TECHNISCHE UNIVERSITÄT WIEN

Dissertation

Advanced Conditional Risk Measurement and Risk Aggregation with Applications to Credit and Life Insurance

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der Naturwissenschaften unter der Leitung von

Univ.-Prof. Dipl.-Math. Dr. rer. nat. habil. Uwe Schmock

Institut für Stochastik und Wirtschaftsmathematik (E105)
Forschungsgruppe Finanz- und Versicherungsmathematik

sowie unter der Betreuung von

Prof. Dr. Pavel V. Shevchenko, MSc BSc

Risk Analytics Group
The Commonwealth Scientific and Industrial Research Organisation

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Jonas Hirz, MSc BSc

Matrikelnummer: 0620033
Wiedner Hauptstraße 8–10
A-1040 Wien

Wien, 19.05.2015

Kurzfassung der Dissertation

Der erste Teil dieser Dissertation beschäftigt sich mit einer sorgfältigen Analyse verschiedener bedingter Risikomaße. Diese sind eine Verallgemeinerung von klassischen Risikomaßen, wie etwa Value at Risk oder expected Shortfall, und können als Basis für dynamisches Risikomanagement verwendet werden. Basierend auf bedingten unteren Quantilen definieren wir Distortion-Risikomaße mittels einer pfadweisen Lebesgue-Stieltjes-Darstellung und geben eine umfangreiche Liste bedingter Eigenschaften an. Bedingter expected Shortfall mit stochastischem Niveau tritt dann als Spezialfall von Distortion-Risikomaßen auf, wobei wir auch eine Definition mit expliziter Dichte und angepasster Indikatorfunktion geben. Letztere Darstellung basiert auf einer verallgemeinerten Definition des bedingten Erwartungswertes mit sigma-integrierbaren Zufallsvariablen. Wir beweisen zusätzliche Eigenschaften und geben weitere alternative Darstellungen für bedingten expected Shortfall. Als dynamisches Risikomaß betrachtet, gilt vor allem die Supermartingaleigenschaft sowie wachsendes Risiko für Submartingale. Gewichteter bedingter expected Shortfall, welcher beta- und alpha-Value-at-Risk miteinschließt, ergibt sich als Spezialfall von bedingten Distortion-Risikomaßen. Darauf aufbauend führen wir Risikobeiträge von gewichtetem bedingten expected Shortfall ein und beweisen mehrere Eigenschaften, wie z.B. bedingte Kohährenz und die Allokation nach Euler. Es ist möglich, Kapitalallokationen sowie Risikobeiträge von Teilportfolios auszurechnen, um den Ursprung der größten Risiken zu identifizieren. Wir geben abschließend einige motivierende Beispiele, wie z.B. eine Anwendung auf diskrete Zeitreihen. Der erste Teil dieser Dissertation bietet, vereinfacht ausgedrückt, eine solide und umfangreiche Analyse von diversen neuen, wie auch bekannten, bedingten Risikomaßen, welche explizit berechnet werden können und somit sowohl für Praktikerinnen und Praktiker, als auch für Forscherinnen und Forscher Verwendung finden.

Der zweite Teil dieser Dissertation beschäftigt sich mit der Entwicklung eines Frameworks zur Schätzung von stochastischen Sterbetafeln und, in weiterer Folge, zur Modellierung von kumulierten Risiken in Kredit-, Pensions- und Lebensversicherungsportfolios, basierend auf einer Erweiterung des Kreditrisikomodells CreditRisk$^+$. Das Ableben jedes einzelnen Versicherungsnehmers wird durch gemeinsame stochastische Risikofaktoren gesteuert, welche direkt mit diversen Todesursachen wie etwa Krebs oder Herz-Kreislauf-Erkrankungen verknüpft werden können. Unser Modell bietet einen äußerst effizienten und numerisch stabilen Algorithmus zur exakten Berechnung der Verlustverteilung des Portfolios zu gegebenen historischen Sterbedaten. Wie von diversen Aufsichtsbehörden gefordert, können Risikomaße wie Value-at-Risk und expected Shortfall dieser Verlustverteilungen leicht berechnet werden. Basierend auf öffentlich zugänglichen Daten entwickeln wir verschiedene Schätzverfahren, wobei, aufgrund der Komplexität des Problems, Markov Chain Monte Carlo (MCMC) von besonderer Bedeutung ist. Basierend auf australischen Daten zeigen wir die Funktionsweise, sowie weitere Anwendungen unseres Modells. Unser Modell erlaubt vor allem die Analyse von Stressszenarien, wodurch Einblicke in den Wirkungsmechanismus diverser unvorhergesehener

Schadensereignisse und die damit einhergehenden Auswirkungen auf Versicherungsleistungen ermöglicht werden. Solche Szenarien können der Ausbruch einer Epidemie, die Verbesserung von medizinischen Behandlungen, sowie die Entwicklung wirkungsvollerer Medikamente sein. Weitere Anwendungen unseres Modells beinhalten die Vorhersage von Sterbewahrscheinlichkeiten und demographischen Verschiebungen.

