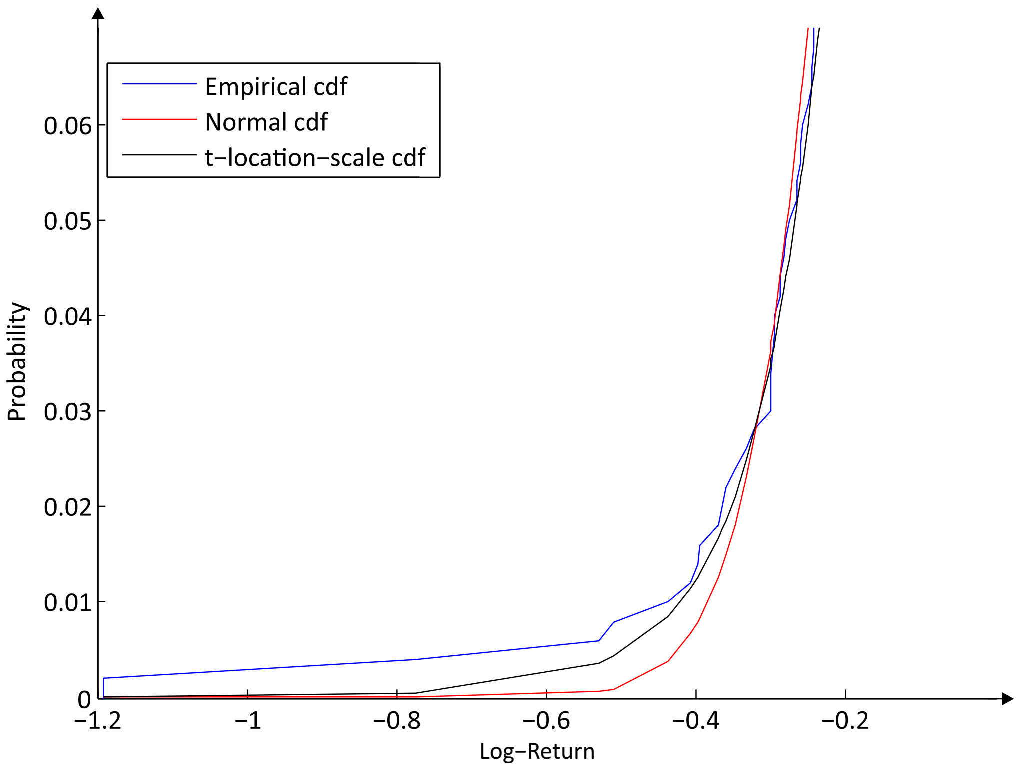

Figure 7.2: Left tail of the empirical and the model distribution functions

The distribution that we use as input to the risk measure can be the model distribution, that is, M can be the identity map on the space of distribution functions. We can compute the value at risk directly, the results are shown in table 7.3. There is a difference between

Model	97.5% value at risk
Historic	3030
Normal	2792
t-location-scale	2828

Table 7.3: Value at risk for the empirical distribution and the models.

the models, but using our decision rule, we go through with the investment in any case.

Example 7.0.8

We want to illustrate the process outlined in chapter 2 with another stylized example, parts of which are taken from Stahl et al. (2012, sect. 4). Consider a market with 10 stocks and a risk-free asset evolving independently from the stocks. We have to decide how much of our wealth we invest in a portfolio Π consisting of the stocks (equally weighted) and in the risk-free asset. Our data set consists of the history of 500 (simulated) daily quotes for each of the stocks as well as for the risk-free rate.

- **Dimension Reduction** We apply Principal Component Analysis to the stock data. We chose the number of factors that we use in the next steps such that they cover 99% of the variance in the data. For our example data set, this turns out to be 6. The true value of the stock portfolio is the sum of the stock prices. Using this value as the dependent variable in a regression against the factors obtained from the PCA, we get an approximation to the portfolio value based on the lower dimensional set of risk factors. If an OLS estimator is used, this implies a small Wasserstein distance between the empirical distribution of the portfolio value on the data set and the approximate portfolio value derived from the risk factors on the data set. For the example data, we obtain $W_2 = 5.54$, and this seems to be reasonably small considering that such a distance would result from moving the distribution by less than a half percent of its mean.

- **Model Building** In order to derive forecasts for our portfolio value and to estimate the risk inherent in the portfolio, we have to build a model for the risk factors. A statistical analysis of the time series of the risk factors shows that, in our example, it is stationary and that for 5 of the 6 factors the hypothesis of normally distributed increments cannot be rejected at a 99% level. To keep matters as simle as possible, we use a normal distribution for the increments of the factors and estimate the parameters with robust methods via the median and the median absolute deviation.

Based on these assumptions and using a Vasicek interest rate model for the risk free rate, we can simulate the development of the asset values over a horizon h of one year to obtain the empirical distribution F of the value of portfolio Π. Figure 7.3 compares the simulation results from the true distribution of the stock portfolio value (in red) with the results from the reduced dimension model (in blue). Obviously, the model does not provide a good fit, which is also indicated by a Wasserstein distance $W_2 = 510$. As the left tail nevertheless fits reasonably well and as the risk measures we use depend mainly this part of the distribution, we do not try to find a better model in this example.

- **Risk Measures and Decision Making** Risk measures such as VaR or TVaR can be computed from the empirical distribution function of the asset values at $t = 1$. These risk measures form the basis for performance measures that can be used for portfolio allocation (see, e.g., Rachev, Menn and Fabozzi (2005, exhibit 13.3)). The performance measures map the empirical distribution to the real numbers, providing us with a way to order portfolios with different distributions. In our example, we use the VaR ratio as performance measure, which relates the mean excess return to the VaR. Via numerical optimization we can find $w \in [0, 1]$ such that investing a fraction w of our wealth in the stocks and $(1 - w)$ in the risk-free asset maximizes the VaR-ratio of the resulting portfolio Π^* at a 99% level. Based on the true distribution, the optimization yields $w = 7.2\%$, based on the reduced dimensional model we obtain $w = 12.5\%$ with VaR ratios of 0.86 and 0.24, respectively. The poor performance of the reduced dimensional model should not be a surprise given the rather large Wasserstein distance obtained in the previous step.

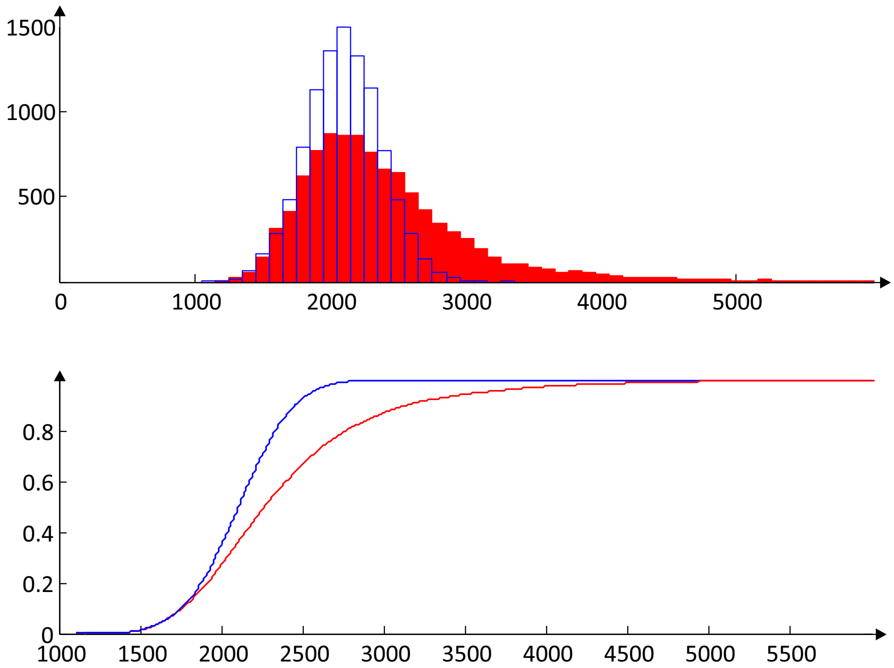

Figure 7.3: Histogram and empirical cdf of the simulated value of the stock portfolio

Part III

The Solvency Capital Requirement

Chapter 8

Mathematical Framework

8.1 Motivation and Definitions

Capital requirements, especially the Solvency Capital Requirement (SCR), are a central part of the Solvency II framework, as we have seen in chapter 3. The calculation as a quantile of own funds over a one-year horizon requires a total balance sheet approach that includes the market-based valuation of assets and liabilities.

Apart from the guidelines in the Solvency II documents, there are several works on market consistent valuation techniques in insurance in the literature. Among them is Wüthrich, Bühlmann and Furrer (2008), who combine results from Buchwalder, Bühlmann, Merz and Wüthrich (2007) (non-life insurance) and Baumgartner, Bühlmann and Koller (2004) (life insurance). These works bring together and integrate the actuarial valuation of the liabilities and the mark-to-market valuation of the assets of an insurance company for the first time. The introductory note by Kalberer (2006) focuses on the Market Consistent Embedded Value (MCEV) and is more concerned with stochastic simulation, although trees, replicating portfolios and closed-form solutions are mentioned as alternatives methods. The calculation of the SCR is closely connected to the notion of the MCEV as calculated according to the CFO Forum (2009) by the present value of future profits (PVFP). Bauer et al. (2011*b*, eq. (1)) present an MCEV-based approximation of the SCR.

$$MCEV = ANAV + PVFP - CoC \tag{8.1}$$

where $ANAV$ is the Adjusted Net Asset Value as the sum of free surplus and required capital, $PVFP$ is the Present Value of Future Profits, and CoC is a Cost of Capital charge, see figure 8.1 for details. These quantities are calculated according to the principles established by the CFO Forum (2009).

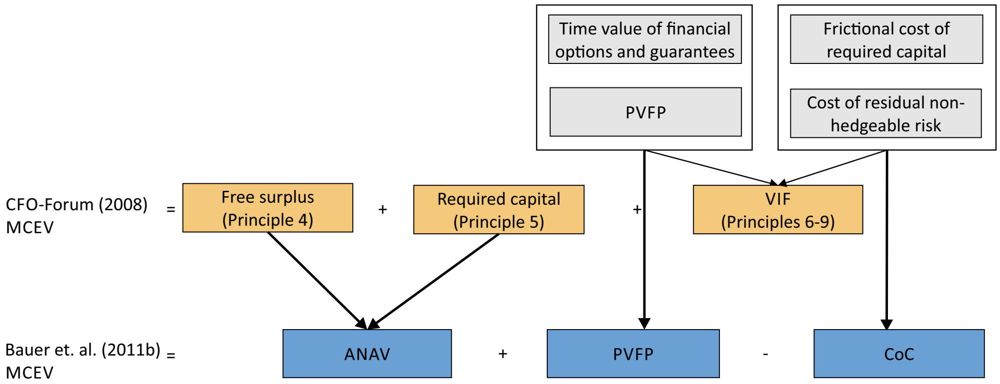

Figure 8.1: Components of the MCEV according to the CFO Forum (2009) and Bauer et al. (2011*b*). VIF is the *value of in-force business.*

The SCR approximation by Bauer et al. (2011*a*) depends on the Available Capital one year into the future (Bauer et al., 2011*a*, eq. (2)):

$$AC_1 =: MCEV_1 + X_1 \tag{8.2}$$

where X_1 is the profit in this year. The Available Capital computed in this way corresponds roughly to the basic own funds in Solvency II. This is sufficient for their—and our—modeling purposes because we do not capture the details in which differences arise anyway[1]. Therefore, we try to calculate the distribution of the available capital at $t = 1$ given the available capital at $t = 0$—especially the 99.5%-quantile which then corresponds to the SCR. The first step in this exercise is the simulation of the market development, in particular of the risk factors that drive the value of the portfolio, from $t = 0$ to $t = 1$. This has to be done under a real-world measure. Given the market development up to $t = 1$, the unknown part of the available capital is the present value of future profits (PVFP). Therefore, the second step consists of finding the PVFP. It can potentially be found by different means: closed formulas, tree methods, stochastic simulation or replicating portfolios (see Kalberer, 2006). Due to the nature of the options and guarantees embedded in many life-insurance contracts, it is usually not possible to find closed-form prices. Tree methods do not seem to be feasible because of the high dimension of the

[1]Additionally, the industry expects the MCEV guidelines to converge to the Solvency II regulations (see, e.g., Towers Watson, 2012)

problem (participation usually depends at least on the yield curve because fixed-income securities are usually a main part of a life insurer's portfolio). Consequently, we consider only the remaining two methods in the following, with a focus on nested simulations. For this method, we generate scenarios under the risk-neutral or pricing measure to estimate the value for each state of the world at $t = 1$ that we have reached in the first simulation step. In the replicating portfolio approach, traded financial instruments are used to mimic the payoffs generated by the insurance portfolio. If this replication is close, the PVFP can be approximated by the value of the replicating portfolio at $t = 1$.

8.1.1 Nested Simulation

One method to evaluate and assess the risk of an insurer's portfolio is nested simulation. While straightforward application of this method, as described above, yields the desired results, it is often too expensive computationally. In the literature, there are mainly two approaches to this problem. The first is to choose the number of inner and outer simulation steps optimally. In this context, Gordy and Juneja (2008) derive the optimal allocation of the computational budget and introduce a jackknife bias-reduction technique, both in the context of derivative portfolios. Broadie et al. (2010) show an algorithm that builds on the results by Gordy and Juneja (2008), but allocates the computational resources sequentially. Screening techniques can be used if the risk measure of interest depends only on the tail of the distribution. Screening was introduced by Lan (2009) and applied to the insurance context by Bauer et al. (2011*b*). The second approach is to use interpolation techniques to reduce the number of necessary steps. Bauer, Bergmann and Reuss (2009) develop a framework in the context of MCEV and the Solvency requirements. They apply a least-squares method, which was originally introduced by Longstaff and Schwartz (2001) for the valuation of American options, to reduce the number of inner steps. Bergmann et al. (2009) use a replicating portfolio as a control variate and try to avoid the change of measure in the inner step by interpolating from the portfolio values in risk neutral scenarios used for MCEV computation in $t = 0$. Liu and Staum (2010) apply stochastic kriging, a method that originated in geostatistics, to the problem.

8.1.2 Replicating Portfolios

The literature on replicating portfolios in (life-) insurance is rather limited. We review some practitioners' accounts first which do not give a rigorous treatment but rather a summary of strengths, weaknesses, and applications of the method. Baxter, Muir and Leung (2008) give a short introduction to replicating portfolios, show some fitting results and list advantages and disadvantages of the approach. They see improved speed, quality, and accuracy of information about insurance liabilities as well as a better understanding of ALM exposures. On the other hand, they mention that a replicating portfolio may be volatile over time and that it is difficult to obtain a good fit. They also note that the application of the method might be limited if liability cash flows depend on the asset allocation or if non-economic assumptions are to be assessed. Erixon and Tubis (2008) suggest the use of replicating portfolios for the valuation and risk management of unit-linked products with investment guarantees. They give an overview of the approach and present a small case study. Mason (2008) focuses on applications in ALM and risk management with regard to Solvency II. Additionally, Clark, Gillespie and Verheugen (2008) and Kalberer (2006) mention replicating portfolios but advocate model compression and stochastic simulation, respectively, as alternative methods to obtain market consistent valuations.

Oechslin et al. (2007) show the application of replicating portfolios to the valuation and management of embedded options and guarantees. In contrast to the publications referenced above, they describe the method in more detail. Starting from the general setup, they find a metric measuring the “distance” of the replicating portfolio to the original portfolio, show how to minimize it with and without constraints, suggest methods to assess the quality of the replication and conduct a small case study. Another work in the spirit of Oechslin et al. (2007) is Seemann (2009). He considers replicating portfolios that match the terminal cash-flows, the continuous cash-flows at each time of the life of the portfolio as well as the terminal cash-flows of intervals. A case study is presented in which different life-insurance products are replicated. Starting from a deterministic savings-only contract, guarantees, repurchase options, participation and mortality risk are are added to the model.

8.2 Algorithms

In this section, we show three different simulation algorithms that can be used to compute the SCR—the nested simulation algorithm, the least-squares Monte-Carlo algorithm, and the least-squares Monte-Carlo algorithm with risk-neutral paths. Before we can work on these algorithms, we have to fill in some details.

Our setup is similar to that of Bauer et al. (2009). Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a filtered probability space with $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ that fulfills the usual hypothesis Protter (2005, ch. 1). T is the maturity of the latest cash flow in the portfolio, $\mathbb{P}$ is the real-world or physical measure. Further, assume that the uncertainty in the market is modeled by a d-dimensional Markov semimartingale[2] $(Y_t)_{t\in[0,T]} = (Y_{t,1}, \dots, Y_{t,d})_{t\in[0,T]}$. The filtration $(\mathcal{F}_t)$ is generated by (Y_t). Additionally, there is a numéraire process $B_t = \int_0^t r_s ds$ with $r_t = r(Y_t)$ the risk-free interest rate. There is a cash flow projection model in place, that is, the cash flow of the portfolio of the insurance company is given by an $\mathcal{F}_t$ measurable process $X_t = f(Y_s, s \in [0,t])$. The process X_t can be discrete or continuous in time, but we will restrict ourselves to the discrete version.

Remark 8.2.1

The cash flow process X_t as defined above is *not* Markov because for any $t \in [0,T]$, X_t usually depends not only on Y_t but on the development of $Y_s, 0 \le s \le t$. A practical solution to this problem can be the introduction of additional "Markov state variables" $D_t \in \mathbb{R}^m$ which, in an insurance context, could represent the bookkeeping, see Bauer et al. (2009). We assume that for any $t \in [0,T]$ the pair (Y_t, D_t) contains all the information we need from $\mathcal{F}_t$, that is, $X_u, t \le u \le T$ and $\mathcal{F}_t$ as well as $B_u, t \le u \le T$ and $\mathcal{F}_t$ are conditionally independent given $\sigma\{(Y_t, D_t)\} \supset \sigma\{Y_t\}$. We write $\mathcal{F}_t \perp_{(Y_t,D_t)} X_u, t \le u \le T$ and $\mathcal{F}_t \perp_{(Y_t,D_t)} B_u, t \le u \le T$.

We assume the existence of a risk-neutral probability measure $\mathbb{Q}$ so that our market is arbitrage free. If the market is not complete, we select one of the risk-neutral measures and label it $\mathbb{Q}$. As a consequence, we can use the risk neutral valuation formula (see Bingham and Kiesel, 2004, thm. 6.1.14) to obtain the value of our portfolio (the PVFP)

$$\begin{aligned} V_t &= \mathbf{E}^{\mathbb{Q}} \left[\sum_{s=t+1}^{T} \exp\left(-(B_s - B_t)\right) X_s \middle| \mathcal{F}_t \right] \\ &= \mathbf{E}^{\mathbb{Q}} \left[\sum_{s=t+1}^{T} \exp\left(-(B_s - B_t)\right) X_s \middle| (Y_t, D_t) \right], 0 \le t \le T. \end{aligned} \tag{8.3}$$

The second equality uses the assumptions from remark 8.2.1.

Proposition 8.2.2

Fix t. Then there exists a measurable function $h_{V_t} : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ such that $V_t =$

[2]For conditions under which a Markov process is a semimartingale, see Çinlar, Jacod, Protter and Sharpe (1980)

$h_{V_t}(Y_t, D_t)$.

Proof. $(Y_t, D_t) : \Omega \to \mathbb{R}^{d+m}$ is a random vector and therefore measurable, V_t is (Y_t, D_t)-measurable by the definition of the conditional expectation. Applying lemma 1.13 from Kallenberg (2002) gives the result. □

Remark 8.2.3

Note that the function h_{V_t} is *deterministic*. The randomness of V_t is in the input (Y_t, D_t), which is a random vector.