Abstract

In the first part of this thesis we deal with a detailed analysis of several classes of conditional risk measures which are natural generalisations of classical risk measures such as value at risk or expected shortfall. They provide the basis for an assessment of acceptable risk in a dynamic environment to cover unexpected losses. Based on the upper envelope and conditional lower quantiles, we define conditional distortion risk measures via a pathwise Lebesgue–Stieltjes integral representation and give a collection of different properties. Conditional expected shortfall arises as a special case of conditional distortion risk measures. We also give a definition via an explicit density with adjusted indicator function on a modelling setup involving stochastic levels and generalised conditional expectations based on sigma-integrability. We prove additional properties and give several alternative representations of conditional expected shortfall. Furthermore, we point out the link to dynamic risk measures and show a supermartingale property, as well as the property of prospective increase in uncertainty for submartingales. Weighted conditional expected shortfall, which includes beta- and alpha-value-at-risk, also arises as a special case of conditional distortion risk measures. We then introduce contributions to weighted conditional expected shortfall and prove several properties, including conditional coherence and Euler allocation. It is possible to derive allocation of capital and, in particular, contributions of subportfolios in order to identify main sources of risk. We end with some motivating examples including a time series application. Thus, the first part provides a sound approach and a thorough analysis of some well-known, as well as new classes of conditional risk measures which can be calculated explicitly. Therefore, we provide a useful toolbox for risk measurement, addressing practitioners, as well as scientists working in this field.

In the second part of this thesis, using an extended version of the credit risk model CreditRisk$^+$, we develop a flexible framework to estimate stochastic life tables and to model credit, life insurance and annuity portfolios, including actuarial reserves. Deaths are driven by common stochastic risk factors which may be interpreted as death causes like neoplasms, circulatory diseases or idiosyncratic components. Our approach provides an efficient, numerically stable algorithm for an exact calculation of the one-period loss distribution where various sources of risk are considered. As required by many regulators, we can then derive risk measures for the one-period loss distribution such as value at risk and expected shortfall. Using publicly available data, we provide estimation procedures for model parameters including classical approaches, as well as Markov chain Monte Carlo methods. We conclude with a real world example using Australian death data. In particular, our model allows stress testing and, therefore, offers insight into how certain health scenarios influence annuity payments of an insurer. Such scenarios may include outbreaks of epidemics, improvement in health treatment, or development of better medication. Further applications of our model include modelling of stochastic life tables with corresponding forecasts of death probabilities and demographic changes.

Acknowledgements

Most importantly, I want to thank my supervisor and mentor Univ.-Prof. Dr. Uwe Schmock for all the efforts he put into my education and our work. Not just his brilliant ideas and far-sightedness, also his generosity to let me work on various projects in different fields—ranging from payoff optimisation and credit risk, including applied studies, to risk measurement—as well as to give me the opportunity to attend numerous international conferences brought me this far. All these circumstances made it possible to gather a vast amount of mathematical techniques and to experience alternative ways of thinking. Therefrom, I have undergone an interesting transformation from a learning and mostly 'believing' student to an always critically thinking and mostly 'nothing-believing' researcher. Fruitful collaborations and discussions with other experts added further insights into interesting mathematics. In particular, I want to emphasise the great support of Prof. Dr. Pavel V. Shevchenko, and of course of CSIRO, who was host supervisor of my research visit in Sydney within the 2014 Endeavour Research Fellowship and second supervisor of my thesis. His knowledge and expertise contributed massively to the results obtained in the second part of this thesis. Furthermore, I want to thank Dr. Karin Hirhager for her rich contributions to the first part of this thesis. Our well-beloved secretary Sandra has made work seem far too easy as she has broadly freed me from administrative tasks so that I could really focus on my research. Moreover, my collective thanks go to all other, not previously mentioned, people involved in my PhD study or in this thesis.

Of course, without money from generous sponsors, this work and my entire PhD study would not have been possible. Therefore, especially for the first part of this thesis, my thanks go to the Christian Doppler Research Association (CDG) where I gratefully acknowledge fruitful collaboration and support by msg life (former COR & FJA), Oesterreichische Kontrollbank AG (OeKB) and UniCredit Bank Austria (BA) through CDG and the CD-Laboratory for Portfolio Risk Management (PRisMa Lab) `http://www.prismalab.at/`. Additionally to the financial support by CDG, I gratefully acknowledge financial support from funds of the Oesterreichische Nationalbank (Anniversary Fund, project number: 14977). For the second part of this thesis, I gratefully acknowledge financial support from the Australian Government via the 2014 Endeavour Research Fellowship, as well as again from the Oesterreichische Nationalbank (Anniversary Fund, project number: 14977) and Arithmetica.

Last but not least, I want to thank my whole family for giving me the opportunity to study in the first place, as well as my friends for making life so much more fun. Without their financial, emotional and social support, studying would have certainly been much harder and much less enjoyable.

In conclusion, I want to emphasise that having access to higher education and to a peaceful, self-fulfilling life is still (sadly) a huge privilege in this world and should never be taken for granted.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past few years risk management has become increasingly important in the financial industry, mainly due to new regulatory requirements such as Basel III and Solvency II. On the other hand, the financial crises of 2007 to 2008 quite dramatically demonstrated that risk management tools had often been chosen wrongly, such that tail risks and dependencies had consistently been underestimated. As a consequence, risk management and risk measurement are active fields of mathematical research with numerous unsolved problems and issues to address. This thesis deals with two issues in the context of risk management and is thus divided into two parts. In the first part we provide a sound approach towards various classes of conditional risk measures and give a thorough analysis of mentionable properties. In the second part we then deal with risk aggregation in credit and life insurance portfolios, as well as risks associated to it. In the following few paragraphs, we will introduce the reader to some classical concepts, as well as to problems we address.

1.1 Introduction to Part I: Advanced conditional risk measurement

Before we start with the introduction to conditional risk measurement, we go a step back and briefly discuss the traditional concept of *risk measures* and recall well-known examples, cf. McNeil, Frey and Embrechts [85] for a comprehensive introduction to risk management. A risk measure should quantify the downside risk of a financial position in monetary units such that if this amount of money is added to the position—often called economic capital—the risk is acceptable.[1] In the financial industry, including banks and insurance companies, capital requirements are increasingly regulated. For example, banking regulation under Basel III requires financial institutions to hold capital above some minimum amount, also known as floor capital, to cover unexpected losses. This additional capital requirement is divided into three components: Tier 1, Tier 2 and Tier 3 capital. Which risk measures have to be used for derivation and technical details can be found on the website of the Basel Committee on Banking Supervision.

Many classes of risk measures, mostly based on economical reasoning, have been defined. For example, *coherent risk measures* form an important class and were first introduced

[1] It should be noted that in our framework losses are positive. This assumption is not consistent in the literature and may lead to some changes of signs, inequalities or other minor technicalities.

in Artzner et al. (1997) [7] and further generalised by Artzner et al. (2002) [8], as well as Delbaen [29]. Coherent risk measures are normalised, translation invariant, monotone, subadditive and positively homogeneous, see Definition 1.1. A famous example of a coherent risk measure is *expected shortfall*, see Footnote 10, or sometimes also referred to as average value at risk and closely related to tail value at risk or conditional value at risk. This risk measure can be derived using an average of *value at risk* at different levels. Value at risk itself is a non-coherent risk measure as it is non-subadditive, in general. Subadditivity, or also often referred to as diversification, simply states that the sum of two stand-alone capital requirements is higher than the overall capital requirement, i.e., merging risks never creates extra risk. Note that for heavy-tailed distributions value at risk can strongly violate subadditivity as, for example, outlined in Embrechts et al. [43]. However, under the assumption of elliptical loss distributions, it can be shown that value at risk is coherent, see Embrechts, McNeil and Straumann [40]. Numerous papers deal with the question whether value at risk or expected shortfall is the most relevant risk measure in practical situations with the fewest shortcomings. Replacing subadditivity and positive homogeneity by convexity gives the class of *convex risk measures* which can be equivalently characterised by convex acceptance sets. As convexity is a weaker property compared to subadditivity and positive homogeneity, this class contains coherent risk measures. The motivation behind convexity is that positive homogeneity has been criticised for ignoring liquidity and risk concentration. Later, also the class of *quasiconvex risk measures* was introduced as discussed in Cerreia-Vioglio et al. [21]. This incomplete list of risk measures should just serve the purpose of providing a few references for further reading and a brief introduction to some classical concepts. Further references are given throughout this thesis.