In practice, we are interested in two objects: the degenerated random variable $V_0 \in \mathbb{R}$ and the $\mathbb{P}$-distribution of V_1, that is, the distribution given by the cumulative distribution function

$$F_{V_1}(x) = \mathbb{P}(V_1 \leq x). \tag{8.4}$$

In the second case, risk measures derived from the distribution are especially important. We considered examples in chapter 6, among them the value at risk at level $\alpha \in (0, 1)$

$$\mathrm{VaR}_\alpha(V_1) = \inf\{x \in \mathbb{R} | F_V(x) \geq \alpha\} \tag{8.5}$$

which is used with $\alpha = 99.5\%$ to compute the solvency capital requirement, and the expected shortfall at level α

$$\mathrm{ES}_\alpha(V_1) = \mathbf{E}[V_1 | V_1 \leq \mathrm{VaR}_\alpha(V_1)]. \tag{8.6}$$

8.2.1 Nested Simulations

We focus our attention on the computation of the distribution of V_1 via nested simulations as described in section 8.1. Recall that the first step is to generate N paths from $t = 0$ up to $t = 1$ under the real-world measure, that is, we have $(Y_1^{(i)}, D_1^{(i)})$ for $i = 1, \ldots, N$. The second step is to obtain an estimate of the function $h_{V_1}(.,.)$, at least at $(Y_1^{(i)}, D_1^{(i)}), i = 1, \ldots, N$. We consider different approaches for estimating h_{V_1}.

Remark 8.2.4

We follow the literature (see, e.g., Gordy and Juneja, 2008; Bauer et al., 2009; Broadie et al., 2010) and assume that the risk neutral measure does not change from $t = 0$ to

$t = 1$. The measure would change if any market price of risk parameters changed. But those are calibrated to external data, and we should not gain any new external information while running simulations which use reasonable amounts of computing time. Therefore, the assumption seems reasonable.

Unfortunately, we do not have a closed formula for $V_1(Y_1, D_1)$. So instead of calculating $V_1(Y_1, D_1)$, which is deterministic given (Y_1, D_1), directly we can only estimate it by Monte Carlo methods, which, in this context, means that we generate samples of the random variable $\tilde{V}_1(Y_1, D_1; K) \equiv V_1(Y_1, D_1) + \varepsilon_K(Y_1, D_1)$ where $\varepsilon_K(Y_1, D_1)$ is an $\mathcal{F}_T$ measurable random variable. We follow Gordy and Juneja (2008) by assuming that $\varepsilon_K(Y_1, D_1)$ has a mean of zero and $\mathbf{E}\left[|\varepsilon_K|\right] < \infty$. Additionally, we assume that

$$\varepsilon_K \to 0 \text{ a.s.}$$

as $K \to \infty$.

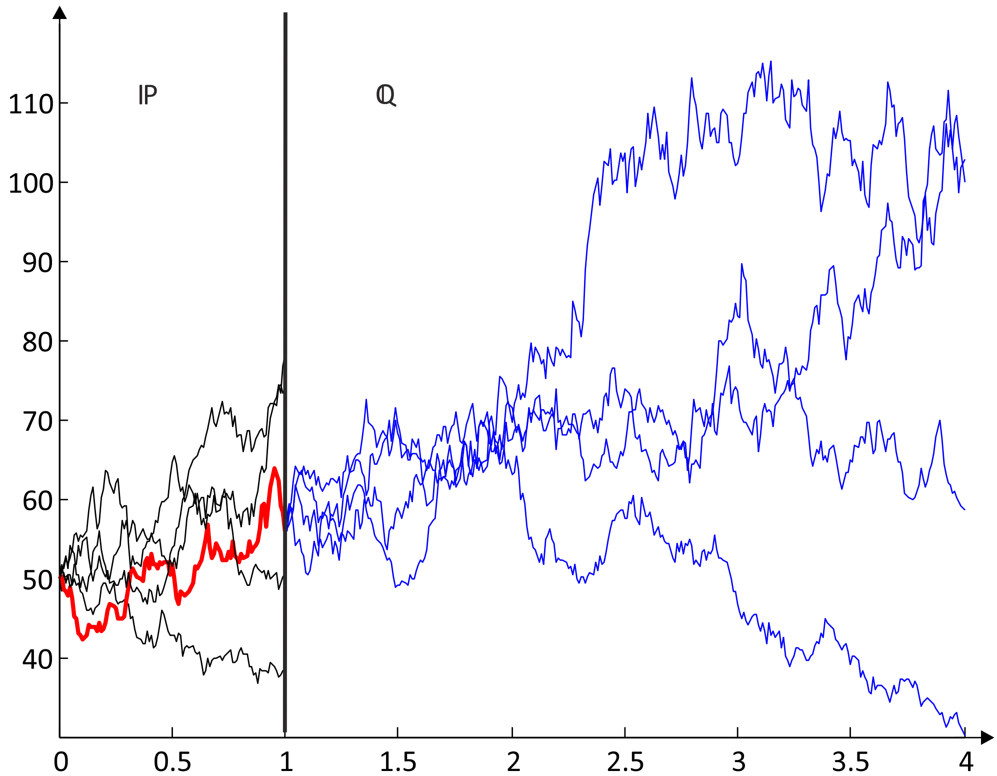

Figure 8.2: Path generation for nested simulation; paths from 0 to 1 are generated under the real-world measure $\mathbb{P}$, paths from 1 to T are generated under the risk-neutral measure

These assumptions fit well into our Monte Carlo framework. To estimate the function at each point $(Y_1^{(i)}, D_1^{(i)})$ of a path i, we generate $K^{(i)}$ paths of the market development from $t = 1$ to $t = T$ under the risk-neutral measure $\mathbb{Q}$ starting at $(Y_1^{(i)}, D_1^{(i)})$, see figure 8.2. Then, evaluating the cash flows on those paths, we get the estimator

$$\hat{V}_1^{(i)}(K^{(i)}) = \frac{1}{K^{(i)}} \sum_{k=1}^{K^{(i)}} \underbrace{\sum_{s=2}^{T} \exp\left(-(B_s^{(i,k)} - B_1^{(i,k)})\right) X_s^{(i,k)}}_{:=PV_1^{(i,k)}} .$$

Here, a superscript (i, k) indicates that we consider the k-th inner path $(k = 1, \ldots, K^{(i)})$ starting from $(Y_1^{(i)}, D_1^{(i)}), i = 1, \ldots, N$. The number of inner paths $K^{(i)}$ can be the same for each outer path i, $K(i) = K$, as in (most parts of) Gordy and Juneja (2008), or it can be chosen dynamically as advocated by Broadie et al. (2010). We use $K^{(i)} \equiv K \,\forall i$ in the following. Note that $PV_1^{(i,k)}$ is an unbiased estimate of $V_1(Y_1^{(i)}, D_1^{(i)})$ which is associated with a mean-zero pricing error $\varepsilon^{(i,k)}$ which can be seen as a realization on the k-th inner path given $(Y_1^{(i)}, D_1^{(i)})$ of $\varepsilon_K(Y_1, D_1)$. As functions of the inner paths, the pricing errors $\varepsilon^{(i,k)}, k = 1, \ldots, K$ are mutually independent samples given $(Y_1^{(i)}, D_1^{(i)})$. Taken together, this means that conducting the nested simulations as described above can be seen as drawing iid realizations from the distribution of $\tilde{V}_1(Y_1, D_1; K) \equiv V_1(Y_1, D_1) + \varepsilon_K(Y_1, D_1)$ for fixed K. By assumption, we have $\tilde{V}_1(Y_1, D_1; K) \to V_1(Y_1, D_1)$ a.s. as $K \to \infty$. In this setting, it makes sense to consider the empirical distribution function

$$\hat{F}_{\tilde{V}_1(K)}(x; N) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{\tilde{V}_1^{(i)}(K) \leq x\}} \tag{8.7}$$

of $\tilde{V}_1(K)$. From the Glivenko-Cantelli theorem (see Karr, 1993, thm. 7.28) we can obtain uniform convergence

$$\lim_{N \to \infty} \sup_{x \in \mathbb{R}} |\hat{F}_{\tilde{V}_1(K)}(x; N) - F_{\tilde{V}_1(K)}(x)| = 0 \text{ a.s.}, \tag{8.8}$$

but this does not help much because we need to know the c.d.f. of V_1, not that of $\tilde{V}_1(K)$. Gordy and Juneja (2008) develop a representation for the mean-squared error (MSE) of

$\hat{F}_{\hat{V}_1(K)}(x;N)$ as follows. Fix $x \in \mathbb{R}$. Then

$$\begin{aligned}
&\mathbf{E}\left[\left(\hat{F}_{\tilde{V}_1(K)}(x;N) - F_{V_1}(x)\right)^2\right] \\
&= \mathbf{E}\left[\left(\hat{F}_{\tilde{V}_1(K)}(x;N) - F_{\tilde{V}_1(K)}(x) + F_{\tilde{V}_1(K)}(x) - F_{V_1}(x)\right)^2\right] \\
&= \mathbf{E}\left[\left(\hat{F}_{\tilde{V}_1(K)}(x;N) - F_{\tilde{V}_1(K)}(x)\right)^2\right] + \left(F_{\tilde{V}_1(K)}(x) - F_{V_1}(x)\right)^2 \\
&= \underbrace{\frac{F_{\tilde{V}_1(K)}\left(1 - F_{\tilde{V}_1(K)}\right)}{N}}_{\text{Variance}} + \underbrace{\left(F_{\tilde{V}_1(K)}(x) - F_{V_1}(x)\right)^2}_{\text{Bias}}
\end{aligned}$$

where the last equality holds because we can see each indicator function in equation (8.7) as an independent realization of a random variable with Bernoulli distribution with parameter $p_x = F_{\hat{V}_1(K)}(x)$. For fixed K, the variance part of the MSE vanishes when we increase the number N of outer simulation steps while the bias part does not change. This observation should not be a surprise given the result in equation (8.8). Gordy and Juneja (2008) also obtain a result on the behavior of the bias part, but only under the following assumptions.

Assumption 8.2.5
Define $\tilde{\varepsilon}(K) := \varepsilon(K) \cdot \sqrt{K}$.

- The joint pdf $g_K(v,z)$ of V_1 and $\tilde{\varepsilon}(K)$ as well as its first and second partial derivatives w.r.t. v exist for each K and any $(v,z) \in \mathbb{R}^2$.
- For each $K \in \mathbb{N}_+$ there exist non-negative functions $p_{0,K}(.), p_{1,K}(.), p_{2,K}(.)$ with

  $$\sup_K \int_{-\infty}^{\infty} |z|^r p_{i,K}(z)dz < \infty$$

  for $i = 0,1,2$ and $0 \leq r \leq 4$ such that

  $$\begin{aligned}
  g_K(v,z) &\leq p_{0,K}(z) \\
  \left|\frac{\partial}{\partial v} g_K(v,z)\right| &\leq p_{1,K}(z) \\
  \left|\frac{\partial^2}{\partial v^2} g_K(v,z)\right| &\leq p_{2,K}(z)
  \end{aligned}$$

  for all $v, z \in \mathbb{R}$.

Proposition 8.2.6

Let f_{V_1} denote the density of V_1, that is, $f_{V_1} = F'_{V_1}$. Define

$$\Theta(u) := \frac{1}{2} f(u) \mathrm{E}\left[\sigma^2(Y_1, D_1) \mid V_1 = u\right], \tag{8.9}$$

$$\theta_u := -\Theta'(u). \tag{8.10}$$

Fix $x \in \mathbb{R}$. Then, under Assumption 8.2.5,

$$F_{\tilde{V}_1(K)}(x) = F_{V_1}(x) - \frac{\theta_x}{K} + O\left(K^{-3/2}\right).$$

Proof. The proof is completely analogous to Gordy and Juneja (2008, Appendix 2). □

Remark 8.2.7

A consequence of Proposition 8.2.6 is that the estimator $\hat{F}_{\tilde{V}_1(K)}(x; N)$ converges to $F_{V_1}(x)$ in probability for each $x \in \mathbb{R}$ as $N, K \to \infty$ (see Karr, 1993, Prop. 5.11). In contrast, we could get almost sure convergence uniform in x from the Glivenko-Cantelli theorem if we were able to compute equation (8.3) explicitly.

We summarize the nested simulation method in algorithm 1.

Algorithm 1 Nested simulation (NS)

Value the portfolio at $t = 0$ using N_0 risk-neutral sample paths
Generate N_1 paths under the real-world measure from $t = 0$ to $t = 1$
for each path **do**
  Value the portfolio at $t = 1$ given the history up to $t = 1$ using K_1 inner simulations
end for

8.2.2 Least-Squares Monte-Carlo

A different approach to the problem is to estimate the function h_{V_1} as a whole. In this case, we have to make some structural assumptions on h_{V_1} which result in a parametrization. Once such a parametrization is found and the parameters are estimated, we can evaluate the function at the points of interest, namely $(Y_1^{(i)}, D_1^{(i)}), i = 1, \ldots, N$.

The least squares method presented by Bauer et al. (2009) belongs to this group of methods. They assume that the function h_{V_1} belongs to the Hilbert space

$$L^2(\Omega, \sigma(Y_1, D_1), \mathbb{P}(Y_1, D_1))$$

Then, there exists a linearly independent and complete sequence $(e_k(Y_1, D_1))$ of basis functions on this space which can be used to approximate h_{V_1} by

$$h_{V_1}(Y_1, D_1) \approx \hat{V}_1^{(M)}(Y_1, D_1) = \sum_{k=1}^{M} \alpha_k \cdot e_k(Y_1, D_1).$$

Consequently, they generate N sample paths from $t = 0$ up to T, under the real-world measure from $t = 0$ to $t = 1$ and continuing under the risk neutral measure from $t = 1$ to $t = T$. From the realized cash flows on these paths, they obtain least-squares estimates $\hat{\alpha}_k$ of the coefficients α_k. Therefore,

$$\hat{h}_{V_1}(Y_1, D_1) \approx \hat{V}_1^{(M)}(Y_1, D_1) \approx \hat{V}_1^{(M,N)}(Y_1, D_1) = \sum_{k=1}^{M} \hat{\alpha}_k \cdot e_k(Y_1, D_1). \tag{8.11}$$

This function can be evaluated at $(Y_1^{(i)}, D_1^{(i)}), i = 1, \ldots, N$ obtained, for example, from the first year sub-paths generated for the calibration. Then, we can estimate the empirical distribution function from these realizations.

We quote the following result from Bauer et al. (2011*a*, Prop. 3.1).

Proposition 8.2.8
$\hat{V}_1^{(M)} \to V_1$ in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{(Y_1,D_1)})$, $M \to \infty$, and $\hat{V}_1^{(M,N)} \to \hat{V}_1^{(M)}$, $N \to \infty$, $\tilde{\mathbb{P}}$ almost surely.

Proof. See Bauer et al. (2011*a*, Prop. 3.1). □

The measure $\tilde{\mathbb{P}}$ is defined via the Radon-Nikodym derivative

$$\frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} = \frac{\frac{\partial \mathbb{Q}}{\partial \mathbb{P}}}{\mathrm{E}\left[\frac{\partial \mathbb{Q}}{\partial \mathbb{P}} \,\middle|\, \mathcal{F}_1\right]}.$$

We summarize the least-squares Monte-Carlo method in algorithm 2.

Algorithm 2 Least-squares Monte-Carlo (LSM)

Value the portfolio at $t = 0$ using N_0 risk-neutral sample paths
Generate N_1 paths under the real-world measure from $t = 0$ to $t = 1$
for each path **do**
  Change to the risk-neutral measure and generate a path up to T
end for
Estimate h_{V_1} using equation (8.11) based on the risk-neutral paths from $t = 1$ to T
Evaluate h_{V_1} at $t = 1$ at the end points of the previously generated real-world paths

For the estimation of the function h_{V_1}, it is not important under which measure $(Y_1^{(i)}, D_1^{(i)})$ are generated. In fact, if the change of measure from $\mathbb{P}$ to $\mathbb{Q}$ in $t = 1$ is difficult in practice, it might be convenient to estimate the function using paths that are generated entirely under the risk neutral measure, as suggested by Bergmann et al. (2009).

We can adapt the convergence result from proposition 8.2.8 to this setting.

Proposition 8.2.9

Assume $V_1 \in L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{(Y_1,D_1)})$. $\hat{V}_1^{(M)} \to V_1$ in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{(Y_1,D_1)})$, $M \to \infty$, and $\hat{V}_1^{(M,N)} \to \hat{V}_1^{(M)}$, $N \to \infty$, $\mathbb{Q}$ almost surely.

Proof. The proof is completely analogous to the proof of Prop. 8.2.8. □

As L^2-convergence implies convergence in probability (see Karr, 1993, section 5.2.1), Prop. 8.2.9 implies the convergence $V_1^{(M)} \xrightarrow{P} V_1$ in $\mathbb{Q}$-probability. Due to the following result, this is equivalent to $V_1^{(M)} \xrightarrow{P} V_1$ in $\mathbb{P}$-probability.

Proposition 8.2.10

Let μ and ν be two probability measures on some measure space $(\Omega, \mathcal{F})$ and let $\nu \ll \mu$. If a sequence of random variables $X_n, n \geq 0$ in $(\Omega, \mathcal{F}, \mu)$ converges to a random variable X in probability, then the sequence in $(\Omega, \mathcal{F}, \nu)$ also converges to X in probability.

Proof. Let $X_n, n \geq 0$ be a sequence of random variables in $(\Omega, \mathcal{F})$. By assumption,

$$\forall \delta > 0 \; \exists n(\delta) \in \mathbb{N} : \mu\left(\{\omega \in \Omega | \, \|X_n(\omega) - X(\omega) \geq \epsilon\|\}\right) < \delta \forall n > n(\delta). \tag{8.12}$$

Since $\mu, \nu < \infty$, $\nu \ll \mu$ is equivalent (see Billingsley (1986, p.443) for details) to

$$\forall \tilde{\delta} > 0 \; \exists \tilde{\epsilon} > 0 : \nu(A) < \tilde{\epsilon} \text{ if } \mu(A) < \tilde{\delta} \text{ for any } A \in \Omega. \tag{8.13}$$

Fix $\overline{\delta} > 0$. The there exists $\overline{\epsilon} > 0$ (by 8.13) and $n(\overline{\epsilon})$ (by 8.12) such that

$$\mu\left(\{\omega \in \Omega | \, \|X_n(\omega) - X(\omega) \geq \epsilon\|\}\right) < \overline{\epsilon} \; \forall n > n(\overline{\epsilon})$$

and therefore

$$\nu\left(\{\omega \in \Omega | \, \|X_n(\omega) - X(\omega) \geq \epsilon\|\}\right) < \overline{\delta} \; \forall n > n(\overline{\epsilon}).$$

□

Proposition 8.2.11

Assume that the Radon-Nikodym density $d\mathbb{Q}/d\mathbb{P}$ is bounded $\mathbb{P}$- almost everywhere. Then Prop. 8.2.9 implies that $\hat{V}_1^{(M)} \to V_1$ in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{(Y_1,D_1)})$.

We summarize the least-squares Monte-Carlo with risk-neutral paths method in algorithm 3.

Algorithm 3 Least-squares Monte-Carlo with risk neutral paths (LSMQ)

Value the portfolio at $t = 0$ using N_0 risk-neutral sample paths
Estimate h_{V_1} using equation (8.11) based on the risk-neutral paths from $t = 1$ to T
Generate N_1 paths under the real-world measure from $t = 0$ to $t = 1$
Evaluate h_{V_1} at $t = 1$ at the end points of the previously generated real-world paths

Chapter 9

Empirical Study

We investigate the performance of the different algorithms for the computation of an insurance company's loss distribution. Following our approach using the risk management procedure, we have to specify its elements. We use the 2-Wasserstein distance d_2 where possible, but also consider other possibilities. The estimator of the loss distribution M is computed using the asset-liability model exhibited in section 9.1. For the probability model G, we consider two alternatives in section 9.2. Due to its robustness properties and its wide-spread use in practice as well as in the Solvency II regulations, we will use the value at risk as the risk measure ρ. Additionally, we also consider the expected shortfall. In this empirical study, we refrain from specifying the decision rule because our focus is on the properties of the other elements.