Next, we recall the definition of *conditional risk measures* as a natural extension to classical risk measures which are just using unconditional information. As it is interesting to consider partial and dynamically changing information for a more risk-sensitive approach, many classes of conditional risk measures have been introduced and considered in the literature where, for example, Acciaio, Föllmer and Penner [1], as well as Cheridito, Delbaen and Kupper [23] provide comprehensive introductions. Many contributions to conditional convex risk measures and their representation in terms of conditional expectation can be found in Detlefsen and Scandolo [35]. In particular, conditional expected shortfall has been widely discussed in the literature in the past few years. For example, it is used in the work of McNeil and Frey [84] where, for continuous distribution functions, their representation coincides with Definition 5.3. In the textbook of McNeil, Frey and Embrechts [85] different methods for measuring market risk in the conditional, as well as in the unconditional case are discussed. Also Peracchi and Tanase [92], as well as Leorato et al. [80] deal with estimation of conditional expected shortfall. Very importantly, conditional risk measures can also be used for portfolio selection and portfolio optimisation. Acciaio and Goldammer [2] study these problems based on conditional convex risk measures, in particular, using conditional expected shortfall and the conditional entropic risk measure. Another application of conditional risk measures arises in the context of systemic risk where *spatial risk measures* are introduced, see Föllmer [48]. Within this thesis, we will study several explicitly defined conditional risk measures based on very general settings, including conditional lower quantiles, conditional distortion risk measures, conditional expected shortfall, weighted conditional expected shortfall and contributions to weighted conditional expected shortfall.

Throughout this thesis, we will work on a probability space denoted by $(\Omega, \mathcal{F}, \mathbb{P})$[2] with

[2] We do not denote the probability in the following results.

σ-algebra $\mathcal{F}$ on Ω and probability measure $\mathbb{P}$. All random variables are assumed to be real-valued unless stated otherwise. Let $\mathcal{G}$ and $\mathcal{H}$ denote further sub-σ-algebras with $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$. Filtrations are denoted by $(\mathcal{F}_t)_{t\in[0,\infty)}$. Using suitable embeddings, such a continuous-time setting covers discrete, finite and infinite cases as well. The following definition introduces conditional counterparts of convex and coherent risk measures.

Definition 1.1 (Conditional risk measures)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $\mathcal{L}(\mathbb{P})$ denote a suitable subset of $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$[3] such that the map $\rho[\cdot\,|\,\mathcal{G}]\colon \mathcal{L}(\mathbb{P}) \to L^0(\Omega, \mathcal{G}, \mathbb{P}, \overline{\mathbb{R}}^+)$ is well-defined. Then, ρ is called a *conditional risk measure* given $\mathcal{G}$ if, for all $X, Y \in \mathcal{L}(\mathbb{P})$, it satisfies the following conditional properties:

(a) Conditional normalisation: $\rho[0\,|\,\mathcal{G}] = 0$ a.s.[4]

(b) Conditional translation invariance:[5] If $Z\colon \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable and $X + Z \in \mathcal{L}(\mathbb{P})$, then $\rho[X + Z\,|\,\mathcal{G}] = \rho[X\,|\,\mathcal{G}] + Z$ a.s.

(c) Conditional monotonicity: If $X \leq Y$ a.s., then $\rho[X\,|\,\mathcal{G}] \leq \rho[Y\,|\,\mathcal{G}]$ a.s.

A conditional risk measure is called *conditional convex risk measure* if, in addition, it satisfies the following property:

(d) Conditional convexity: If $Z\colon \Omega \to [0, 1]$ is $\mathcal{G}$-measurable with $X\,Z + Y\,(1 - Z) \in \mathcal{L}(\mathbb{P})$, then $\rho[X\,Z + Y\,(1 - Z)\,|\,\mathcal{G}] \leq \rho[X\,|\,\mathcal{G}]\,Z + \rho[Y\,|\,\mathcal{G}]\,(1 - Z)$ a.s.

A conditional convex risk measure is called *conditional coherent risk measure* if, in addition, it satisfies the following property:

(e) Conditional positive homogeneity: If $Z \geq 0$ a.s. is $\mathcal{G}$-measurable such that $X\,Z \in \mathcal{L}(\mathbb{P})$, then $\rho[X\,Z\,|\,\mathcal{G}] = \rho[X\,|\,\mathcal{G}]\,Z$ a.s.

Remark 1.2. For technical, as well as consistency reasons, we sometimes have to assume that conditional risk measures are defined for infinite losses—thus the use of $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$—and that they map into a space with infinite values. Economically, this can be interpreted as capital requirements for losses which cannot be compensated simply by money anymore. A setting including the value $-\infty$, additionally, is avoided as this leads to complications in many arguments and would correspond to unlimited gains which is not possible in practical situations.

To go one step Furthermore, *dynamical risk measures* can be introduced which deal with the evolution of risk over time. If a filtered probability space is given, the theory of conditional risk measures can be used to consider the evaluation of risk at different moments in time by conditioning on the corresponding σ-algebra, i.e., dynamic risk measures can be interpreted as a sequence of conditional risk measures.