9.1 Asset and Liability Model

For our empirical studies, we use the simple model of an insurance undertaking which was proposed by Zaglauer and Bauer (2008). The model is based on a simplified balance sheet that is valued using a market-based approach. The balance sheet items in the model are the assets A_t, the liabilities L_t, and the reserves R_t, each at time t, see figure 9.1.

The contracts in the undertaking's portfolio are fixed-term (life-time of the contracts is $[0, T]$) single premium $P = L_0$ life-insurance contracts with guaranteed annual return g. The company has to share additional profits on top of the guaranteed rate with the policy holders according to the MUST-case exhibited by Bauer, Kiesel, Kling and Ruß (2006), which corresponds to the legal requirements in Germany. The participation amounts to a fixed fraction δ of earnings on book value. The portion of earnings on book value that

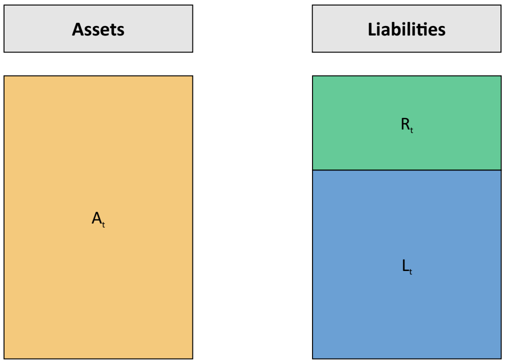

Figure 9.1: Balance sheet in the model, source: Zaglauer and Bauer (2008).

is not used for the participation is paid to to the shareholders as dividends d_t. If the earnings on the assets and the reserves are not sufficient to cover the guaranteed return of the contract, shareholders have to inject capital c_t.

The asset value A_t evolves according to some stochastic model. At the end of the period $[t-1,t]$, when the asset value is A_t^-, dividends d_t are paid to the shareholders or the shareholders might have to inject capital c_t. The asset value after these corporate actions is then $A_t^+ = A_t^- - d_t + c_t$ (see Bauer et al., 2011*b*, p. 20). Assuming that the earnings on book value are a fraction y of the earning on market value $E_t = A_t^- - A_{t-1}^+$. Thus, given the development of asset prices from $t-1$ to t, we can compute L_t, c_t, and d_t according to the following formulas provided by Bauer et al. (2011*b*, pp. 20-21):

$$L_t = (1+g)L_{t-1} + (\delta y E_t - gL_{t-1})^+ \, , t = 1, \dots, T \tag{9.1}$$

$$d_t = (1-\delta)yE_t\mathbf{1}_{\{\delta y E_t > gL_{t-1}\}} + (yE_t - gL_{t-1})\,\mathbf{1}_{\{\delta y E_t \leq gL_{t-1} \leq yE_t\}} \tag{9.2}$$

$$c_t = \left(L_t - A_t^-\right)^+ \tag{9.3}$$

Bauer et al. (2011*b*) also show how to compute the available capital in this model. Setting

$$X_t = \begin{cases} d_t - c_t & t \in [0, T-1] \\ d_T - c_T + R_T & t = T, \end{cases}$$

they obtain from a shareholder perspective

$$AC_s = \mathbf{E}^{\mathbb{Q}}\left[\sum_{t=1}^{T} \exp\left(-(B_t - B_s)\right) X_t\right], s = 0, 1, \tag{9.4}$$

and from a policyholder perspective

$$AC_0 = A_0 - \mathbf{E}^{\mathbb{Q}}\left[\exp\left(-B_T\right) L_T\right] \tag{9.5}$$

$$AC_1 = A_1^+ - \mathbf{E}^{\mathbb{Q}}\left[\exp\left(-(B_T - B_1)\right) L_T | \mathcal{F}_1\right] + X_1. \tag{9.6}$$

9.2 Asset Price Models

9.2.1 Black-Scholes-Vasicek Model

It remains to specify a stochastic model for the asset prices and discount factors. One possibility is to again follow Bauer et al. (2011*b*) and model the asset price process (A_t) as a geometric Brownian motion and the short rate as an Ornstein-Uhlenbeck process r_t—as in the Vasicek model—that is correlated with (A_t):

$$dA_t = \mu A_t dt + \sigma_A(\rho dW_t + \sqrt{1-\rho^2} dZ_t), A_0 > 0$$
$$dr_t = \kappa(\theta - r_t)dt + \sigma_r dW_t, r_0 > 0.$$

Here $\rho \in [-1, 1]$ is the correlation, $\mu \in \mathbb{R}$ and $\sigma_A > 0$ are the drift and volatility, respectively, of the asset price process, and κ, θ, σ_r are the speed of mean reversion, long-term mean, and volatility, respectively, of the short-rate process. W_t and Z_t are two independent Brownian motions under the real world measure $\mathbb{P}$. A change of measure to the risk-neutral pricing measure $\mathbb{Q}$ leads to

$$dA_t = r_t A_t dt + \sigma_A(\rho dW_t + \sqrt{1-\rho^2} dZ_t), A_0 > 0$$
$$dr_t = \kappa(\tilde{\theta} - r_t)dt + \sigma_r dW_t, r_0 > 0.$$

with $\tilde{\theta} = \theta - \frac{\lambda \sigma_r}{\kappa}$, the market price of interest rate risk λ, and independent $\mathbb{Q}$-Brownian motions W, Z (see Bauer et al., 2011*b*, sect. 6.1.3).

An advantage of this model is its tractability. First, the stochastic differential equations of both the asset and the short rate process can be solved explicitly, see Glasserman (2004, sect. 3.2.1 & 3.3.1):

$$A_t = A_s \exp\left(\left(\mu - \frac{\sigma_A^2}{2}\right) + \sigma_A(\rho(W_t - W_s) + \sqrt{1-\rho^2}(Z_t - Z_s)\right) \tag{9.7}$$

$$r_t = r(s)\exp(-\kappa(t-s)) + \theta\,(1 - \exp(-\kappa(t-s))) + \sigma_r \int_s^t \exp(-\kappa(t-u))dW(u) \tag{9.8}$$

for $0 \leq s < t \leq T$. In particular, the short-rate as well as the log-asset price process have normal distributions. Second, the distribution of the integrated short-rate process $I_t = \int_0^t r_s ds$ conditional on $r_s, s < t$ and its covariance with the short-rate itself are known, see Glasserman (2004, pp. 113f):

$$\begin{aligned} I_t - I_s | r_s \sim & N\left(\frac{r_s - \theta}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) + \theta(t-s),\right. \\ & \left.\frac{\sigma_r}{\kappa}\sqrt{(t-s) - \frac{2}{\kappa}\left(1 - e^{-\kappa(t-s)}\right) + \frac{1}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right)}\right) \\ \mathrm{Cov}(r_r, I_t) = & \frac{\sigma_r^1}{2\kappa}\left(1 + e^{-2\kappa(t-s)} - 2e^{-\kappa(t-s)}\right). \end{aligned}$$

These results allow exact simulation of the process at any time t of interest. The model allows the computation of bond and option prices in closed form, a property which is very convenient for the calibration of the model to data. The price at time t of a European call option with strike K and maturity T is given by (see Kim, 2002, sect. 2.1.1)

$$V_c(t) = A_t \Phi(d_1^V(t,T)) - P(t,T)K\Phi(d_2^V(t,T))$$

where $P(t,T)$ is the time t price of a zero coupon bond in the Vasicek model,

$$P(t,T) = \exp(\frac{1}{2}\Sigma_{22}(t,T) - B(t,T))$$

and d_1^V, d_2^V are, in analogy to the Black-Scholes model with deterministic rates, given by

$$\begin{aligned} d_1^V(t,T) &= \frac{(C(t,T) + \Sigma_{12}(t,T))}{\sqrt{D(t,T)}} \\ d_2^V(t,T) &= d_1^V(t,T) - \sqrt{D(t,T)} \end{aligned}$$

and auxiliary functions

$$
\begin{aligned}
\Sigma_{11}(t,T) &= \sigma^2(T-t) \\
\Sigma_{12}(t,T) &= \frac{\sigma\sigma_r\rho}{\kappa}\left(\frac{e^{-\kappa(T-t)}-1}{\kappa}+(T-t)\right) \\
\Sigma_{22}(t,T) &= \frac{\sigma_r^2}{\kappa^2}\left((T-t)-\frac{(3-4e^{-\kappa(T-t)}+e^{-2\kappa(T-t)})}{2\kappa}\right) \\
B(t,T) &= -\frac{1}{\kappa}\left[\left(r_t-\tilde{\theta}\right)\left(e^{-\kappa(T-t)}-1\right)-\tilde{\theta}(T-t)\right] \\
C(t,T) &= \log\left(\frac{A_t}{K}\right)+\frac{\Sigma_{11}(T-t)}{2}+B(t,T) \\
D(t,T) &= \Sigma_{11}(T-t)+2\Sigma_{12}(T-t)+\Sigma_{22}(T-t).
\end{aligned}
$$

9.2.2 Heston-Cox-Ingersoll-Ross Model

A weakness of the Black-Scholes-Vasicek model is that it is difficult to obtain a reasonable fit to option prices, due to the assumption of constant volatility. Market prices usually show different implied volatilities across time and strikes (see Cont and da Fonseca (2002) among many others). To this end, we therefore consider an alternative specification, which is more general. We model the asset price process with a Heston stochastic volatility model (see Heston, 1993), the interest rate with a Cox-Ingersoll-Ross process (see Cox, Ingersoll and Ross, 1985):

$$
\begin{aligned}
dA_t &= \mu A_t dt + \sqrt{v_t} A_t dW_1 \\
dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_2 \\
dr_t &= \kappa_r(\theta_r - r_t)dt + \sigma_r\sqrt{r_t}dW_3
\end{aligned}
\tag{9.9}
$$

where we assume that $dW_1dW_2 = \rho dt$ and $dW_1dW_3 = dW_2dW_3 = 0$. Here, $\mu \in \mathbb{R}$ is the expected return on the asset[1], $\kappa > 0$ the volatility's speed of mean reversion, $\theta > 0$ the volatility's long term mean, $\sigma > 0$ the volatility of volatility, and ρ the correlation between asset and volatility. The corresponding parameters for the short interest rate process are $\kappa_r, \theta_r > 0$, and $\sigma_r > 0$. A change to an equivalent martingale measure for valuation purposes can be done such that only the long-term means of both volatility and

[1]For an arbitrage free model, μ should be a function $\mu(r_t, V_t)$ such that $\mu(r_t, 0) = r_t$, see as argued by Alan L. Lewis in the discussion `http://www.wilmott.com/messageview.cfm?catid=4&threadid=71093`. We follow this suggestion.

short-rate are affected (see Brigo and Mercurio, 2006, sect. 3.2.3). Consequently, we have risk neutral dynamics given by

$$
\begin{aligned}
dA_t &= r_t A_t dt + \sqrt{v_t} A_t dW_1 \\
dv_t &= \kappa(\tilde{\theta} - v_t)dt + \sigma\sqrt{v_t} dW_2 \\
dr_t &= \kappa_r(\tilde{\theta}_r - r_t)dt + \sigma_r\sqrt{r_t} dW_3
\end{aligned}
$$

where $\tilde{\theta}$ and $\tilde{\theta}_r$ are the long-term means of volatility and short-rate, respectively, and the other parameters are as in equation (9.9).

The simulation of paths is more involved in this model than in the Black-Scholes-Vasicek model. This is due to the two square root diffusion processes involved. A simple Euler discretization scheme would allow for negative interest rates and negative volatilities (see Lord, Koekkoek and Dijk, 2010) that are not possible in the continuous time model. The conditional distribution of the square-root diffusion process is a non-central chi-square distribution (see Glasserman, 2004, sect. 3.4)—this makes exact simulation on a discrete time grid possible, albeit with high computational cost. The caveat, however, is that exact simulation of the asset price process then requires sampling from the conditional distribution of the integrated volatility process, which is only available via its characteristic function combined with Fourier-Inversion techniques, as developed by Broadie and Kaya (2006). Observing that these methods are computationally rather slow and difficult to implement correctly, Andersen (2008) suggests approximate simulation schemes that belong to a class of methods based on moment matching. We take a closer look at his QE-scheme that presents a good compromise between speed, accuracy, and ease of implementation.

We present a summary of the QE-scheme (see Andersen, 2008, sect. 3.2) for the volatility process. As the interest rate process is of the same type, it can be applied analogously. The scheme is based on the observation that the conditional distribution of v_t given $v_s, s < t$ is a non-central chi-square distribution with a non-centrality parameter proportional to v_s. For large v_s, Andersen (2008) approximates this distribution by that of a moment-matched squared Gaussian random variable, while for small v_s, he uses a density approximation based on a Dirac-mass at 0 combined with an exponential tail. Andersen (2008) also supplies us with a switching rule to decide when v_s is "small" or "large" and with a discretization scheme for the asset price. He derives the latter from the exact representation of the asset price process by approximating the integrated volatility process with a generalized trapezoidal quadrature rule $\int_t^s v_u du \sim (t-s)(\gamma_1 v_t + \gamma_2 v_s)$ where the constants γ_1, γ_2 are either set to $\gamma_1 = \gamma_2 = \frac{1}{2}$ or are obtained via moment matching.

The considerations regarding the simulation suggests that we cannot hope for closed-form options prices in the Heston-Cox-Ingersoll-Ross model. As the calibration of the model parameters to option prices potentially requires the re-pricing of options in the model under many different combinations of parameters, we need an alternative pricing method that is faster than simulation. We can compute the characteristic function of the log-asset price in the model. Using our assumption that $dW_1 dW_3 = dW_2 dW_3 = 0$, we can easily extend the results by Mikhailov and Nögel (2003) to obtain[2]

$$\varphi_{HCIR,j}(u; S_t, \tau = T - t) = \exp(C_j(u) + D_j(u)Vt + iu\log(S_t)) \cdot \varphi_{CIR}(u), j = 1, 2$$

where

$$\begin{aligned}
C_i(u) &= \frac{\kappa\theta}{\sigma^2}\left((b_j - \rho\sigma ui + d)\tau - 2\log\left(\frac{(1-ge^{d\tau})}{(1-g)}\right)\right)\\
D_i(u) &= \frac{b_j - \rho\sigma ui + d}{\sigma^2}\left(\frac{1-e^{d\tau}}{(1-ge^{d\tau}}\right)\\
g &= \frac{b_j - \rho\sigma ui + d}{b_i - \rho\sigma ui - d}\\
d &= \sqrt{(\rho\sigma ui - b_j)^2 - \sigma^2(2u_j ui - u^2)}\\
u_1 &= 0.5 = -u_2\\
b_1 &= \kappa - \rho\sigma\\
b_2 &= \kappa
\end{aligned}$$

and where φ_{CIR} is the characteristic function of the integrated Cox-Ingersoll-Ross process. It is given by

$$\varphi_{CIR}(u) = A(\tau)e^{iur_t B(\tau)}$$

[2]There are also results for non-zero correlation available in the literature, see, e.g., Grzelak and Oosterlee (2011), Scott (1997), and Ahlip and Rutkowski (2009).

with

$$\begin{aligned} A(\tau) &= \left(\frac{2h \exp{(\kappa_r + h)\tau/2}}{2h + (\kappa_r + h)\exp(\tau h) - 1} \right)^{2\kappa_r\theta_r/\sigma_r^2} \\ B(\tau) &= \frac{2(\exp(\tau h) - 1}{2h + (\kappa_r + h)\exp(\tau h) - 1} \\ h &= \sqrt{\kappa_r^2 + 2\sigma_r^2}. \end{aligned}$$

This formula can be derived directly from the pricing function for zero-coupon bonds in the CIR model, which is provided by Brigo and Mercurio (2006, sect. 3.2.3). The characteristic function can then be used for option pricing based on Inverse Fourier Transform introduced to option valuation by Stein and Stein (1991) and Heston (1993). The method, also used in case of the Heston model with deterministic rates by Mikhailov and Nögel (2003), can be adapted starting from the Black-Scholes-type equation

$$V_c(t) = SP_1 - KP_{CIR}(t,T)P_2$$

where P_1, P_2 are the probabilities of the option being in-the-money at maturity with the asset and the bond as numéraire, respectively (see Schmelzle, 2010, sect. 4.1), and $P_{CIR}(t,T)$ is the zero-coupon bond price in the CIR model given by $\varphi_{CIR}(u)|_{u=i}$. P_1 and P_2 can then be computed using a Fourier inversion theorem (Shephard, 1991, thm. 3)

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-iuK}\varphi_{HCIR,j}(u;S_t;T-t)}{iu}\right]du, j = 1,2.$$

9.3 Data and Calibration

For our experiments, we fit the two models to market data. The asset is represented by an index that is generated from a portfolio of the German REX and DAX indices, with weights of 90% and 10%, respectively. We compile a price history for the index from the beginning of 2007 up to the valuation date of October 11, 2012, from the Bloomberg information system (see figure 9.2), which also provides us with option prices for different strikes and maturities at the valuation date. As a proxy for the historic short-rate, we use the time series of the 1-week EONIA interest rate[3] As shown in figure 9.3, there is a

[3]We use the 1-week rate instead of the overnight rate because the latter tends to spike at the end of ECB reserve maintenance periods, see Fecht, Nyborg and Rocholl (2008).

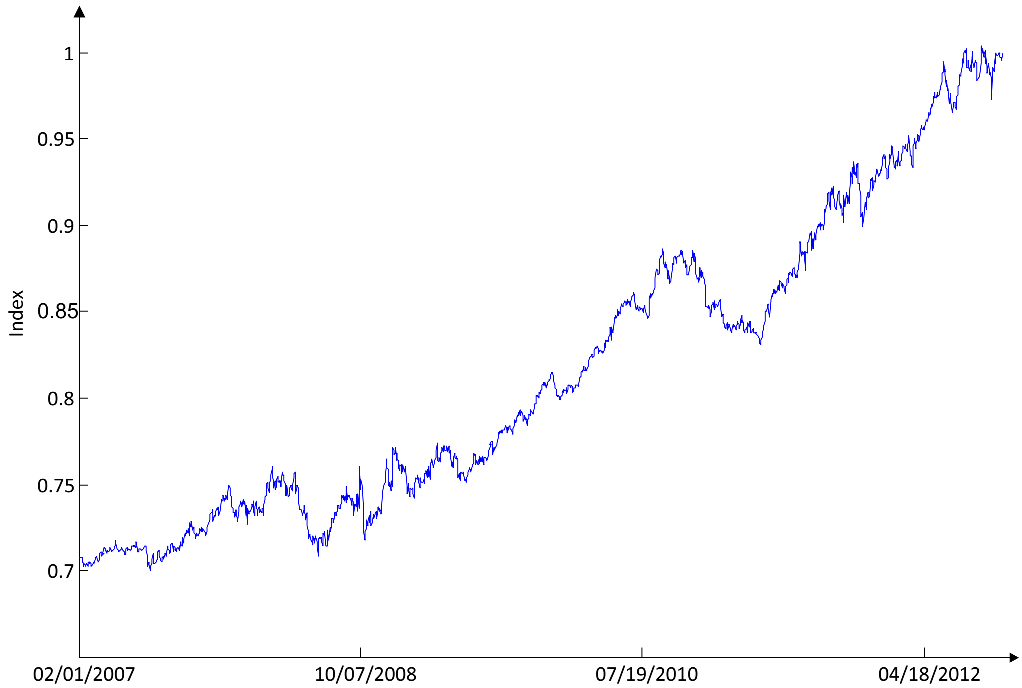

Figure 9.2: Portfolio consisting of investments in the REX (portfolio weight 90%) and DAX (portfolio weight 10%) indices

sharp drop in the EONIA rates in October and November 2010. Computing the means of the process up to 04/16/2009 and from this date on to the end of the time series gives values of 3.5% and 1%, respectively. Welch's t-test (see Sawilowsky, 2002) confirms that the two means differ significantly at any reasonable level of confidence. As we do not want to complicate matters by modeling a regime switch, we use only the data starting from this date. The corresponding term-structure of interest rates is the Euro overnight indexed swap (OIS) curve, see figure 9.4.