Definition 1.3 (Dynamic risk measures)**.** Given an interval $I \subset [0, \infty)$, as well as a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P})$, consider a suitable subset of random variables $\mathcal{L}(\mathbb{P}) \subset L^0(\Omega, \mathcal{F}, \mathbb{P})$ such that conditional risk measures $\rho_t[\cdot\,|\,\mathcal{F}_t]\colon \mathcal{L}(\mathbb{P}) \to L^0(\Omega, \mathcal{F}_t, \mathbb{P})$, for $t \in I$, are well-defined. Then, $(\rho_t[\cdot\,|\,\mathcal{F}_t])_{t\in I}$ is called *dynamic risk measure.*

[3] The set of equivalence classes of all $\mathbb{P}$-a.s. equal $\mathcal{F}$-measurable random variables with values in $\mathbb{R} \cup \{\infty\}$ is denoted by $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$. $L^0(\Omega, \mathcal{F}, \mathbb{P})$ denotes the set of all elements of $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$ which are real-valued. A 'suitable' subset of $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$ at least covers the set of bounded elements in $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$ or may equal $L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$, depending on the specific risk measure.

[4] Abbreviation for $\mathbb{P}$-almost surely.

[5] Sometimes also referred to as conditional cash invariance.

Dynamic risk measures provide a rich range of possible applications. In particular, they fit with dynamic financial models and thus can be perfectly used for stress testing in risk management. In credit risk, such a consideration of an evolution of acceptable risk corresponds to the term 'point-in-time'. Dynamical settings may lead to results which heavily fluctuate over time. Fluctuations are often reasonable as the riskiness of a portfolio of securities, insurance contracts or credit contracts varies with market fundamentals, such as the current level of market volatility or other sources of uncertainty. The work of Acciaio, Föllmer and Penner [1] provides a useful basis and an extensive collection of references concerning dynamic risk measures in a discrete-time setting. Furthermore, the authors show how to identify a conditional risk measure of a stochastic process with a conditional risk measure of a random variable defined on an appropriate product space. In continuous-time settings, Delbaen [31] provides an extensive and technically sound study of dynamic risk measures. Another important concept for dynamic risk measures which is broadly discussed in the two papers mentioned above is *time-consistency*, see Definition 10.6. The basic idea of time-consistency is that if a financial position is preferable to a benchmark at some future point in time, then it should also be preferable to this benchmark today. Unfortunately, just a limited range of conditional risk measures satisfy strong or weak versions of time-consistency which is why we are aiming for alternative dynamic properties.

In Chapter 3 we start with the definitions of *$\mathcal{G}$-measurable upper envelope* and *conditional lower quantiles* based on the slightly less general definitions as, for example, given in Cheridito and Stadje [24] or McNeil and Frey [84, Section 2]. The latter is a slightly generalised concept of conditional lower quantiles used in quantile regression, see Koenker [74]. Our definitions do not rely on the existence of regular conditional distributions as their existence is not guaranteed in general. For the definition of conditional lower quantiles and other subsequent conditional risk measures, including conditional expected shortfall, we use the notion of *essential suprema* and *essential infima* for which we refer to the textbook of He, Wang and Yan [65, Chapter I.3] or Föllmer and Schied [49, Chapter A.5]. Moreover, we use a general version of conditional expectation which is defined for *σ-integrable* random variables. For the definition of σ-integrability, as well as several basic results we again refer to [65, Chapter I.4]. As in the classical case, note that for existence of conditional expectation it suffices to assume σ-integrability of either X^- or X^+ using Remark 5.6(d).

In Chapter 4 conditional lower quantiles are then used to define *conditional distortion risk measures* via a pathwise Lebesgue–Stieltjes representation. The theory of distortion risk measures goes back to Yaari's dual theory [132] and the axiomatic definition in Wang, Young and Panjer [126]. Our definition is based on the work of Dhaene et al. [37]. We avoid the use of Choquet integrals since our definition fits nicely with the theory of conditional lower quantiles. For completeness note that there exists the concept of spectral risk measures, cf. Acerbi [4] and [62], which is closely related to distortion risk measures. We prove several properties of conditional distortion risk measures including coherence under concavity of the distortion process. Recently, distortion risk measures gained popularity as they are extensively used in conic finance, see Madan and Cherny [81] and related literature. We provide a conditional framework for distortion risk measures which can potentially be used to provide a sound approach for conic finance in a dynamic setting. Moreover, similarly as in Acciaio and Goldammer [2], conditional distortion risk measures can potentially be used for portfolio selection in a dynamic setting. For an approach in the classical, unconditional case see Sereda et al. [112].

In Chapter 5 we define *conditional expected shortfall* via an *adjusted indicator function*. Obtaining the *conditional lower quantile representation*, we see that conditional expected

shortfall arises as a special case of conditional distortion risk measures. Subsequently, we list several properties of conditional expected shortfall which are mostly immediate, as the previously obtained results for conditional distortion risk measures apply. All these properties are based on a general setting including discrete-time and continuous-time models. Some alternative, direct proofs of properties for conditional expected shortfall are given in Section 10.2. Extending the idea of conditional expected shortfall we could also introduce higher moment conditional coherent risk measures where the classical, unconditional case is, for example, analysed in Krokhmal [76]. Moreover, this would give rise to conditional versions of Kusuoka representations of various conditional risk measures, cf. Dentcheva, Penev and Ruszczyński [34]. These topics are not discussed in this thesis. Moreover, we will not discuss elicitable risk measures which were introduced by Bellini and Bignozzi [10].