We start the parameter estimation in the Black-Scholes-Vasicek model with the short-rate parameters. In this step, we try to fit the bond prices in the Vasicek model to zero bond prices derived from the term-structure of interest rates, minimizing the squared Euclidean distance between the two. This provides us with the parameters under the risk-neutral measure. The real-world drift is then estimated from the time-series data via maximum-likelihood, taking into account the distributional properties exhibited in equation (9.8).

Concerning the parameters for the Black-Scholes model, we follow the same route and first estimate the implied volatility by least-squares from the option prices and then estimating the drift via maximum-likelihood. The results are summarized in table 9.1.

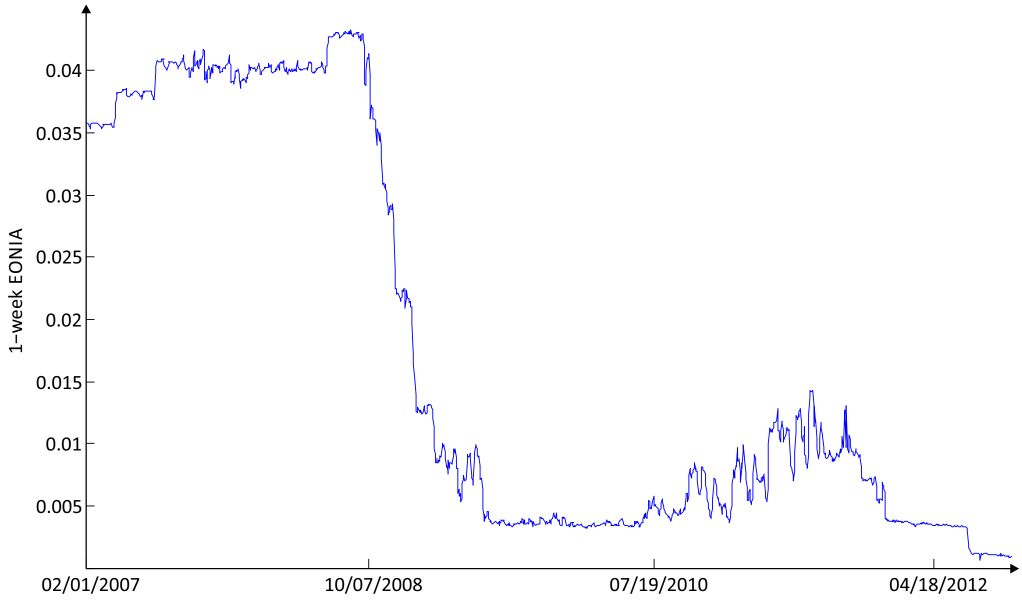

Figure 9.3: 1-week EONIA interest rate from 01/02/2007 to 10/11/2012, data from Bloomberg

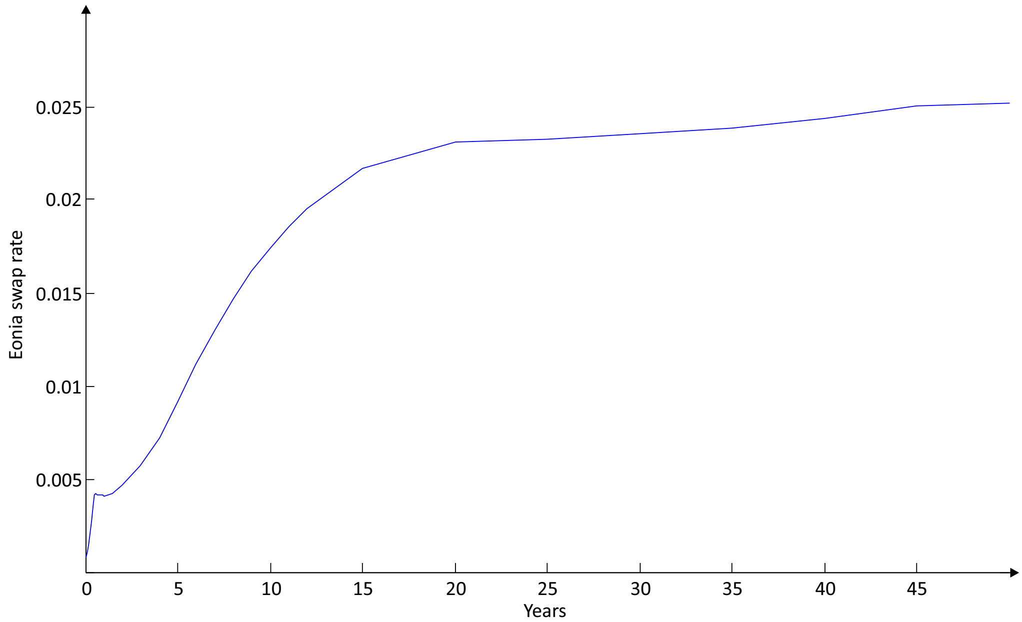

Figure 9.4: Euro overnight indexed swap curve, data from Bloomberg (curve S122)

κ_r	3.178	σ	0.084
θ_r	0.005	μ	0.063
σ_r	0.007	ρ	-0.038
$\tilde{\theta}_r$	0.016	r_0	0.001

Table 9.1: Parameters for the Black-Scholes-Vasicek model

κ_r	2.148	V_0	0.009	κ	0.310
θ_r	0.004	μ	0.031	θ	0.001
σ_r	0.085	ρ	0.969	σ	0.027
$\tilde{\theta}_r$	0.016	r_0	0.001	$\tilde{\theta}$	0.023

Table 9.2: Parameters for the Heston-Cox-Ingersoll-Ross model

For the Heston-Cox-Ingersoll-Ross model, we replicate this strategy. The only substantial difference is the estimation of the real-world parameters for the asset process. Here, direct maximum-likelihood estimation is not possible because the volatility is an unobserved parameter. We implement a dual particle filter for the joint estimation of the volatility state and the parameters as in Olivier, Huang and Craig (2012), using results from Aihara, Bagchi and Saha (2009) on particle filtering for the Heston model. The results are exhibited in table 9.2.

9.4 Empirical Study

9.4.1 Model Comparison

In this section, we compare the performance of algorithms 1-3, that is, nested simulation (NS), least-squares Monte-Carlo (LSM), and least-squares Monte-Carlo with risk-neutral paths (LSMQ) using the asset and liability model from section 9.1 and the Black-Scholes-Vasicek (BSV) and Heston-Cox-Ingersoll-Ross (HCIR) models from section 9.2. Following Bauer et al. (2011*a*), we use the policyholder perspective throughout this section, as Bauer et al. (2011*a*) observe that this leads to better results (we are able to confirm this observation in section 9.4.2). The regressions in the LSM and LSMQ algorithms are conducted with their full model (see Bauer et al., 2011*a*, tab. 2). In this section, we aim for small errors by generating large numbers of paths. We set $N_0 = N_1 = 1.5MM$, $K_1 = 400$.

The computation times are exhibited in table 9.3. Even in our simple model, the NS takes three quarters of an hour with the BSV and more than two and a half hours with the HCIR asset model, a result that suggests that this algorithm is not feasible for application in practice. Compared to 29.2 and 41.8 seconds, respectively, for the LSM with the two asset models, the NS needs about 90 and 200 times as much computation time, respectively. The difference in these factors can be explained by the more complex simulation scheme for the HCIR model, which takes more time relative to the calculations for the asset-liability model and overhead. This also suggests that the LSM scales better to advanced models than nested simulations. As expected, the LSMQ is the fastest of the three algorithms, saving about 35% of computation time compared to the LSM with each model. This improvement is sufficient for the algorithm to be faster with the HCIR asset model than the LSM algorithm is with the much simpler BSV model.

	BSV	HCIR
LSM	29.2s	41.8s
LSMQ	19.6s	27.0s
NS	2700.5s	9466.3s

Table 9.3: Computation times for the LSM, LSMQ, and NS algorithms using the Black-Scholes-Vasicek (BSV) and Heston-Cox-Ingersoll-Ross (HCIR) models; MATLAB2012b on a PC with Core2Quad Q8400, 4GB RAM; $N_0 = N_1 = 1.5MM$, $K_1 = 400$

Next, we consider the distance of the resulting distributions as measured by the 2-Wasserstein, L^2, and Kolmogorov distances introduced in chapter 4. The distances for the BSV and HCIR models are displayed in table 9.4 and table 9.5, respectively. We use the distributions generated using NS as the references. The Kolmogorov distance is small across the board and does not help us in distinguishing between the algorithms. The distance of the LSMQ to the NS is larger than the distance of the LSM to NS in the case of the BSV model, but it is the other way around in for he HCIR model. Visual inspection of the empirical cdfs—we show two examples in figure 9.5 and figure 9.6—suggests that the Wasserstein and L^2 distances are small, too. To give a comparison, the Wasserstein distances obtained here would also be the result of shifting the distribution mean by less than 0.5% of the distribution's standard deviation.

Algorithm		d_2	L_2	K
LSM	NS	3.08	0.06	0.01
LSMQ	NS	5.32	0.11	0.01
LSM	LSMQ	3.56	0.06	0.00

Table 9.4: Distances between the distributions computed by the LSM, LSMQ, and NS algorithms in the Black-Scholes-Vasicek model

The comparison of the cdfs across asset models—figure 9.7 shows the results for the NS—suggests larger distances between the distributions. This impression is confirmed by the data in table 9.6. We investigate and explain this difference in section 9.4.3. However, the algorithms perform similarly.

Algorithm		d_2	L_2	K
LSM	NS	4.69	0.08	0.01
LSMQ	NS	3.36	0.06	0.01
LSM	LSMQ	3.11	0.03	0.00

Table 9.5: Distances between the distributions computed by the LSM, LSMQ, and NS algorithms in the Heston-Cox-Ingersoll-Ross model

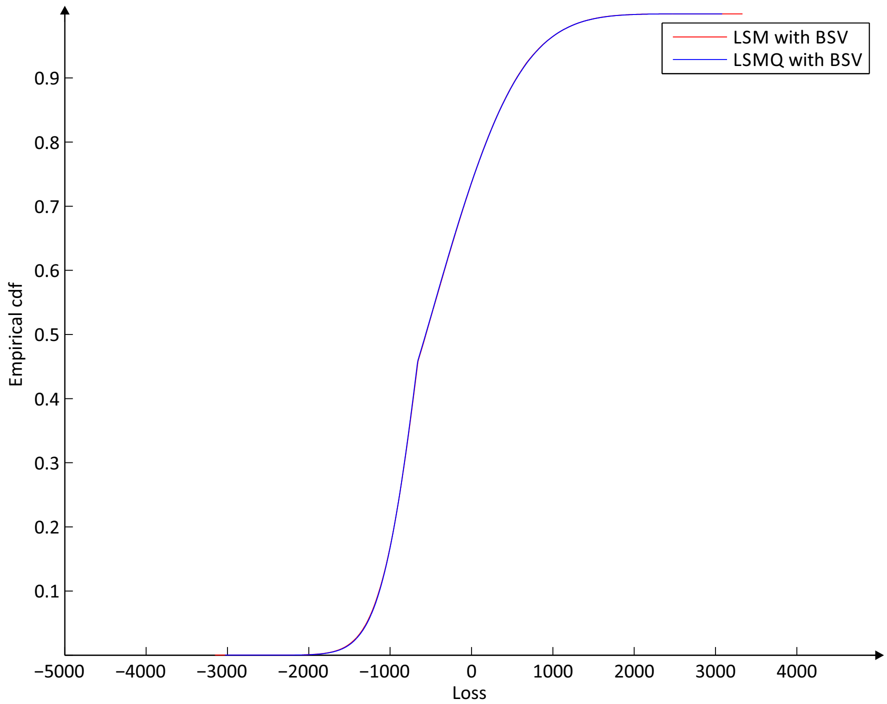

Figure 9.5: Distribution functions of the AC at $t = 1$ in the BSV model, computed with the LSM and LSMQ algorithms

Algorithm	Model		d_2	L_2	K
NS	BSV	HCIR	503.08	10.01	0.28
LSM	BSV	HCIR	503.14	10.01	0.28
LSMQ	BSV	HCIR	504.87	10.05	0.28

Table 9.6: Distances between the distributions computed computed in the BSV and HCIR models by the LSM, LSMQ, and NS algorithms

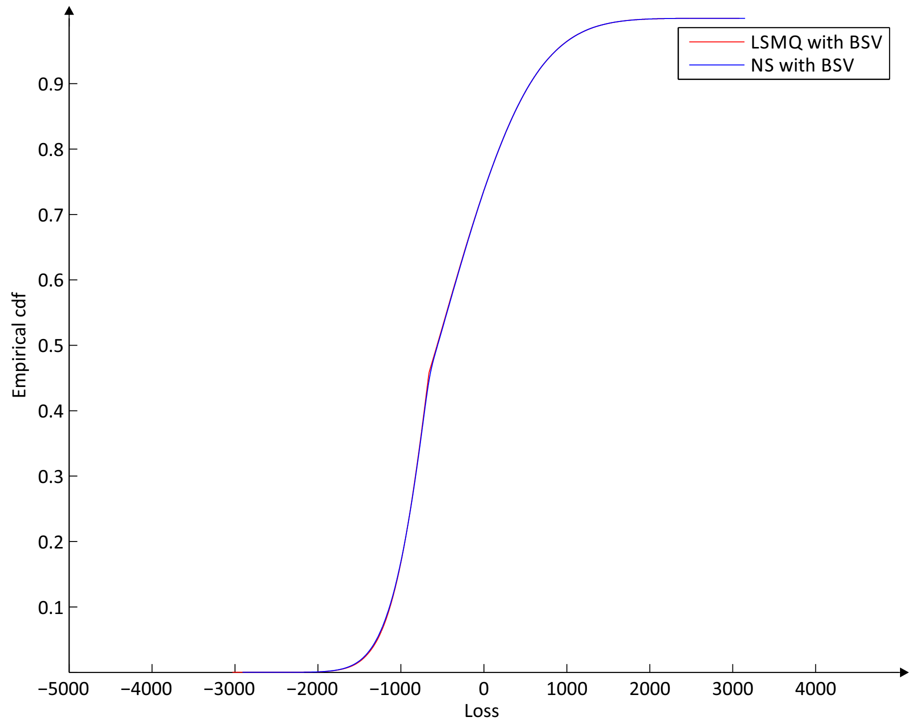

Figure 9.6: Distribution functions of the AC at $t = 1$ in the BSV model, computed with the LSM and NS algorithms

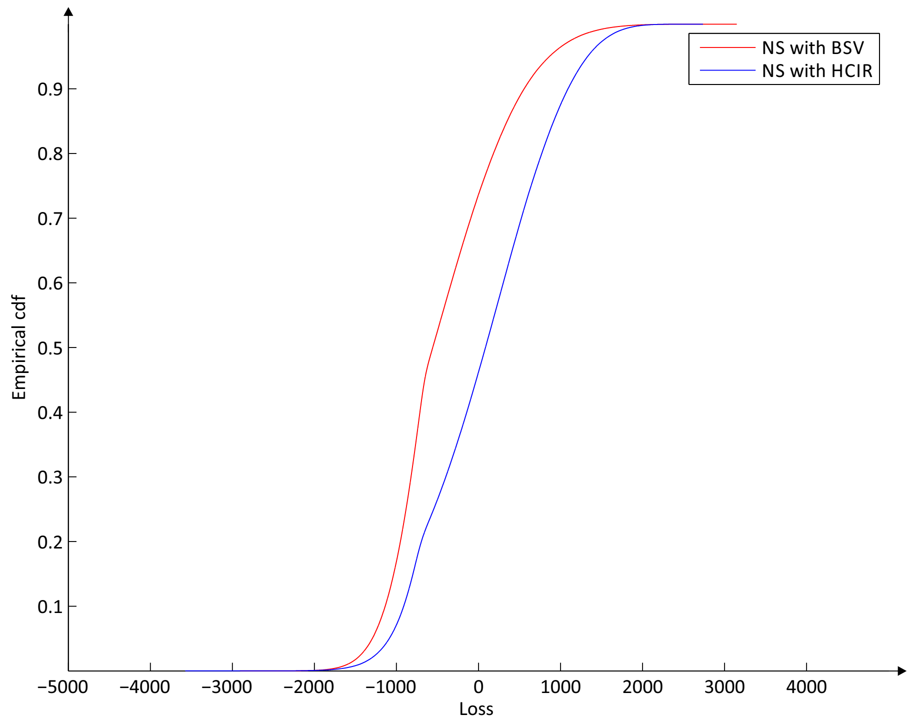

Figure 9.7: Distribution functions of the AC at $t = 1$ in the BSV and HCIR models, computed with the NS algorithm

Finally, we consider the value at risk and the expected shortfall computed from the simulations (table 9.7 and table 9.8). The results from the three algorithms within the models are quite close, with the largest relative difference less than 0.5%. While larger Wasserstein and L^2 distances suggest larger deviations in the risk measures, the results are too close together in this case to empirically confirm our argument that the Wasserstein distance is better indicator of deviations in tail-based risk measures.

Model / Algorithm	$\mathrm{VaR}_{0.95}$	$\mathrm{VaR}_{0.9}$	$\mathrm{VaR}_{0.995}$
BSV			
LSM	862.03	1425.64	1618.27
LSMQ	865.09	1429.14	1630.06
NS	865.45	1430.30	1626.73
HCIR			
LSM	1328.48	1722.15	1852.52
LSMQ	1327.39	1720.89	1846.09
NS	1330.71	1724.49	1852.41

Table 9.7: Value at risk

Model / Algorithm	$\mathrm{ES}_{0.95}$	$\mathrm{ES}_{0.9}$	$\mathrm{ES}_{0.995}$
BSV			
LSM	1206.24	1684.17	1855.14
LSMQ	1208.61	1693.55	1866.99
NS	1209.19	1693.33	1868.27
HCIR			
LSM	1568.95	1889.53	1997.86
LSMQ	1566.68	1885.79	1993.65
NS	1570.92	1889.77	1996.94

Table 9.8: Expected shortfall

9.4.2 Validation for the LSM and LSMQ Methods

In section 9.4.1 we use large numbers of paths and, consequently, obtain rather stable results for all algorithms under consideration. In practice, we would probably be restricted to much fewer paths (maybe 1000 instead of 1.5MM paths) due to restrictions of memory and computation time. Such a restriction essentially rules out the NS algorithm, as configurations such as 400 outer paths and 3 inner paths each cannot be near sufficient to obtain a stable estimate of a 99.5% quantile like the SCR. We investigate how many paths are necessary to obtain reasonably stable results with the LSM and LSMQ algorithms and whether we can find any clues regarding robustness by applying regression diagnostics. Such a test is required under the Solvency II regulations, see section 3.2.1.

We start by considering single runs of the LSM and LSMQ algorithms, each for $N_0 = N_1 = \{1000, 10000, 100000\}$. A crucial part in both algorithms is the estimation of the function h_{V_1} by least-squares regression. Table 9.9 and table 9.10 show regression diagnostics[4] for the exemplary runs of the algorithms, both for calculations from the shareholders' (SH) and policyholders' (PH) point-of-view. Recall that the reference values for the SCR obtained from the NS are 1626.73 for the BSV model and 1852.41 for the HCIR model.

The first statistic we consider is the F-statistic for goodness of fit of the linear model (see Weisberg, 1980, p. 48). Under the test-assumptions of independent normal residuals with zero mean, the critical value for the test at the 99% level is between 3 and 4 for all regression models used in this section which would indicate rejection of the H_0-hypothesis that the regression coefficients are jointly equal to zero. However, we do not report p-values, as the residuals in all models turn out to be far from normally distributed. As the F-statistic is increasing with the number of observations, we cannot deduce anything about the number of simulation paths necessary to obtain good results. There are notable differences between the F-statistics of the models from the shareholders' and policyholders' point-of-view, though. The values for the latter are about two to five times higher than for the former, all else being equal, indicating a better model fit for the PH calculations.