As another special case of conditional distortion risk measures, in Chapter 6 we introduce *weighted conditional expected shortfall*. It is a weighted integral over all levels of conditional expected shortfall and, therefore, allows a weighted consideration of different levels of risk aversion at the same time. It is a straight-forward generalisation of traditional weighted expected shortfall and, in particular, includes conditional versions of *beta-* and *alpha-value-at-risk*, cf. Cherny and Madan [25, Example 2.9].

The section about weighted conditional expected shortfall is followed by the introduction of *contributions to weighted conditional expected shortfall* in Chapter 7. These contributions give the impact of a subportfolio to weighted conditional expected shortfall of the total portfolio. This is, for example, required if an overall bank capital should be allocated amongst various lines of business or other levels. It is a straight-forward generalisation of traditional contributions to expected shortfall as in Kalkbrener [70] and Schmock [111, Section 7.3]. We prove several properties including subportfolio continuity and a representation by a directional derivative which gives rise to the *Euler allocation* in continuous settings.

In Chapter 8 we give some illustrative examples to show applications of the provided theory. Among other things, we take a look at a time-series example taken from McNeil and Frey [84]. We also give an application to an extended version of the credit risk model CreditRisk$^+$, see Schmock [111, Section 6], which can be used for scenario analysis.

Remarks 1.4. (Preliminary comments).

(a) Note that throughout this thesis losses are assumed to be positive. Thus our results are based on lower quantiles since we do not want to artificially increase losses. These conventions vary in the literature and mainly result in changes in signs, technical details, as well as left- and right-continuity.

(b) Note that some risk measures depend on a given risk level δ. In practical situations δ takes values close to one, e.g., $\delta \in \{0.9, 0.95, 0.995\}$.

(c) Since the level of risk aversion depends on previous developments on the market, δ can be chosen $\mathcal{G}$-measurable. For an example with varying delta involving conditional expected shortfall see Acciaio and Penner [3, Example 1.38(2)].

Given sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, Table 1.1 provides a summary of important conditional properties which are satisfied by different conditional risk measures, including $\mathcal{G}$-measurable upper envelope $X^{\mathcal{G}}$ as given in Chapter 3.1, conditional lower quantiles $q_{\mathcal{G},\delta}(X)$ as given in at Section 3.2, conditional distortion risk measures $\rho_g[X\,|\,\mathcal{G}]$ as given in Section 4.1, conditional expected shortfall $\mathrm{ES}_\delta[X\,|\,\mathcal{G}]$ as given in Section 5.1 and weighted conditional expected shortfall $\mathrm{ES}_G[X\,|\,\mathcal{G}]$ as given in Section 6.1. Details and proofs can be found in respective sections.

Table 1.1: Collection of conditional properties of different conditional risk measures. Answers are meant in a general sense and do not apply to special cases such as $\delta = 0$ a.s.

	$X^{\mathcal{G}}$	$q_{\mathcal{G},\delta}(X)$	$\rho_g[X\,\vert\,\mathcal{G}]$	$\mathrm{ES}_\delta[X\,\vert\,\mathcal{G}]$	$\mathrm{ES}_G[X\,\vert\,\mathcal{G}]$
Coherence	yes	no	no[6]	yes	yes
Convexity	yes	no	no[6]	yes	yes
Com. additivity	yes	yes	yes	yes	yes
Cond. com. additivity	yes	yes	yes	yes	yes
Monotonicity $\leq_{\mathrm{st}(\mathcal{G})}$	yes	yes	yes	yes	yes
Monotonicity $\leq_{\mathrm{icx}(\mathcal{G})}$	yes	yes	no[6]	yes	yes
Law-determined	yes	yes	yes	yes	yes
Regularity	yes	yes	yes	yes	yes
Time-consistency	yes	no[7]	no	no	no
Supermartingale	yes	no	no	yes	yes
Fatou	yes	yes	yes	yes	yes
Continuity from below	yes	yes	yes	yes	yes

To summarise, the main objectives of the first part of this thesis include the following bullet points in the context of conditional and dynamic risk measures:

(a) Give and collect mathematically rigorous and explicit definitions for sometimes intuitively used conditional versions of notions like essential supremum and quantile, distortion risk measure, expected shortfall, weighted expected shortfall and risk contributions. Definitions should be kept as general as possible so that restricting assumptions like boundedness of random variables are not required.

(b) Transfer usual properties to conditional versions.

(c) Prove notable properties of conditional versions.