The adjusted R^2 (adjusted for the number of explanatory variables in the model, see Weisberg (1980, p. 188)) is generally quite low, which is not unusual in least-squares Monte-Carlo applications. Again, the values for the PH perspective are about two to five times higher than for the SH perspective.

Next, we examine whether single paths have a large influence on regression results. The leverage, given by the diagonal elements of the projection matrix (also called hat matrix), gives an indication. Huber and Ronchetti (2011, p. 160) state that values up to 0.2 are safe, values between 0.2 and 0.5 risky, and that values above 0.5 should be avoided. We see that the maximum leverage for regressions based on BSV paths falls in the lower part of the "risky" interval, while it is above 0.5 for all regressions based on HCIR paths, with improvement only for $N_0 = N_1 = 100000$.

The discussion of the regression diagnostics suggests that the regression works better

[4] Some additional results that are not needed for our discussion are relegated to Appendix D.

Model / No. paths	Perspective	F-statistic	Adj. R^2	Leverage	SCR
BSV					
1000	SH	5.14	0.04	0.24	1891.48
	PH	26.38	0.19		1748.94
10000	SH	48.81	0.04	0.24	1634.77
	PH	315.52	0.22		1664.01
100000	SH	497.51	0.04	0.25	1598.31
	PH	3065.72	0.22		1624.79
HCIR					
1000	SH	3.97	0.03	0.98	1871.42
	PH	6.46	0.05		1841.89
10000	SH	41.42	0.04	0.88	1983.73
	PH	92.87	0.08		1872.89
100000	SH	385.15	0.03	0.53	1862.94
	PH	798.32	0.07		1845.60

Table 9.9: Regression diagnostics for LSM algorithm, shareholder (SH) and policyholder (PH) perspective; number of paths $N_0 = N_1$

Model / No. paths	Perspective	F-statistic	Adj. R^2	Leverage	SCR
BSV					
1000	SH	4.88	0.03	0.25	968.91
	PH	14.58	0.11		1697.25
10000	SH	62.02	0.05	0.24	1362.47
	PH	157.70	0.12		1682.80
100000	SH	570.44	0.05	0.24	1554.96
	PH	1451.47	0.12		1618.17
HCIR					
1000	SH	6.55	0.05	0.98	1761.41
	PH	11.55	0.09		1670.80
10000	SH	49.00	0.04	0.88	1926.74
	PH	142.78	0.11		1852.12
100000	SH	430.83	0.04	0.53	1894.52
	PH	1264.00	0.10		1865.85

Table 9.10: Regression diagnostics for LSMQ algorithm, shareholder (SH) and policyholder (PH) perspective; number of paths $N_0 = N_1$

	SH		PH	
Model / $N_0 = N_1$	Mean SCR	STD SCR	Mean SCR	STD SCR
BSV				
1000	1690.04	294.29	1586.97	130.06
10000	1622.62	91.23	1619.81	44.24
100000	1619.66	28.55	1622.67	13.21
CIR				
1000	1892.79	247.61	1842.28	102.33
10000	1849.04	68.06	1845.99	32.63
100000	1849.09	21.23	1847.58	10.17

Table 9.11: Mean and standard deviation of the SCR computed from 500 simulation runs with the LSM algorithm

for the PH perspective than for the SH perspective and possibly better in the BSV model than in the HCIR model. To get some further information on the stability, we run the model 500 times in each configuration. Table 9.11 and table 9.12 show means and standard deviations for both asset models, from SH and PH perspective for the LSM and the LSMQ method, respectively.

The means for $N_0 = N_1 > 1000$ are within about 0.5% of the previously computed values (cf. table 9.7), but for $N_0 = N_1 = 1000$ the deviations are larger, ranging from about 1% (HCIR,PH) to 34% (BSV,SH). The means obtained from PH calculations are closer in the latter case, while there is no recognizable pattern in the former case. The picture is clearer when we consider the standard deviations: computations from the PH perspective have only between about 20% to 50% of the corresponding value for SH perspective computations. This confirms the conjecture we formed from the regression diagnostics and the observations by Bauer et al. (2011*a*).

The results for the two asset models do not agree with what we expected from the regression diagnostics: the simulations based on the HCIR model are more stable than those based on the BSV model, with standard deviations being about 15% to 20% lower. The somewhat unstable regressions do not seem to have an overriding influence on the resulting distributions.

The standard deviation in relation to the number of paths turns out to be close to what can be derived from the theory on Monte-Carlo methods: increasing the number of paths by a factor of n decreases the standard deviation of the estimator by about a factor of $\sqrt{n}$.

	SH		PH	
Model / $N_0 = N_1$	Mean SCR	STD SCR	Mean SCR	STD SCR
BSV				
1000	2072.68	1382.33	1602.36	257.51
10000	1612.94	264.20	1619.49	76.76
100000	1624.26	87.61	1621.23	24.50
CIR				
1000	1901.39	542.18	1828.01	200.03
10000	1845.52	172.47	1846.32	64.93
100000	1850.21	55.31	1847.67	20.08

Table 9.12: Mean and standard deviation of the SCR computed from 500 simulation runs with the LSMQ algorithm

Parameter	% change	SCR	SCR % change
μ	+5.00	1595.90	−1.70
μ	−5.00	1649.41	1.60
σ	+5.00	1710.74	5.38
σ	−5.00	1534.13	−5.50
$\hat{\theta}_r$	+5.00	1627.26	0.24
$\hat{\theta}_r$	−5.00	1623.54	0.01

Table 9.13: Sensitivity analysis for the BSV model (SCR with standard parameters: 1623.43)

9.4.3 Sensitivity Analysis

A sensitivity analysis to changes of model parameters is required for the robustness check under the Solvency II for internal model approval, see section 3.2.1. We perform such an analysis for exemplary parameters for both the BSV and the HCIR model.

We start with the BSV model, examining the influence on the SCR of changes in the expected asset return μ, the asset volatility σ, and the long-term mean of the short-rate under the risk-neutral measure $\hat{\theta}_r$. We increase and decrease each parameter by 5% of its respective value. The results, computed with $N_0 = N_1 = 1.5MM$ paths in the LSM algorithm, are shown in table 9.13. From our results in section 9.4.2, we can infer a standard deviation of about 3.41 for the SCR computations. In relation to this, the changes in $\hat{\theta}_r$ barely have an influence on the final result. This should not be surprising, though, as the interest-rate is used as the log-asset drift in the risk-neutral model, and payoffs are discounted based on the same rate. The SCR changes induced by altering the asset volatility have roughly the same size in relative terms as the parameter change. In particular, increasing the asset volatility increases the SCR. This can be explained as

Parameter	% change	SCR	SCR % change
θ	+5.00	1856.34	0.49
θ	−5.00	1845.45	−0.10
$\hat{\theta}_r$	+5.00	1843.04	−0.23
$\hat{\theta}_r$	−5.00	1843.62	−0.20
σ_r	+5.00	1829.55	−0.96
σ_r	−5.00	1868.04	1.12

Table 9.14: Sensitivity analysis for the HCIR model (SCR with standard parameters: 1847.26)

follows: As the guarantee in the contract that is modeled in the asset-liability framework is essentially a strip of put options (with strike set one year before maturity) sold to the policyholders, and put options have a positive vega, the value of the guarantees to the policyholders increases with σ. The real-world drift μ is only used for the generation of paths over the first year of the time horizon. Increasing μ relative to the short-rate increases the probability of positive asset returns in the model and therefore lowers the SCR. This is what we can see in the results here. While the magnitude of the influence is smaller than that of the asset volatility, the difference observed here is still significantly larger than the Monte-Carlo noise at any standard confidence level[5]. This observation helps us to explain the difference in SCR between the BSV and the HCIR models. The no-arbitrage argument we used in section 9.2 leads to real-world asset drifts that are close to the short-rate ($\sim 0.1\%$) in the HCIR model, while the real world asset drift in the BSV model is estimated to be 6.3%. Indeed, using $\mu = r_0\%$ in the BSV model, we obtain an SCR estimate of 2186.29, confirming the significant influence the parameter has on the result. In practice, we could discuss whether and how we can change the structure of the drift term in the HCIR model to get consistent results.

We repeat the exercise for the HCIR model, this time considering the influence of the long-term mean θ of the volatility process, the risk-neutral long-term mean $\hat{\theta}_r$ of the short-rate, and the short-rate volatility σ_r. The other parts of the setup remain the same: we use $N_0 = N_1 = 1.5MM$ paths in the LSM algorithm and permute each parameter by plus and minus 5% of its respective value. The results are shown in table 9.14. For both θ and $\hat{\theta}_r$ the changes in the SCR we observe are smaller than the Monte-Carlo error. In the case of θ, the direction of the changes can be explained using the same line of arguments as for

[5]Based on the normal error distribution for Monte Carlo simulations (see Glasserman, 2004, p. 2), the difference is more than seven standard deviations away from zero.

the volatility in the BSV model, while the Monte-Carlo error seems to be the overriding influence for $\hat{\theta}_r$, as it was the case in the BSV model.[6] The influence of the short-rate volatility is more pronounced: altering it by $\pm 5\%$ leads to a change of the SCR of about 1% in the opposite direction. Increasing σ_r leads to higher probabilities of higher interest rates, and as the short-rate enters into the drift from $t = 0$ to $t = 1$, this explains—in analogy to the BSV model—why the SCR decreases.

[6]This impression is confirmed be rerunning the simulations.

Chapter 10

Conclusions

In this thesis, we investigated how to improve the robustness of quantitative risk management procedures, especially in the light of the Solvency II regulations for the insurance industry. The importance of this task was recently highlighted during the financial crisis. Our starting points were the definitions of risk and risk management, and the Solvency II framework. We sought to identify important elements of the risk management process and their required properties—confirming that robustness is one of them—and to combine these elements into a structure that facilitates a coherent risk management procedure. Additionally, we tried to show how the Solvency capital requirements can be computed with a simulation algorithm that circumvents the practical problem of a change of measure within the simulation paths.

In the first part, the distinction between risk and uncertainty (see chapter 2) provided a first impulse to consider robustness of statistical methods used in quantitative risk management—methods that work under uncertainty about models and their parameters. Added to that are the regulatory requirements in Solvency II regulations (see chapter 3, and in particular section 3.2.1) which explicitly require robustness of the statistical and probabilistic methods used on several occasions. The reasoning behind this turned out to be practical considerations, as alluded to in the introduction: an insurance company's management should be able to understand and base decisions on the results of risk management activities, and this is greatly facilitated by the use of robust methods.

Our approach to robustness, laid out in the second part, was based on the idea that any probabilistic model that we use can only be an approximation. This necessitated a facility which can measure the quality of such an approximation. In our probabilistic setup, probability distances are the solution (see chapter 4). Among the possible choices, the Wasserstein metric turned out to be a good candidate for risk management: it focuses on differences in the tails of distributions—this is the important area of a loss or value distribution for risk management—its properties are well understood and it has a simple formula on the real line, making interpretations and communication of results easier. Probability distances in general have the advantage that they are widely used in classical robustness literature in statistics (see chapter 5). In particular, we were able to apply these classic results to risk measures, which are central to quantitative risk management (see chapter 6). Consequently, we collected the elements necessary for robust quantitative risk management, united them in a structure we called "risk management procedure" and illustrated its application in two examples.

In the third part, we were concerned with the computation of the Solvency capital requirement, specifically by simulation algorithms. Following up on a proposal from practice, we showed that the least-squares algorithm proposed by Bauer et al. (2011*a*) can be modified so that a change of measure is not necessary within the simulation paths. We proved its convergence theoretically and investigated its performance empirically using a simple asset-liability model of a life-insurance company to compare it to the standard least-squares and the nested simulation algorithms. The empirical study included model validation exercises required by Solvency II regulations. The results of the new algorithm were comparable to those of the unmodified least-squares algorithm, and it could be serviceable if the change of measure is a serious problem in practice.

Our studies suggest areas for further research. While the robustness properties of the most widely used risk measures, value at risk and expected shortfall, are known, it would be an improvement to devise a risk measure that is designed with this aspect in mind while still fulfilling some sensible axiomatic system. Additionally, in order to employ the Wasserstein metric in practice, calibration methods based on this metric would be needed for more than the most basic models. In particular, it could potentially be used to find replicating portfolios, which are another promising method for the calculation of the Solvency capital requirement.

Robustness is an important property of any quantitative risk management process for theoretical, regulatory, and especially practical reasons. In this thesis, we showed how this aspect can be integrated into a risk management procedure that—if applied in practice—could lead to more reliable results that are easier to interpret by stakeholders, especially in the complex context of rule-based Solvency II regulations.

Appendix A

Geometry of the Space $(\mathcal{M}^2, d_2)$

We consider the metric space $(\mathcal{M}^2, d_2)$ of probability distributions on the Borel σ-field $\mathcal{B}(X)$ of some separable metric space (X, d).

Remark A.0.1
The space $(\mathcal{M}^2, d_2)$ is a metric space but not a vector space. Consequently, classical results such as the Krein-Milman theorem (see theorem B.2.4) cannot be applied.

Proposition A.0.2
If (X, d) is a separable and complete metric space and $p > 0$, then the metric space $(\mathcal{M}^p, d_p)$ is separable and complete. Additionally, $K \subset \mathcal{M}^p(X)$ is relatively compact w.r.t. the topology induced by d_p if and only if it is tight and p-uniformly integrable.

Proof. See Bolley (2008) and Ambrosio and Gigli (2010). □

In the following we restrict ourselves to $X = \mathbb{R}^n$ for some $n \in \mathbb{N}$.

Proposition A.0.3 (Gigli (2004, Prop. 2.5))
A subset of $\mathcal{M}^2(\mathbb{R}^n)$ is relatively compact if and only if it is 2-uniformly integrable.

Definition A.0.4 (Length Space, (see Papadopoulos, 2005, def. 2.1.2))
Let (X, d) a metric space. It is called a *length space* if for any $x, y \in X$ we have

$$d(x, y) = \inf_{\gamma} L(\gamma)$$

where $L(\gamma)$ denotes the length of the path γ and the infimum is taken over all paths connecting x and y.

Definition A.0.5 (Geodesics, see Ambrosio and Gigli (2010))
A curve $\gamma : [0,1] \to X$ is called a *constant speed geodesic* if

$$d(\gamma_t, \gamma_s) = |t-s|\, d(\gamma_0, \gamma_1) \qquad \forall\, t, s \in [0,1].$$

A metric space (X, d) is called *geodesic* if for every $x, y \in X$ there exists a constant speed geodesic connecting them.

Example A.0.6
The Euclidean spaces $(\mathbb{R}^d, \|.\|)$ with the distance induced by the norm $\|.\|$ are geodesic. Let $x, y \in \mathbb{R}^d$. Then

$$\gamma_t := x + t(y-x), t \in [0,1]$$

with

$$\begin{aligned} \|\gamma_t - \gamma_s\| &= \|x - x + t(y-x) - s(y-x)\| \\ &= \|(t-s)(y-x)\| \\ &= |t-s|\, \|y-x\| = |t-s|\, \|\gamma_0 - \gamma_1\| \end{aligned}$$

is a constant speed geodesic connecting x and y.

Proposition A.0.7 (see Papadopoulos (2005, Prop. 2.4.2))
A geodesic space is a length space.

Theorem A.0.8
Let (X, d) be Polish and geodesic. Then $(\mathcal{M}^2(X), d_2)$ is geodesic and, in particular, $(\mathcal{M}^2(\mathbb{R}^n), d_2)$ is geodesic.

Proof. See Ambrosio and Gigli (2010, Th. 2.10). □

Proposition A.0.9 (see Ambrosio and Gigli (2010, Prop. 2.26))
Let $(\mu_t) \subset \mathcal{M}^2(\mathbb{R}^d)$ be a geodesic and assume that μ_0 is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^d$. Then, μ_t is absolutely continuous w.r.t. the Lebesgue measure and the set of absolutely continuous measures is geodesically convex.

Let $F_1, F_2 \in \mathcal{M}^2(\mathbb{R})$ be two distributions on the real line. Then for any geodesic $\gamma^{1\to 2}$ connecting F_1 and F_2 induced by a distribution γ for which the infimum in equation (4.8) is attained, we have

$$F^{-1}_{\gamma_t^{1\to 2}} = (1-t)F^{-1}_{F_1} + tF^{-1}_{F^2}$$

a.e. w.r.t. the Lebesgue measure on the real line, see Ambrosio, Gigli and Savaré (2008).

Definition A.0.10 (Interpolation, see Gigli (2004, def. 2.6))
Let $F, G \in \mathcal{M}^2(\mathbb{R}^n)$, $\gamma \in \tau(F, G)$. The *interpolation* from F to G is the curve $[\gamma](t) : [0.1] \to \mathcal{M}^2(\mathbb{R}^n)$ defined as

$$[\gamma](t) := \big((1-t)\pi^1 + t\pi^2\big)_{\#}\,\gamma$$

where π^i denotes the projection operator to the i-th component and $\nu_{\#}\mu$ denotes the push forward of μ through ν, see definition B.1.3.

Proposition A.0.11 ($(\mathcal{M}^2(\mathbb{R}^n), d_2)$ is positively curved, see Gigli (2004, Prop. 2.13))
Let $F_0, F_1, G \in \mathcal{M}^2(\mathbb{R}^d)$ and assume that the minimum in equation (4.8) is attained for $\gamma \in \tau(F_0, F_1)$. Then

$$d_2(G, [\gamma](t)) \geq (1-t)d_2(G, F_0) + td_2(G, F_1) - t(1-t)d_2(F_0, F_1), \tag{A.1}$$

that is, $(\mathcal{M}^2(\mathbb{R}^n), d_2)$ is positively curved in the sense of Alexandrow (1957).

We follow Ambrosio et al. (2008, Ch. 12.3) and define a triangle $\Delta \in \mathcal{M}^2(\mathbb{R}^d)$ to be the image of three geodesics $\mu^{1\to 2}, \mu^{2\to 3}, \mu^{3\to 1}$ connecting three points μ^1, μ^2, $\mu^3 \in \mathcal{M}^2(\mathbb{R}^d)$. A reference triangle $\hat{\Delta}$ in $\mathbb{R}^2$ is given by the image of the geodesics $x^{1\to 2}, x^{2\to 3}, x^{3\to 1}$ connecting some points x^1,x^2,x^3 in $\mathbb{R}^2$ where $d(\mu^i, \mu^j) = \left\|x^i - x^j\right\|, i, j = 1, 2, 3$ (unique up to isometric transformation). Positive curvature means that for any points $\nu_1, \nu_2 \in \Delta$ and $y_1, y_2 \in \hat{\Delta}$, we have $d(\nu_1, \nu_2) \geq \|y_1 - y_2\|$.

Appendix B

Measure Theory and Probability Theory

B.1 Definitions

Definition B.1.1 (Weak Convergence of Measures, see Billingsley (1999, p. 2))
Let (M, d) be a metric space. Let $P, P_n, n \in \mathbb{N}$ be probability measures on the Borel σ-field $\mathcal{B}(M)$ of M. The sequence P_n *converges weakly* to P if the probability measures satisfy $\int_S f dP_n \to \int_S f dP$ for any bounded, continuous real-valued function f on S. We write $P_n \Rightarrow P$.

Definition B.1.2 (Continuity Sets, see Billingsley (1999, p. 2))
Let (M, d) be a metric space and let P be a probability measure on M. A set $A \in \mathcal{B}(M)$ is called a *P-continuity set* if its boundary ∂A has measure zero, that is, $P(\partial A) = 0$.

Definition B.1.3 (Push Forward, see, e.g., Sturm (2006, sect. 2.3))
Let (X, σ_X) and (Y, σ_Y) be two measurable spaces and let $f : X \to Y$ be a measurable function. If μ is a measure on (X, σ_X) then the measure

$$f_{\#}\mu(B) := \mu(f^{-1}(B)), \text{ for any } B \in \sigma(Y)$$

is called the *push forward of μ through f*.