(d) Find alternatives to time-consistency.

(e) Provide a useful risk management toolbox for practitioners and researchers as conditional and dynamic risk measures provide the basis for an assessment of acceptable risk in a dynamic environment.

1.2 Introduction to Part II: Risk aggregation with applications to credit and life insurance

Risk aggregation of large portfolios in credit, life insurance or related fields typically is a very challenging task due to high computational complexity. Thus, in applications, Monte Carlo is the most commonly used approach to approximate loss distributions of such portfolios as it is easy to implement for all different kinds of stochastic settings but lacks finesse and speed. In this work we propose a new approach to model *aggregated risk in annuity and life insurance portfolios over one period* , as well as a possibility to stochastically model mortality, considering several sources of risk. Coming from credit risk, this model allows flexible handling of dependence structures within a portfolio via common stochastic risk

[6] Yes, for pathwise concave g.

[7] Just middle and weakly acceptance, as well as rejection consistent.

factors. Extensions to multi-period settings are possible but just partially analysed in this thesis. The setting and algorithm used here are based on *extended CreditRisk*$^+$ as introduced in Schmock [111, Section 6]. No simulation is required, which, unlike Monte Carlo, allows a very efficient implementation to derive loss distributions exactly given the input data and the chosen granularity associated with stochastic rounding, see Schmock [111, Section 6.2.2].

Two further observations have led us to the study given in the second part of this thesis. First, life insurers and pension funds usually use deterministic first-order life tables to derive premiums, forecasts, risk measures for portfolios and other related quantities. These first-order life tables are derived from second-order life tables[8] plus artificially added risk margins associated with longevity, size of the company, selection phenomena, estimation and various other sources, see, for example, Pasdika and Wolff [91]. The risk margins described there often lack stochastic foundation and are certainly not consistently appropriate for all companies due to a possibly twisted mix of these risks, see Chapter 13. We are aiming for a unified and stochastically sound approach to tackle these risks. Secondly, we have

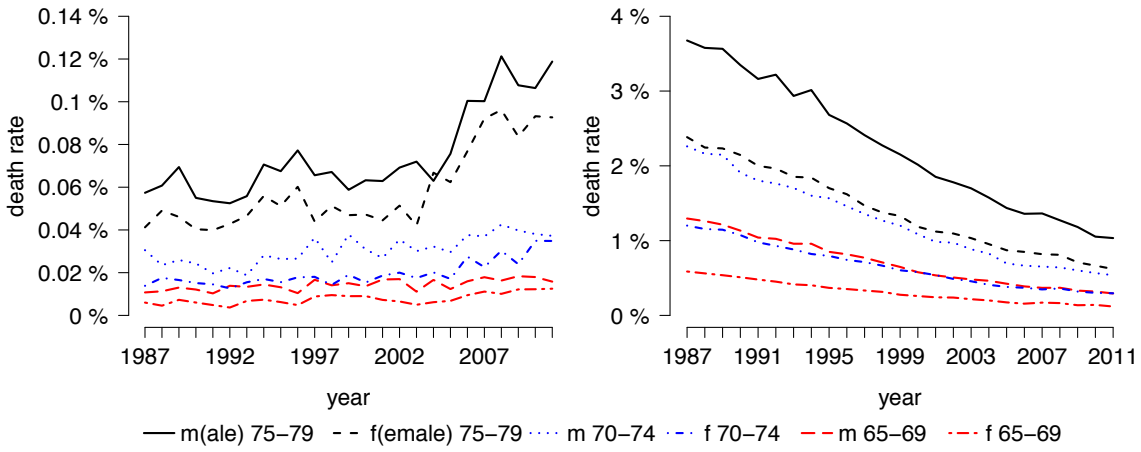


Figure 1.1: Australian death rates for mental and behavioural disorders (left), as well as for circulatory diseases (right) from from 1987 to 2011 for age categories 75–79 years, 70–74 years and 65–69 years, as well as both genders.

observed drastic shifts in death rates due to certain death causes over the past decades. This phenomenon is usually not captured by generation life tables which incorporate only an overall trend in death probabilities. As an illustration of this fact, Figure 1.1 shows death rates based on Australian data[9] for death causes, such as mental and behavioural

[8] Best estimates of the current mortality of a population.