Definition B.1.4 (2-uniform integrability, 2-boundedness, see Gigli (2004))
Let X be some separable metric space with Borel σ-field $\mathcal{B}(X)$. Denote by

$$\Gamma_2(X) := \left\{ \mu \middle| \mu \text{ a probability measure on } \mathcal{B}(X), \int_X \|x\|^2 \, d\mu < \infty \right\}$$

Some subset $A \subset \Gamma_2(X)$ is called 2-uniformly integrable if for every $\epsilon > 0$ there exists K_ϵ such that

$$\int_{X \setminus K_\epsilon} \|x\|^2 \, d\mu < \epsilon \forall \mu \in A.$$

A set $A \subset \Gamma_2(X)$ is called *2-bounded* if

$$\sup_{\mu \in A} \int \|x\|^2 \, d\mu < \infty.$$

B.2 Results from Measure Theory

Theorem B.2.1 (Disintegration Theorem, see Ambrosio et al. (2008, Th. 5.3.1))
Let X, Y be Radon separable metric spaces, $\mu \in \Gamma(X)$, let $\pi : X \to Y$ be a Borel map and let $\nu = \pi_\# \mu \in \Gamma(Y)$. Then there exists a ν-a.e. unique family of Borel measures $\{\mu_y\}_{y \in Y} \subset \Gamma(X)$ such that

$$\mu_y(X \setminus pi^{-1}(y)) = 0 \ \nu\text{-a.e. } y \in Y$$

and

$$\int_X f(x) d\mu(x) = \int_Y \left(\int_{\pi^{-1}(y)} f(x) d\mu_y(x) \right) d\nu(y)$$

for every Borel map $f : X \to [0, +\infty]$.

Theorem B.2.2 (The Portmanteau Theorem)
Let (M, d) be a metric space. Let $P, P_n, n \in \mathbb{N}$ be probability measures on the Borel σ-field $\mathcal{B}(M)$ of M. The following conditions are equivalent.

(i) $P_n \Rightarrow P$.

(ii) $\int f dP_n \to \int f dP$ for all bounded, uniformly continuous functions f.

(iii) $\limsup_n P_n(A) \leq P(A)$ for all closed sets $A \in \mathcal{B}(M)$.

(iv) $\liminf_n P_n(A) \geq P(A)$ for all open sets $A \in \mathcal{B}(M)$.

(v) $P_n(A) \to P(A)$ for all P-continuity sets A.

Proof. See Billingsley (1999, theorem 2.1). □

Theorem B.2.3 (Ulam's theorem)
On any complete separable metric space (M, d) every finite Borel measure is tight and therefore regular.

Proof. See Dudley (2002). □

Theorem B.2.4 (Krein-Milman Theorem (Royden, 1968, thm. 26))
Let K be a compact convex set in a locally convex topological vector space X. Then K is the closed hull of its extreme points.

Proof. The theorem was originally proved by Krein and Milman (1940). □

B.3 Stochastic Dominance

Definition B.3.1 (see Rachev et al. (2008))
A random variable X dominates a random variable Y in the sense of

- *first order stochastic dominance*,

$$X \succeq_{FSD} Y \text{ iff } F_X(x) \leq F_Y(x) \forall x \in \mathbb{R};$$

- *second order stochastic dominance*,

$$X \succeq_{SSD} Y \text{ iff } \int_{-\infty}^{x} F_X(t)dt \leq \int_{-\infty}^{x} F_Y(t)dt, \forall x \in \mathbb{R}.$$

Corollary B.3.2
We have $X \succeq_{SSD} Y \Rightarrow X \succeq_{FSD} Y$ directly from the definitions.

Appendix C

Convex Analysis

In this appendix, we collect some facts on convex analysis that are needed to derive the robustified version of convex risk measures in section 6.1.2 from Rockafellar (1997, sect. 1,2,6).

Definition C.0.3 (Affine and Convex Sets)
Let $\langle ., . \rangle$ denote the inner product on $\mathbb{R}^n$.

- A set $A \subseteq \mathbb{R}^n$ is called *affine* if $(1-\lambda)x + \lambda y \in A \forall x, y \in A, \lambda \in \mathbb{R}$.
- For any set $S \subseteq \mathbb{R}^n$, its *affine hull* is defined as

$$\operatorname{aff} S = \bigcap_{A \supset S, A \text{ affine}} A$$

- A subset $C \subset \mathbb{R}^n$ is called *convex* if $(1-\lambda)x + \lambda y \in C$ for all $\lambda \in (0,1), x, y \in C$.
- The *relative interior* of a convex set C is the interior of C as a subset of its affine hull,

$$\operatorname{ri} C = \{x \in \operatorname{aff} C | \exists \epsilon > 0, (x + \epsilon B) \cap (\operatorname{aff} C) \subset C\}.$$

- A half-space in $\mathbb{R}^n$ is characterized by a hyperplane

$$H = \{x | \langle x, b \rangle = \beta\}, b \in \mathbb{R}^n \setminus \{0\}, \beta \in mR.$$

There are the *closed half-spaces*

$$\{x|\langle x,b\rangle \leq \beta\} \quad \text{and} \quad \{x|\langle x,b\rangle \geq \beta\}$$

and the *open half-spaces*

$$\{x|\langle x,b\rangle < \beta\} \quad \text{and} \quad \{x|\langle x,b\rangle > \beta\}\,.$$

Half-spaces are convex sets.

Definition C.0.4 (Convex Functions (Rockafellar, 1997, sect. 4 & 7))
Let $f : S \to \overline{\mathbb{R}}$ be a function on the domain $S \subseteq \mathbb{R}^n$. Its *epigraph* is defined as

$$\operatorname{epi} f := \{(x,\mu)|x \in S, \mu \in \mathbb{R}, \mu \geq f(x)\}\,.$$

- Such a function f is called *convex* if its epigraph $\operatorname{epi} f$ is a convex set.
- The projection of the epigraph on $\mathbb{R}^n$ is called the *effective domain* of f:

$$\operatorname{dom} f := \{x|f(x) < \infty\}\,.$$

- A convex function f is called *proper* if

$$\exists x \in \mathbb{R}^n : f(x) < +\infty \text{ and } \forall x \in \mathbb{R}^n : f(x) > -\infty,$$

otherwise it is called *improper*.

- A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *lower semi-continuous* if

$$f(x) = \liminf_{y \to x} f(y).$$

- The *lower semi-continuous hull* of a convex function f is given by the function $\overline{f}$ whose epigraph equals the closure of the epigraph of f:

$$\operatorname{epi} \overline{f} = \operatorname{cl} \operatorname{epi} f.$$

- The *closure* of a convex function f is defined as

$$\operatorname{cl} f(x) := \begin{cases} \overline{f}(x); & f(x) > -\infty \forall x \in \mathbb{R}^n \\ -\infty; & else. \end{cases}$$

  A convex function is *closed* if $\operatorname{cl} f = f$. For proper convex functions, lower semi-continuity and closedness are actually the same (see Rockafellar, 1997, sect. 7).

The dual characterization of convex risk measures is based the following theorem.

Theorem C.0.5 (Rockafellar (1997, thm. 12.1))
A closed convex function f is the pointwise supremum of the collection of all affine functions h such that $h \leq f$.

The affine functions from $\mathbb{R}^n$ to $\mathbb{R}$ are characterized by points $(x^*, \mu^*), x^* \in \mathbb{R}^n, \mu^* \in \mathbb{R}$ and therefore $h(x) = \langle x, x^* \rangle - \mu^* \leq f(x) \forall x \in \mathbb{R}^n$ if and only if

$$\mu^* \geq \sup \{\langle x, x^* \rangle - f(x) | x \in \mathbb{R}\}$$

The set on the left-hand side of this equation is the epigraph of the function

$$f^*(x^*) = \sup_x \{\langle x, x^* \rangle - f(x)\}$$

which is called the *(convex) conjugate* of f. By definition,

$$f(x) = \sup_{x^*} \{\langle x, x^* \rangle - f(x^*)\},$$

so the conjugate f^{**} of f^* is equal to f.

Theorem C.0.6 (Rockafellar (1997, thm. 12.2 & Corr. 12.2.2))
Let f be a convex function. The conjugate function f^* is then a closed convex function, proper if and only if f is proper. Moreover, $(\operatorname{cl} f)^* = f^*$ and $f^{**} = \operatorname{cl} f$. For any convex function f on $\mathbb{R}^n$, one actually has

$$f^*(x^*) = \sup \{\langle x, x^* \rangle - f(x) | x \in \operatorname{ri}(\operatorname{dom} f)\}.$$

Definition C.0.7 (Subgradient and Subdifferential)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. $x^* \in \mathbb{R}^n$ is called a *subgradient* of f at x if

$$f(z) \geq f(x) + \langle x^*, z - x \rangle , \quad \forall z.$$

The set of all subgradients of f at x is said to be its *subdifferential* $\partial f(x)$.

Theorem C.0.8 (Rockafellar (1997, thm. 23.5))

For any proper convex function $f; \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}$, the following conditions are equivalent for $x^* \in \mathbb{R}^n$:

- $x^* \in \partial f(x)$;
- $\langle z, x^* \rangle - f(z)$ attains its supremum in z at $z = x$;
- $f(x) + f^*(x^*) \leq \langle x, x^* \rangle$;
- $f(x) + f^*(x^*) = \langle x, x^* \rangle$.
- if f is closed, also $x \in \partial f^*(x^*)$ is equivalent to the conditions above.

Appendix D

Validation of the LSM and LSMQ

We use a scaled version of the full regression model from Bauer et al. (2011*a*):

$$V_1 = \alpha + \beta_1 \frac{A_1}{10000} + \beta_2 \left(\frac{A_1}{10000}\right)^2 + \beta_3(1+r_1) + \beta_4(1+r_1)^2 + \beta_5 \frac{L_1}{10000} + \beta_6 \frac{R_1}{L_1}$$
$$+ \quad \beta_7 \frac{A_1}{10000} \exp(r_1) + \beta_8 \frac{L_1}{10000} \exp(r_1) + \beta_9 \exp\left(\frac{A_1}{10000}\right).$$

Exemplary coefficient estimates and regression diagnostics can be found in table D.1 and table D.2.

	SH						PH					
	1000		10000		100000		1000		10000		100000	
Model/i	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value
BSV												
1	-2.265E+07	0.61	1.750E+07	0.23	7.006E+05	0.88	1.001E+07	0.36	-1.300E+06	0.72	3.497E+05	0.76
2	-8.337E+04	0.97	1.486E+05	0.80	-6.748E+04	0.71	-4.430E+05	0.35	-5.304E+04	0.72	-1.749E+04	0.70
3	-1.655E+05	0.81	-2.042E+04	0.85	3.068E+04	0.42	-1.246E+05	0.46	-1.423E+04	0.61	2.224E+04	0.02
4	4.042E+07	0.65	-3.405E+07	0.24	-1.394E+06	0.88	-2.090E+07	0.34	2.453E+06	0.74	-7.118E+05	0.75
5	-1.798E+07	0.68	1.653E+07	0.25	7.299E+05	0.87	1.073E+07	0.33	-1.171E+06	0.75	3.872E+05	0.73
6	4.312E+06	0.48	-9.530E+05	0.62	1.201E+05	0.84	1.125E+06	0.45	1.921E+05	0.69	8.180E+04	0.59
7	-8.364E+04	0.91	-9.643E+03	0.94	-9.806E+03	0.83	2.157E+05	0.23	1.226E+04	0.73	4.022E+02	0.97
8	3.143E+05	0.86	-1.273E+05	0.82	7.086E+04	0.68	2.434E+05	0.58	4.680E+04	0.74	3.860E+03	0.93
9	-4.375E+06	0.46	9.262E+05	0.63	-1.456E+05	0.81	-8.734E+05	0.55	-1.645E+05	0.73	-7.108E+04	0.64
10	7.424E+04	0.87	1.409E+04	0.83	-1.736E+04	0.46	8.814E+04	0.42	8.639E+03	0.61	-1.169E+04	0.05
HCIR												
1	2.660E+07	0.05	3.151E+06	0.37	-1.001E+06	0.38	6.827E+06	0.14	-6.454E+05	0.57	-2.361E+05	0.53
2	1.148E+06	0.31	8.679E+04	0.78	-1.029E+05	0.30	3.347E+05	0.39	-2.053E+04	0.84	-8.133E+04	0.01
3	-7.263E+05	0.12	2.109E+04	0.82	-3.746E+04	0.24	-1.266E+05	0.43	2.445E+04	0.42	4.486E+03	0.67
4	-5.160E+07	0.05	-8.001E+06	0.25	1.996E+06	0.37	-1.292E+07	0.16	6.291E+05	0.78	5.428E+05	0.46
5	2.412E+07	0.07	4.881E+06	0.16	-1.047E+06	0.35	5.943E+06	0.19	4.151E+04	0.97	-3.048E+05	0.41
6	-3.719E+06	0.45	1.692E+06	0.16	2.574E+04	0.95	-1.239E+06	0.46	7.279E+05	0.06	3.473E+04	0.79
7	3.905E+05	0.56	-1.858E+05	0.21	4.070E+04	0.40	1.705E+05	0.46	-2.704E+04	0.57	1.827E+04	0.25
8	-1.116E+06	0.23	1.416E+05	0.60	9.436E+04	0.27	-4.613E+05	0.15	4.153E+04	0.64	5.934E+04	0.04
9	4.198E+06	0.39	-1.898E+06	0.11	1.369E+04	0.97	1.440E+06	0.39	-7.420E+05	0.05	-1.674E+03	0.99
10	3.994E+05	0.17	-2.854E+04	0.61	2.012E+04	0.30	8.084E+04	0.41	-1.604E+04	0.37	-1.697E+03	0.79

Table D.1: Regression results

	SH						PH					
	1000		10000		100000		1000		10000		100000	
Model/i	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value	$\hat{\beta}_i$	p-value
BSV												
1	-7.952E+06	0.26	3.907E+06	0.09	6.824E+05	0.35	-1.965E+06	0.27	6.231E+05	0.27	4.619E+04	0.80
2	-1.655E+06	0.05	1.029E+05	0.68	9.796E+04	0.27	-5.613E+05	0.01	1.718E+04	0.78	-6.625E+03	0.77
3	-9.272E+05	0.18	2.369E+05	0.03	8.531E+04	0.23	-7.506E+04	0.67	4.062E+04	0.13	3.032E+04	0.09
4	9.278E+06	0.49	-6.200E+06	0.16	-1.175E+06	0.40	2.868E+06	0.40	-8.656E+05	0.43	-2.134E+04	0.95
5	-2.490E+06	0.71	2.595E+06	0.24	5.836E+05	0.40	-1.012E+06	0.55	2.960E+05	0.59	7.552E+03	0.97
6	6.241E+06	0.03	-1.180E+06	0.21	-1.049E+05	0.72	1.390E+06	0.06	-2.932E+05	0.22	6.045E+03	0.94
7	9.307E+05	0.24	-1.196E+05	0.44	-7.831E+04	0.32	2.362E+05	0.24	-9.234E+03	0.81	-8.392E+03	0.67
8	1.017E+06	0.07	-1.071E+05	0.57	-1.896E+04	0.75	3.029E+05	0.03	-3.645E+04	0.43	4.578E+03	0.76
9	-5.123E+06	0.07	9.973E+05	0.27	1.167E+04	0.97	-1.110E+06	0.11	2.858E+05	0.20	-3.441E+03	0.96
10	5.894E+05	0.19	-1.296E+05	0.05	-6.031E+04	0.19	6.363E+04	0.57	-1.997E+04	0.23	-1.863E+04	0.11
HCIR												
1	-7.103E+05	0.85	-8.298E+05	0.46	2.642E+05	0.45	3.419E+04	0.98	1.378E+05	0.71	2.354E+05	0.04
2	-1.016E+06	0.29	-1.926E+05	0.36	1.285E+05	0.04	-1.352E+05	0.67	-5.145E+04	0.45	1.141E+04	0.59
3	-1.344E+06	0.13	-2.182E+04	0.84	-2.625E+04	0.27	-5.260E+05	0.08	-2.297E+04	0.51	3.646E+03	0.65
4	-1.133E+06	0.88	1.212E+06	0.57	-3.101E+05	0.65	-7.750E+05	0.75	-3.435E+05	0.62	-4.059E+05	0.07
5	2.596E+05	0.94	-4.097E+05	0.70	1.615E+04	0.96	1.274E+05	0.92	1.762E+05	0.61	1.734E+05	0.13
6	9.820E+05	0.65	5.399E+05	0.34	-3.783E+05	0.03	-9.618E+04	0.89	9.800E+04	0.60	-4.276E+04	0.47
7	1.397E+06	0.16	6.334E+04	0.68	-1.524E+04	0.70	4.382E+05	0.19	3.000E+04	0.56	-1.987E+02	0.99
8	1.319E+04	0.98	1.263E+05	0.37	-7.501E+04	0.09	-1.150E+05	0.46	2.975E+04	0.52	-9.887E+03	0.50
9	5.691E+05	0.75	-4.835E+05	0.38	3.513E+05	0.04	6.047E+05	0.31	-5.048E+04	0.78	5.557E+04	0.33
10	8.715E+05	0.13	2.044E+04	0.75	9.680E+03	0.48	3.278E+05	0.09	1.452E+04	0.49	-2.871E+03	0.53

Table D.2: Regression results

Bibliography

Acerbi, C. (2004), Coherent Representations of Subjective Risk Aversion, *in* G. Szegö, ed., 'Risk Measures for the 21st Century', Wiley, New York, 147–207.

Acerbi, C., Nordio, C. and Sirtori, C. (2001), Expected Shortfall as a Tool for Financial Risk Management. Accessed 12/14/2011.
http://www.pacca.info/public/files/docs/public/finance/Active%20Risk%20Management/ncs.pdf

Acerbi, C. and Tasche, D. (2002), 'On the Coherence of Expected Shortfall', *Journal of Banking and Finance* **26**(7), 1487–1503.

Ahlip, R. and Rutkowski, M. (2009), 'Forward Start Options Under Stochastic Volatility and Stochastic Interest Rates', *International Journal of Theoretical & Applied Finance* **12**(2), 209–225.

Aihara, S., Bagchi, A. and Saha, S. (2009), 'On Parameter Estimation of Stochastic Volatility Models from Stock Data Using Particle Filter - Application to AEX Index', *International Journal of Innovative Computing, Information and Control* **5**(1), 17–27.

Ale, B. (2002), 'Risk Assessment Practices in The Netherlands', *Safety Science* **40**(1-4), 105–126.

Alexandrow, A. D. (1957), Über eine Verallgemeinerung der Riemannschen Geometrie, *in* 'Der Begriff des Raumes in der Geometrie - Bericht von der Riemann-Tagung des Forschungsinstituts für Mathematik', Vol. 1 of *Schriftenreihe des Forschungsinstituts für Mathematik*, Deutsche Akademie der Wissenschaften zu Berlin, Akademie-Verlag Berlin, 33–84.

Ambrosio, L. and Gigli, N. (2010), A User's Guide to Optimal Transport. Accessed 06/03/2011.
http://cvgmt.sns.it/cgi/get.cgi/papers/ambgig11/users_guide@-@final.pdf

Ambrosio, L., Gigli, N. and Savaré, G. (2008), *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zürich, Birkhäuser, Boston.

Andersen, L. B. G. (2008), 'Efficient Simulation of the Heston Stochastic Volatility Model', *The Journal of Computational Finance* **11**(3), 1–42.

Aragonés, J. R., Blanco, C. and Dowd, K. (2008), Stress Tests, Market Risk Measures and Extremes: Bringing Stress Tests to the Forefront of Market Risk Management, *in* 'Stress Testing for Financial Institutions', Risk Books, chapter 2, 17–33.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999), 'Coherent Measures of Risk', *Mathematical Finance* **9**(3), 203 – 228.

Aven, T. (2007), 'A Unified Framework for Risk and Vulnerability Analysis Covering Both Safety and Security', *Reliability Engineering & System Safety* **92**(6), 745–754.

Aven, T. (2010), 'On the Need for Restricting the Probabilistic Analysis in Risk Assessments to Variability.', *Risk Analysis: An International Journal* **30**(3), 354–360.

Aven, T. (2011*a*), 'On how to Conceptualise and Describe Risk', *Reliability & Risk Analysis: Theory & Application* **2**(2), 28–37. Issue (20).

Aven, T. (2011*b*), *Quantitative Risk Assessment: The Scientific Platform*, Cambridge University Press.

Aven, T. and Renn, O. (2009), 'On Risk Defined as an Event where the Outcome is Uncertain', *Journal of Risk Research* **12**(1), 1–11.

Balbás, A., Garrido, J. and Mayoral, S. (2009), 'Properties of Distortion Risk Measures', *Methodology and Computing in Applied Probability* **11**, 385–399.

Barrieu, P. and Karoui, N. E. (2009), Pricing, Hedging and Optimally Designing Derivatives via Minimization of Risk Measures, *in* R. Carmona, ed., 'Indifference Pricing - Theory and Application', Princeton University Press, chapter 3, 77–146.

Basel Committee on Banking Supervision (2006), The IRB Use Test: Background and Implementation, Technical report, Basel Committee on Banking Supervision. Basel Committee Newsletter No. 9.

Basel Committee on Banking Supervision (2011), 'Basel III: A Global Regulatory Framework for more Resilient Banks and Banking Systems'. Revised version of the Basel III capital rules, accesssed 08/08/2012.
`http://www.bis.org/publ/bcbs189.pdf`

Bauer, D., Bergmann, D. and Kiesel, R. (2008), On the Risk-Neutral Valuation of life insurance contract with numerical methods in view. AFIR2008 Paper, Accessed 07/28/2009.
`http://www.italian-actuaries.org/afir/cd1/Papers/AFIR2008_Paper_Bauer%20Bergmann%20Kiesel.pdf`

Bauer, D., Bergmann, D. and Reuss, A. (2009), Solvency II and Nested Simulations - a Least-Squares Monte Carlo Approach. Preprint Series, Ulm University.
`http://numerik.uni-ulm.de/preprints/2009/200905_Solvency_Preprint-Server.pdf`

Bauer, D., Bergmann, D. and Reuss, A. (2011*a*), A Least-Squares Monte Carlo Approach to the Calculation of Capital Requirements. Forthcoming.
`http://numerik.uni-ulm.de/preprints/2009/200905_Solvency_Preprint-Server.pdf`

Bauer, D., Bergmann, D. and Reuss, A. (2011*b*), 'On the Calculation of the Solvency Capital Requirement based on Nested Simulations', *ASTIN Bulletin* . Forthcoming.

Bauer, D., Kiesel, R., Kling, A. and Ruß, J. (2006), 'Risk Neutral Valuation of With-Profits Life Insurance Contracts', *Insurance: Mathematics and Economics* **39**, 171–183.

Baumgartner, G., Bühlmann, H. and Koller, M. (2004), ‘Multidimensional Valuation of Life Insurance Policies and Fair Value’, *Mitteilungen der Schweiz. Aktuarvereinigung* **1**, 27–64.

Baxter, A., Muir, M. and Leung, V. (2008), Implementing Replicating Portfolios: Benefits and Pitfalls. 2008 Life Convention.

Bennemann, C., Oehlenberg, L. and Stahl, G., eds (2011), *Handbuch Solvency II*, Schäffer-Poeschel Verlag Stuttgart.

Bergmann, D., Reuss, A., Siebert, A., Stahl, G. and Zwiesler, H.-J. (2009), Computational Aspects of Nested Monte-Carlo Simulations for Risk Management Purposes. Working Paper.

Bickel, P. J. and Freedman, D. A. (1981), ‘Some Asymptotic Theory for the Bootstrap’, *The Annals of Statistics* **9**(6), 1196–1217.

Bickel, P. J. and Lehmann, E. L. (1975), ‘Descriptive Statistics for Nonparametric Models II. Location’, *The Annals of Statistics* **3**(5), 1045–1069.

Billingsley, P. (1986), *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, 2nd edition, Wiley, New York.

Billingsley, P. (1999), *Convergence of Probability Measures*, Wiley, New York.

Bingham, N. H. and Kiesel, R. (2004), *Risk-neutral valuation*, Springer finance, 2nd edn., Springer, New York.

Boissard, E. and Gouic, T. L. (2011), On the Mean Speed of Convergence of Empirical and Occupation Measures in Wasserstein Distance. Available at arXiv:1105.5263v1.

Bolley, F. (2008), Seperability and Completeness for the Wasserstein Distance, *in* C. Donati-Martin, M. Émery, A. Rouault and C. Stricker, eds, ‘Séminaire de Probabilités XLI’, Vol. 1934 of *Lecture Notes in Mathematics*, Springer, New-York.

Bolley, F., Guillin, A. and Villani, C. (2007), ‘Quantitative Concentration Inequalities for Empirical Measures on Non-compact Spaces.’, *Probability Theory & Related Fields* **137**(3/4), 541–593.

Boos, D. D. (1979), ‘A Differential for L-Statistics’, *The Annals of Statistics* **7**(5), 955–959.

Brennan, M. J. and Schwartz, E. (1976), ‘The pricing of equity-linked life insurance policies with an asset value guarantee’, *Journal of Financial Economics* **3**(3), 195–213.

Brigo, D. and Mercurio, F. (2006), *Interest Rate Models - Theory and Practice*, Springer Finance, 2nd edn, 3rd corr. printing 2007, Springer, New-York.

Bris, A., Welch, I. and Ning, Z. (2006), ‘The Costs of Bankruptcy: Chapter 7 Liquidation versus Chapter 11 Reorganization.’, *Journal of Finance* **61**(3), 1253–1303.

Broadie, M., Du, Y. and Moallemi, C. D. (2010), Efficient Risk Estimation via Nested Sequential Simulation. Accessed 04/30/2010.
`http://www.pfp.gsu.edu/Research/downloads/2010/Broadie_seqrisk-2010.pdf`

Broadie, M. and Kaya, z. (2006), ‘Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes’, *Operations Research* **54**(2), 217–231.

Buchwalder, M., Bühlmann, H., Merz, M. and Wüthrich, M. V. (2007), 'Valuation Portfolio in Non-Life Insurance', *Scandinavian Actuarial Journal* **2007**(2), 108–125.

CEA Groupe Consultatif (2007), Solvency II Glossary. Accessed 12/16/2012. `http://ec.europa.eu/internal_market/insurance/docs/solvency/impactassess/annex-c08d_en.pdf`

CEIOPS (2008*a*), Advice to the European Commission on the Principle of Proportionality in the Solvency II Framework Directive Proposal, Technical report, CEIOPS. CEIOPS-DOC-24/08.

CEIOPS (2008*b*), Own Risk and Solvency Assessment (ORSA), Technical report, CEIOPS. Issues Paper, CEIOPS-IGSRR-09/08.

CEIOPS (2009*a*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Articles 120 to 126 Tests and Standards for Internal Model Approval, Technical report, CEIOPS.

CEIOPS (2009*b*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Own funds - Article 130 Calculation of the MCR, Technical report, CEIOPS. CEIOPS-DOC-47/09, former Consultation Paper 55.

CEIOPS (2009*c*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Own funds - Article 97 and 99 - Classification and eligibility, Technical report, CEIOPS. CEIOPS-DOC-39/09, former Consultation Paper 46.

CEIOPS (2009*d*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula - Article 111 Non-Life Underwriting Risk, Technical report, CEIOPS. CEIOPS-DOC-41/09, former CP48.

CEIOPS (2009*e*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula - Counterparty Default Risk Module, Technical report, CEIOPS. CEIOPS-DOC-23/09, former CP28 and CP51.

CEIOPS (2009*f*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula - Health Underwriting Risk Module, Technical report, CEIOPS. CEIOPS-DOC-43/09, former CP50.

CEIOPS (2009*g*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula - The Procedure to be Followed for the Approval of an Internal Model, Technical report, CEIOPS. CEIOPS-DOC-28/09, former CP37.

CEIOPS (2009*h*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Article 109 - Structure and Design of Market Risk Module, Technical report, CEIOPS. CEIOPS-DOC-40/09, former CP47.

CEIOPS (2009*i*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Article 109(c) - Life Underwriting Risk, Technical report, CEIOPS. CEIOPS-DOC-42/09, former CP49.

CEIOPS (2009*j*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Supervisory Reporting and Public Disclosure Requirements, Technical report, CEIOPS. CEIOPS-DOC-50/09.

CEIOPS (2009*k*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: System of Governance, Technical report, CEIOPS. CEIOPS-DOC-29/09.

CEIOPS (2009*l*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Technical Provisions - Article 86 b - Risk-free interest rate term structure, Technical report, CEIOPS. CEIOPS-DOC-34/09, former Consultation Paper 40.

CEIOPS (2009*m*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Technical Provisions - Article 86 d - Calculation of the Risk Margin, Technical report, CEIOPS. CEIOPS-DOC-36/09, former Consultation Paper 42.

CEIOPS (2009*n*), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Valuation of Assets and "Other Liabilities", Technical report, CEIOPS. CEIOPS-DOC-31/09, former Consultation Paper 35.

CEIOPS (2010), CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Article 111(d) Correlations, Technical report, CEIOPS. CEIOPS-DOC-70/10, former CP74.

CFO Forum (2009), Market Consistent Embedded Value Principles. Accessed 05/20/2010. `http://cfoforum.nl/downloads/MCEV_Principles_and_Guidance_October_2009.pdf`

Cheridito, P. and Li, T. (2009), 'Risk Measures on Orlicz Hearts', *Mathematical Finance* **19**(2), 189–214.

Çinlar, E., Jacod, J., Protter, P. and Sharpe, M. J. (1980), 'Semimartingales and Markov processes', *Probability Theory and Related Fields* **54**, 161–219.

Clark, D., Gillespie, O. and Verheugen, H. (2008), 'Solvency II - Accelerating the Calculations: Replicating Portfolios and Model Compression', Milliman White Paper.

Cont, R. (2006), 'Model Uncertainty And Its Impact On The Pricing Of Derivative Instruments', *Mathematical Finance* **16**(3), 519–547.

Cont, R. and da Fonseca, J. (2002), 'Dynamics of implied volatility surfaces', *Quantitative Finance* **2**, 45–60.

Cont, R., Deguest, R. and He, X. D. (2011), Loss-Based Risk Measures. Working Paper, Available at SSRN.
`http://ssrn.com/paper=1755291`

Cont, R., Deguest, R. and Scandolo, G. (2010), 'Robustness and Sensitivity Analysis of Risk Measurement Procedures', *Quantitative Finance* **10**(6), 593 – 606.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985), 'A Theory of the Term Strucuture of Interest Rates', *Econometrica* **53**(2), 385–407.

CRO Forum (2009), Insurance Risk Management Response to the Financial Crisis. Accessed 11/02/2011.
`http://www.thecroforum.org/assets/files/publications/croforum-insuranceriskmanagementresponsetothefinancialcrisis20090423%282%29.pdf`

Crouhy, M., Galai, D. and Mark, R. (2001), *Risk management*, McGraw-Hill.

Dall'Aglio, G. (1956), 'Sugli estremi dei momenti delle funzioni di ripartizione doppia', *Annali della Scuola Normale Superiore di Pisa - Scienze Fisiche e Matematiche* **10**, 35–74. In Italian.

Daníelsson, J., Jørgensen, B. N., Mandira, S., Samorodnitsky, G. and de Vries, C. G. (2005), Subadditivity Re-Examined: the Case for Value-at-Risk. LSE Discussion Paper 549, Accessed 08/29/2011.
`http://eprints.lse.ac.uk/id/eprint/24668`

Daníelsson, J., Jørgensen, B. N., Mandira, S., Samorodnitsky, G. and de Vries, C. G. (2011), Fat Tails, VaR and Subadditivity. Accessed 08/29/2011.
`http://www.riskresearch.org/files/JD-Cd-BJ-SM-GS-35.pdf`

Davies, P. L. (1995), 'Data Features', *Statistica Neerlandica* **49**(2), 185–245.

de Rocquigny, E., Devictor, N. and Tarantola, S., eds (2008), *Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management*, Wiley, New York.

Delbaen, F. (2002), Coherent Risk Measures on General Probability Spaces, *in* K. Sandmann and P. J. Schönbucher, eds, 'Advances in Finance and Stochastics', Springer.

Denneberg, D. (1994), *Non-additive Measure and Integral*, Vol. 27 of *Theory and decision library, Series B, Mathematical and statistical methods*, Kluwer Academic.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2006), *Actuarial Theory for Dependent Risks*, John Wiley & Sons, Ltd.

Dhaene, J., Denuit, M., Goovaerts, M., Kaas, R. and Vyncke, D. (2002), 'The concept of comonotonicity in actuarial science and finance: theory', *Insurance: Mathematics and Economics* **31**(1), 3–33.

Dhaene, J., Laeven, R. J. A., Vanduffel, S., Darkiewicz, G. and Goovaerts, M. J. (2008), 'Can a Coherent Risk Measure Be Too Subadditive?', *Journal of Risk & Insurance* **75**(2), 365–386.

Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M., Kaas, R. and Vyncke, D. (2004), 'Capital requirements, risk measures and comonotonicity', *Belgian Actuarial Bulletin* **4**(1), 53–61.

Doherty, N. A. (1985), *Corporate Risk Management*, McGraw-Hill.

Dowd, K. and Blake, D. (2006), 'After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures', *Journal of Risk and Insurance* **73**(2), 193–229.

Dudley, R. M. (1969), 'The Speed of Mean Glivenko-Cantelli Convergence', *The Annals of Mathematical Statistics* **40**(1), 40–50.

Dudley, R. M. (2002), *Real Analysis and Probability*, Cambridge University Press.

Durante, F. and Sempi, C. (2010), Copula Theory: An Introduction, *in* P. Jaworski, F. Durante, W. Härdle and T. Rychlik, eds, 'Copula Theory and Its Applications', Vol. 198 of *Lecture Notes in Statistics*, Springer, 3–31.

Eidgenössische Finanzmarktaufsicht FINMA (2008), Swiss Solvency Test (SST), Technical report, FINMA. Circular 2008/44.

Erixon, E. and Tubis, J. (2008), 'Replicating Portfolios for Variable Annuity Hedging', *Emphasis* **4**, 6–9.

Escaniano, J. C. and Olmo, J. (2008), Robust Backtesting for Value-at-Risk Models. Working Paper, Dept. of Economics, Indiana University.
`http://www.finance-innovation.org/risk09/work/1569411.pdf`

European Commission Directorate-General for Translation (2012), 'English Style Guide'. Seventh edition: August 2011, Last updated: May 2012.
`http://ec.europa.eu/translation/english/guidelines/documents/styleguide_english_dgt_en.pdf`

Faulhaber, M. (2004), Risikosteuerung in Versicherungsunternehmen — Die Entwicklung der Aufgaben des Verantwortlichen Aktuars seit 1994, *in* K. Spremann, ed., 'Versicherungen im Umbruch', Springer Verlag Berlin, 97–114.

Fecht, F., Nyborg, K. G. and Rocholl, J. (2008), 'Liquidity management and overnight rate calendar effects: Evidence from German banks', *North American Journal of Economics and Finance* **19**, 7–21.

Filipović, D. and Svindland, G. (2008), 'Optimal Capital and Risk Allocations for Law- and Cash-Invariant Convex Functions', *Finance and Stochastics* **12**, 423–439.

Financial Services Authority (2008), Insurance Risk Management: The Path to Solvency II, Technical report, FSA. Discussion Paper 08/4, Accessed 05/12//2012.

Föllmer, H. (2008), Financial Uncertainty, Risk Measures and Robust Preference, *in* M. Yor, ed., 'Aspects of Mathematical Finance', Springer, Berlin, 3–14.

Föllmer, H. (2012), Risk, Model Risk, and Knightian Uncertainty: On the Role of Probability in Finance. Talk at the Hausdorff Center for Mathematics Probability Day 2012 in Bonn.

Föllmer, H. and Penner, I. (2011), 'Monetary Valuation Of Cash Flows Under Knightian Uncertainty', *International Journal of Theoretical & Applied Finance* **14**(1), 1–15.

Föllmer, H. and Schied, A. (2002*a*), 'Convex Measures of Risk and Trading Constraints', *Finance and Stochastics* **6**, 429–447.

Föllmer, H. and Schied, A. (2002*b*), Robust Preferences and Convex Measures of Risk, *in* 'Advances in Finance and Stochastics', Springer-Verlag, 39–56.

Föllmer, H. and Schied, A. (2011), *Stochastic Finance*, 3rd edn., Walter de Gruyter, Berlin.

Fréchet, M. (1951), 'Sur les tableaux de corrélation dont les marges sont données', *Ann. Univ. Lyon Sect. A Série 3* **14**, 53–77.

Gallagher, R. B. (1956), 'Risk Management: New Phase of Cost Control.', *Harvard Business Review* **34**(5), 75–86.

Gigli, N. (2004), On the Geometry of the Space of Probability Measures Endowed with the Quadratic Optimal Transport Distance, PhD thesis, Scuola Normale Superiore di Pisa.

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering*, Vol. 53 of *Applications of Mathematics*, Springer, New York.

Gordy, M. B. and Juneja, S. (2008), Nested Simulation in Portfolio Risk Management. Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.

Green, L. (2009), Legal Positivism, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', fall 2009, The Metaphysics Research Lab, Stanford University.

Grosen, A. and Jørgensen, P. L. (2000), 'Fair Valuation of Life Insurance Liabilities: The Impact of Interest Rate Guarantees, Surrender Options, and Bonus Policies', *Insurance: Mathematics and Economics* **26**(1), 37–57.

Grzelak, L. and Oosterlee, C. (2011), 'On the Heston Model with Stochastic Interest Rates', *SIAM Journal on Financial Mathematics* **2**(1), 255–286.

Hampel, F. R. (1971), 'A General Qualitative Definition of Robustness', *The Annals of Mathematical Statistics* **42**(6), 1887–1896.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986), *Robust statistics: The Approach Based on Influence Functions*, Wiley series in probability and mathematical statistics, John Wiley & Sons.

Hansen, R. G. and Lott, Jr., J. R. (1996), 'Externalities and Corporate Objectives in a World with Diversified Shareholder/Consumers', *Journal of Financial and Quantitative Analysis* **31**(1), 43–68.

Hart, H. L. A. (1997), *The Concept of Law*, Clarendon Law Series, 2nd edn., paperback, Clarendon Press, Oxford.

Hennessy, C. A. and Whited, T. M. (2007), 'How Costly Is External Financing? Evidence from a Structural Estimation', *The Journal of Finance* **62**(4), 1705–1745.

Heston, S. L. (1993), 'A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options', *The Review of Financial Studies* **6**(2), 327–343.

Heyde, C. C., Kou, S. and Peng, X. H. (2007), What Is a Good External Risk Measure: Bridging the Gaps between Robustness, Subadditivity, and Insurance Risk Measures. Preprint, Columbia University.
`http://www.fdic.gov/bank/analytical/cfr/2007/apr/KOU_hkpv1.pdf`

Heyde, C. and Kou, S. (2004), 'On the controversy over tailweight of distributions', *Operations Research Letters* **32**(5), 399–408.

Hoeffding, W. (1940), 'Masstabinvariante Korrelationstheorie', *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin* **5**, 179–233.

Holton, G. A. (2009), 'The Case for Incoherence'. Accessed 12/12/2011.
`http://glynholton.com/2008/09/the-case-for-incoherence/`

Horowitz, J. and Karandikar, R. L. (1994), 'Mean Rates of Convergence of Empirical Measures in the Wasserstein Metric', *Journal of Computational and Applied Mathematics* **55**(3), 261–273.

Huber, P. J. and Ronchetti, E. M. (2011), *Robust Statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons.

Hull, J. (2011), *Risikomanagement*, 2nd edn., Pearson Prentice Hall, Upper Saddle River.

Hull, J. C. (2009), 'The Credit Crunch of 2007: What Went Wrong? Why? What Lessons Can Be Learned?', *Journal of Credit Risk* **5**(2), 3–18.

Hull, J. C. (2012), *Options, Futures and other Derivatives*, eighth edn., Pearson Prentice Hall, Upper Saddle River.

International Association of Insurance Supervisors (2007), Guidance Paper on the Use of Internal Models for Risk and Capital Management Purposes by Insurers, Technical report, International Association of Insurance Supervisors. Guidance Paper No. 2.2.7.

Jaschke, S. and Küchler, U. (2001), 'Coherent risk measures and good-deal bounds', *Finance and Stochastics* **5**, 181–200.

Jorion, P. (2009), *Value at Risk*, 3rd edn., McGraw-Hill, New-York.

Kahneman, D., Slovic, P. and Tversky, A. (1982), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Kahneman, D. and Tversky, A. (1979), 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica* **47**(2), 263–291.

Kalberer, T. (2006), 'Market Consistent Valuation of Insurance Liabilities', *Der Aktuar* **12**(1), 7–14.

Kallenberg, O. (2002), *Foundations of Modern Probability*, Probability and its Applications, 2nd edn., Springer, New York.

Karr, A. F. (1993), *Probability*, Springer Texts in Statistics, Springer, New York.

Kellerer, H. G. (1985), Duality Theorems and Probability Metrics, *in* M. Iosifescu, ed., 'Proceedings of the 7th Conference on Probability Theory', The Centre of Mathematical Statistics of the National Institute of Metrology Bucharest, 211–220.

Kim, Y.-J. (2002), 'Option Pricing under Stochastic Interest Rates: An Empirical Investigation', *Asia-Pacific Financial Markets* **9**(1), 23–44.

Kloman, H. F. (1990), 'Risk Management Agonistes', *Risk Analysis* **10**(2), 201–205.

Klügel, J.-U. (2007), 'Comment On: "Why do Modern Probabilistic Seismic-Hazard Analysis Often Lead to Increased Hazard Estimates?" by Julian J. Brommer and Norman A. Abrahamson', *Bulletin of the Seismological Society of America* **97**(6), 2198–2207.

Knauth, K.-W. (2005), Systemwandel in der Finanzdienstleistungs- und Versicherungsaufsicht, *in* 'Solvency II & Risikomanagement', Gabler, Wiesbaden, 3–34.

Knight, F. H. (1971), *Risk, Uncertainty and Profit*, reprint of the 1921 edn., University of Chicago Press, Chicago.

Krätschmer, V., Schied, A. and Zähle, H. (2012*a*), Comparative and qualitative robustness for law-invariant risk measures. Available at arXiv:1204.2458v1.
`http://arxiv.org/abs/1204.2458`

Krätschmer, V., Schied, A. and Zähle, H. (2012*b*), 'Qualitative and Infinitesimal Robustness of Tail-Dependent Statistical Functionals', *Journal of Multivariate Analysis* **103**(1), 35–47.

Krätschmer, V. and Zähle, H. (2011), 'Sensitivity of Risk Measures with Respect to the Normal Approximation of Total Claim Distributions', *Insurance: Mathematics and Economics* **49**(3), 335–344.

Krein, M. and Milman, D. (1940), 'On Extreme Points of Regular Convex Sets', *Studia Mathematica* **9**, 133–138.

Lamfalussy, A., Herkströter, C., Rojo, L. A., Ryden, B., Spaventa, L., Walter, N. and Wicks, N. (2001), Final Report of the Committee of Wise Men on the Regulation of European Securities Markets, Technical report, The Committee of Wise Men, Brussels.

Lan, H. (2009), Tuning The Parameters Of A Two-level Simulation Procedure With Screening. Working Paper.
`http://users.iems.northwestern.edu/~staum/Lan%20Parameter%20Tuning.pdf`

Lan, H., Nelson, B. L. and Staum, J. (2007), Two-level Simulation for Risk Management, *in* 'Proceedings of the 2007 INFORMS Simulation Society Workshop'.
`http://www.informs-sim.org/2007informs-csworkshop/23.pdf`

Liu, M. and Staum, J. (2010), Stochastic Kriging for Efficient Nested Simulation of Expected Shortfall. Accessed 04/30/2010.
`http://users.iems.northwestern.edu/~staum/skes.pdf`

Loebnitz, K. and Roorda, B. (2011), 'Liquidity Risk Meets Economic Capital and RAROC'.
`http://ssrn.com/paper=1853233`

Longstaff, F. A. and Schwartz, E. (2001), 'Valuing American Options by Simulation: a simple Least-Squares Approach', *The Review of Financial Studies* **14**, 113–147.

Lord, R., Koekkoek, R. and Dijk, D. V. (2010), 'A Comparison of Biased Simulation Schemes for Stochastic Volatility Models', *Quantitative Finance* **10**(2), 177–194.

Lüthi, H.-J. and Doege, J. (2005), 'Convex Risk Measures for Portfolio Optimization and Concepts of Flexibility', *Mathematical Programming* **104**, 541–559.

Major, P. (1978), 'On the Invariance Principle for Sums of Independent Identically Distributed Random Variables', *Journal of Multivariate Analysis* **8**(4), 487 – 517.

Mallows, C. L. (1972), 'A Note on Asymptotic Joint Normality', *The Annals of Mathematical Statistics* **43**(2), 508–515.

Markowitz, H. (1952), 'Portfolio Selection', *The Journal of Finance* **7**(1), 77–91.

Mason, L. (2008), 'Builing Smart Internal Models', *Emphasis* **4**, 2–5.

McNeil, A. J., Frey, R. and Embrechts, P. (2005), *Quantitative Risk Management*, Princeton Series in Finance, Princeton University Press, Princeton and Oxford.

Merkle, M. (2000), 'Topics in Weak Convergence of Probability Measures', *Zbornik Radova. Nova Serija. Math. Inst., Belgrade* **9**(17), 235–274.

Meyburg, A. H. (2006), Evaluation of Practice Today: Data Needs, *in* 'Freight Demand Modeling: Tools for Public-Sector Decision Making', Vol. 40 of *Conference Proceedings*, Transportation Reserach Board of the National Academies.
`http://onlinepubs.trb.org/onlinepubs/conf/CP40.pdf#page=39`

Mikhailov, S. and Nögel, U. (2003), 'Heston's Stochastic Volatility Model Implementation, Calibration and Some Extensions', *Wilmott Magazine* , 74–79.

Mosteller, F. and Tukey, J. W. (1977), *Data Analysis and Regression: a Second Course in Statistics*, Addison-Wesley Series in Behavioral Science, Addison-Wesley Pub. Co., Reading.

Oechslin, J., Aubry, O., Aellig, M., Käppeli, A., Brönnimann, D., Tandonnet, A. and Valois, G. (2007), 'Replicating Embedded Options', *Life & Pensions* , 47–52.

Oehlenberg, L., Stahl, G. and Bennemann, C. (2011), Von der Standardformel zum internen Modell - ein Überblick über Solvency II, *in* C. Bennemann, L. Oehlenberg and G. Stahl, eds, 'Handbuch Solvency II', Schäffer-Poeschel Verlag Stuttgart, 3 – 36.

Olivier, L., Huang, B. and Craig, I. (2012), Dual Particle Filters for State and Parameter Estimation With Application to a Run-of-mine Ore Mill. Accessed 12/11/2012.
`http://repository.up.ac.za/bitstream/handle/2263/19105/Olivier_Dual%282012%29.pdf?sequence=1`

Papadopoulos, A. (2005), *Metric Spaces, Convexity and Nonpositive Curvature*, Vol. 6 of *IRMA Lectures in Mathematics and Theoretical Physics*, European Mathematical Society.

Pfeiffer, D. and Strassburger, D. (2008), 'Solvency II Stability Problems with the SCR aggregation formula', *Scandinavian Actuarial Journal* (1), 61–77.

Pflug, G. C., Pichler, A. and Wozabal, D. (2012), 'The 1/N investment strategy is optimal under high model ambiguity', *Journal of Banking & Finance* **36**(2), 410–417.

Pflug, G. and Römisch, W. (2007), *Modeling, Measuring and Managing Risk*, World Scientific.

Prokopczuk, M., Rachev, S. T., Schindlmayr, G. and Trück, S. (2007), 'Quantifying Risk in the Electricity Business: A RAROC-based Approach', *Energy Economics* **29**(5), 1033–1049.

Protter, P. E. (2005), *Stochastic Integration and Differential Equations*, Vol. 21 of *Stochastic Modelling and Applied Probability*, version 2.1, 2nd edn., Springer, New York.

Rachev, S. T. (1991), *Probability Metrics and the Stability of Stochastic Models*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Rachev, S. T., Menn, C. and Fabozzi, F. J. (2005), *Fat-Tailed and Skewed Asset Distributions*, John Wiley & Sons.

Rachev, S. T., Stoyanov, S. V. and Fabozzi, F. J. (2008), *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization*, The Frank J. Fabozzi Series, John Wiley & Sons.

Reed, M. and Simon, B. (2005), *Functional Analysis*, Vol. 1 of *Methods of Modern Mathematical Physics*, rev. and enlarged edn., Academic Press, San Diego.

Reeds, J. A. (1976), On the Definition of Von Mises Functionals, PhD thesis, Harvard University.

Renz, M. and Best, G. (2005), Was bedeutet Solvency II für die Lebensversicherung?, *in* 'Solvency II & Risikomanagement', Gabler, Wiesbaden, 317–336.

Rittmann, M. (2009), *Neuausrichtung der Versicherungsaufsicht im Rahmen von Solvency II*, Gabler, Wiesbaden.

Rockafellar, R. T. (1997), *Convex Analysis*, Princeton Landmarks in Mathematics, 10th printing, Princeton University Press, Princeton.

Rousseeuw, P. J. (1981), 'A new infinitesimal approach to robust estimation', *Probability Theory and Related Fields* **56**, 127–132.

Rowe, D. (2010), 'Don't distort the message', *Risk Magazine* **23**(5), 77.

Royden, H. L. (1968), *Real Analysis*, 2nd edn, Macmillan & Co., London.

Rüschendorf, L. (2002), Wasserstein Metric, *in* 'Encyclopaedia of Mathematics', online edition, Springer.
`http://eom.springer.de/w/w120020.htm`

Sandström, A. (2007), 'Solvency II: Calibration for Skewness', *Scandinavian Actuarial Journal* (2), 126–134.

Sawilowsky, S. S. (2002), 'Fermat, Schubert, Einstein, and Behrens-Fisher: The Probable Difference Between Two Means When $\sigma_1^2 \neq \sigma_2^2$', *Journal of Modern Applied Statistical Methods* **1**(2), 461–472.

Schäl, I. (2011), *Management von Operationellen Risiken*, Gabler Verlag - Springer Fachmedien, Wiesbaden.

Schmelzle, M. (2010), Option Pricing Formulae using Fourier Transform: Theory and Application. Accessed 12/11/2012.

Scott, L. O. (1997), 'Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods', *Mathematical Finance* **7**(4), 413–426.

Seemann, A. (2009), Replizierende Portfolien in der Lebensversicherung. PrePrint, Ulm University, Accessed 05/20/2010.
`http://numerik.uni-ulm.de/preprints/2009/ReplicatingPortfolios.pdf`

Sereda, E. N., Bronshtein, E. M., Rachev, S. T., Fabozzi, F. J., Sun, W. and Stoyanov, S. V. (2010), Distortion Risk Measures in Portfolio Optimization, *in* J. B. Guerard, ed., 'Handbook of Portfolio Construction', Springer, New York, chapter 25, 649–673.
`http://dx.doi.org/10.1007/978-0-387-77439-8_25`

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Shao, J. (1989), 'Functional Calculus and Asymptotic Theory for Statistical Analysis', *Statistics & Probability Letters* **8**, 397–405.

Shao, J. (1993), 'Differentiability of Statistical Functionals and Consistency of the Jackknife', *The Annals of Statistics* **21**(1), 61–75.

Shao, J. and Tu, D. (1996), *The Jackknife and Bootstrap*, Springer Series in Statistics, 2nd corrected printing, Springer.

Sharma, P. (2002), Prudential Supervision Of Insurance Undertakings, Technical report, Conference of Insurance Supervisory Services of the Member States of the European Union.

Shephard, N. G. (1991), 'From Characteristic Function to Distribution Function: A Simple Framework for the Theory', *Econometric Theory* **7**(4), 519–529.

Shreve, S. E. (2004), *Stochastic Calculus for Finance II*, Springer, New York.

Stahl, G., Zheng, J., Kiesel, R. and Rühlicke, R. (2012), Conceptualizing Robustness in Risk Management. Working Paper, available at SSRN.
`http://ssrn.com/abstract=2065723`

Stein, E. M. and Stein, J. C. (1991), 'Stock Price Distributions with Stochastic Volatility: An Analytic Approach', *The Review of Financial Studies* **4**(4), 727–752.

Sturm, K.-T. (2006), 'On the geometry of metric measure spaces. I', *Acta Mathematica* **196**, 65–131.

The Council of the European Communities (1973), 'Council Directive 73/239/EEC', Official Journal of the European Communities.

The Council of the European Communities (1979), 'Council Directive 79/267/EEC', Official Journal of the European Communities.

The Council of the European Communities (1999), 'Council Decision 1999/468/EC', Official Journal of the European Communities.
`http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1999:184:0023:0026:EN:PDF`

The Council of the European Communities (2006), 'Council Decision 2006/512/EC', Official Journal of the European Communities.
`http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:200:0011:0013:EN:PDF`

The European Commission (1999), 'The Review of the Overall Financial Position of an Insurance Undertaking (Solvency II) Review'. MARKT/2095/99-EN, Accessed 10/14/2011.
`http://ec.europa.eu/internal_market/insurance/docs/markt-2095/markt-2095-99_en.pdf`

The European Commission (2002*a*), 'Considerations on the Design of a Future Prudential Supervisory System'. MARKT/2535/02-EN, Accessed 05/12/2012.
`http://ec.europa.eu/internal_market/insurance/docs/markt-2535/markt-2535-02_en.pdf`

The European Commission (2002*b*), 'Note to the Solvency Subcommittee: Risk Models of Insurance Companies or Groups'. MARKT/2515/02-EN, Accessed 05/12/2012.
http://www.gcactuaries.org/documents/2515.pdf

The European Commission (2004*a*), 'Commission Decision of 5 November 2003 Establishing the Committee of European Insurance and Occupational Pensions Supervisors'. Commission Decision 2004/6/EC, Accessed 10/14/2011.
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:003:0030:0031:EN:PDF

The European Commission (2004*b*), 'Note to the IC Solvency II Subcommittee'. MARKT/2543/03-EN, Accessed 10/14/2011.
http://ec.europa.eu/internal_market/insurance/docs/markt-2543-03/markt-2543-03_en.pdf

The European Commission (2007), 'Commission Staff Working Document Accompanying Document to the Proposal for a Directive of the European Parliament and of the Council on the Taking-up and Pursuit of the Business of Insurance and Reinsurance Solvency Ii Impact Assessment Report'. COM(2007) 361 final, Accessed 10/14/2011.
http://ec.europa.eu/internal_market/insurance/docs/markt-2519/markt-2519-02-solv2-lamfalussy_en.pdf

The European Commission (2010), 'QIS5 Technical Specifications', European Commission webpage.
http://ec.europa.eu/internal_market/insurance/docs/solvency/qis5/201007/technical_specifications_en.pdf

The European Commission (2011), 'Proposal For A Directive Of The European Parliament And Of The Council Amending Directives 2003/71/EC and 2009/138/EC In Respect Of The Powers Of The European Insurance And Occupational Pensions Authority And The European Securities And Markets Authority', EUR-Lex Celex ID 52011PC0008. COM(2011) 8 final 2011/0006 (COD).
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52011PC0008:EN:NOT

The European Parliament and the Council of the European Union (2002*a*), 'Council Directive 2002/13/EC', Official Journal of the European Communities.

The European Parliament and the Council of the European Union (2002*b*), 'Council Directive 2002/83/EC', Official Journal of the European Communities.

The European Parliament and the Council of the European Union (2005), 'Council Directive 2005/1/EC', Official Journal of the European Communities.

The European Parliament and the Council of the European Union (2006*a*), 'Council Directive 2006/48/EC', Official Journal of the European Communities.

The European Parliament and the Council of the European Union (2006*b*), 'Council Directive 2006/49/EC', Official Journal of the European Communities.

The European Parliament and the Council of the European Union (2009), 'Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance', Official

Journal of the European Union.
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32009L0138:EN:NOT

The European Parliament and the Council of the European Union (2011), 'Regulation (EU) No 1094/2010', Official Journal of the European Union.
https://eiopa.europa.eu/fileadmin/tx_dam/files/Legal_Framework/EIOPA-Regulation.pdf

Towers Watson (2012), Insights. Accessed 09/06/2012.
http://www.towerswatson.com/assets/pdf/7170/TW-EU-2012-25988.pdf

Tsanakas, A. (2004), 'Dynamic Capital Allocation with Distortion Risk Measures', *Insurance: Mathematics and Economics* **35**(2), 223–243. Papers presented at the 7th IME Conference.

Vallender, S. (1973), 'Calculation of the Wasserstein distance between probability distributions on the line', *Theory of Probability and its Applications* **18**(4), 784–786.

Vasershtein, L. N. (1969), 'Markov Processes Over Denumerable Products Of Spaces, Describing Large Systems Of Automata', *Problemy Peredachi Informatsii* **5**(3), 64–72.

Vaughan, E. J. and Vaughan, T. (2008), *Fundamentals of Risk and Insurance*, tenth edn., John Wiley & Sons.

Villani, C. (2009), *Optimal Transport Old and New*, Vol. 338 of *Grundlehren der mathematischen Wissenschaften*, Springer, New York.

von Mises, R. (1947), 'On the Asymptotic Distribution of Differentiable Statistical Functions', *The Annals of Mathematical Statistics* **18**(3), 309–348.

Wagner, F. (n.d.), 'Gabler Wirtschaftslexikon, "Solvency I"', online version. version 6.
http://wirtschaftslexikon.gabler.de/Archiv/145567/solvency-i-v6.html

Walker, A. M. (1968), 'A Note on the Asymptotic Distribution of Sample Quantiles', *Journal of the Royal Statistical Society. Series B (Methodological)* **30**(3), 570–575.

Wang, S. (1996), 'Premium Calculation by transforming the layer premium density', *Astin Bulletin* **26**, 71–92.

Wang, S. S., Young, V. R. and Panjer, H. H. (1997), 'Axiomatic Characterization of Insurance Prices', *Insurance: Mathematics and Economics* **21**(2), 173 – 183. in Honor of Prof. J.A. Beekman.

Weis, U. (2009), *Risikomanagement nach ISO 31000*, WEKA Media.

Weisberg, S. (1980), *Applied Linear Regression*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

Wirch, J. L. and Hardy, M. R. (1999), 'A Synthesis of Risk Measures for Capital Adequacy', *Insurance: Mathematics and Economics* **25**(3), 337–347.

Wirch, J. L. and Hardy, M. R. (2003), Ordering of Risk Measures for Capital Adequacy, *in* 'Proceedings of the AFIR Colloquium'.
http://www.actuaries.org/AFIR/Colloquia/Tromsoe/Wirch_Hardy.pdf

Wozabal, D. (2012), Robustifying Convex Risk Measures: A Non-Parametric Approach. Technical Report, Version January 13, 2012, Accessed 01/19/2012. `http://www.optimization-online.org/DB_FILE/2011/11/3238.pdf`

Wüthrich, M. V., Bühlmann, H. and Furrer, H. (2008), *Market-Consistent Actuarial Valuation*, EAA Lecture Notes, Springer, New-York.

Yaari, M. E. (1987), 'The Dual Theory of Choice under Risk', *Econometrica* **55**(1), 95–115.

Zaglauer, K. and Bauer, D. (2008), 'Risk-neutral valuation of participating life insurance contracts in a stochastic ineterest rate environment', *Insurance: Mathematics and Economics* **43**, 29–40.

Ehrenwörtliche Erklärung

Ich versichere an Eides statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe. Die Arbeit hat in dieser oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Essen, den 11.3.2013

(Robin Rühlicke)