[9] Same data are used in Section 15.1. Annual number of registered deaths in Australia for calender years 1922 to 2011 for different death causes based on the International Statistical Classification of Diseases and Related Health Problems (ICD) is available at the Australian Institute of Health and Welfare (AIHW). There, deaths are categorised by underlying cause of death, i.e., a disease or injury that initiated the train of morbid events leading directly to death. Australian population data are available at the Australian Bureau of Statistics, given annually for June 30 including estimates for births, deaths and migration. In this thesis, motivated by the approach of the AIHW, death rates are then defined as the number of deaths for a given calender year and cause of death divided by the estimated resident population of Australia on June 30 of that year. Due to suitably rich Australian data, this approach suffices and coincides with Definition 11.2. But note that estimation of crude death rates is a delicate issue due to non-constant population and deaths occurring randomly throughout each calender year, cf. Gerber [52]. Death rates do not coincide with death probabilities obtained by some statistical model as death rates always contain statistical fluctuations.

disorders and circulatory diseases, from 1987 to 2011 for various age categories and both genders. Diseases of the circulatory system, such as ischaemic heart disease, have been clearly reduced throughout the past years while death rates due to mental and behavioural disorders, such as dementia, have doubled for older age groups. This observation nicely illustrates the existence of serial dependence amongst different death causes.

In Chapter 11 we develop a framework which stochastically incorporates death probabilities into the model and which, simultaneously, accounts for *longevity risk* in various ways. Longevity risk essentially reflects any potential risk associated with increasing expected future life times of policyholders. This can result in higher than expected annuity payments from an insurer's perspective since policyholders will 'outlive their savings'. Motivated by regulatory risk management standards, our model is used to derive all annuity payments of an insurer for the next period and thus we are, at first glance, not aiming for long-term forecasts or pricing. Annuity payments in our model can range from fixed annual pension payments to variable annuities or index-linked annuity payments with any kind of optionality. Since many approaches and their implementations, in particular Monte Carlo, for deriving loss distributions of large portfolios are very slow, our aim is to provide an alternative, faster, yet flexible approach. Under all these criteria, we choose the application of an extension of a collective risk model, called extended CreditRisk$^+$. As the name suggests, it is a credit risk model used to derive loss distributions of credit portfolios and originates from the classical CreditRisk$^+$ model which was introduced by Credit Suisse First Boston [16] in 1997. Within credit risk models it is classified as a Poisson mixture model. Identifying default with death makes the model perfectly applicable for various kinds of life insurance portfolios and annuity portfolios. For the latter, the situation is more elaborate in contrast to typical credit portfolios as we are interested in the tail of the distribution where only few deaths (defaults) happen. This, together with a special interest in longevity risk, has led us to an argumentation based on annuity portfolios. Nevertheless, generalisations to other life insurance contracts are straight-forward. Extended CreditRisk$^+$ provides a flexible basis for modelling multi-level dependencies and allows a *fast and numerically stable algorithm for risk aggregation*, even in settings with large portfolios. For a more theoretical background, the reader is referred to the Schmock [111] and the references therein. The algorithms described there, originally due to Giese [54] for which Haaf, Reiß and Schoenmakers [63] proved numerical stability, use multivariate *iterated Panjer's recursions*, as well as stochastic rounding for efficient and exact results. The relation to Panjer's recursion was first pointed out by Gerhold, Schmock and Warnung [53, Section 5.5]. Panjer's recursion is an iterative procedure to derive exact distributions of certain random sums, such as Poisson sums, up to a desired cumulative probability. As we are going to see, we are also able to derive value at risk and expected shortfall[10] of the whole portfolio loss for arbitrary levels exactly. In our model, deaths are driven by independent *stochastic risk factors* which are associated with different *underlying causes of death*, see Assumption 12.3, in such a way that variation in

[10] Whenever losses are positive, value at risk (VaR) at level $\delta \in [0,1]$ of a random variable $X\colon \Omega \to \mathbb{R}$ is defined by $q_\delta(X) = \inf\{x \in \mathbb{R} \cup \{\infty\} \,|\, \mathbb{P}(X \le x) \ge \delta\}$ for $\delta > 0$ and $q_0(X) := \inf_{\delta' \in (0,1)} q_{\delta'}(X)$ for $\delta = 0$, i.e., value at risk is the lower δ-quantile of the distribution function of X or the corresponding infimum for $\delta = 0$. Given that losses are positive, expected shortfall at level $\delta \in (0,1)$ of X is then defined by

$$\mathrm{ES}_\delta[X] = \frac{1}{1-\delta}\int_\delta^1 q_t(X)\,dt = \frac{\mathbb{E}\big[X\,1_{X>q_\delta(X)}\big] + q_\delta(X)\big(\mathbb{P}(X \le q_\delta(X)) - \delta\big)}{1-\delta}\,,$$

for $\delta = 0$ by $\mathrm{ES}_0[X] := \inf_{\delta' \in (0,1)} \mathrm{ES}_{\delta'}[X]$, as well as for $\delta = 1$ by $\mathrm{ES}_1[X] := \inf\{z \in \mathbb{R} \cup \{\infty\} \,|\, X \le z \text{ a.s.}\}$. See, for example, [85] or [111] for further information about these risk measures.

these risk factors represents unforeseen changes in mortality, e.g., due to advances in medical treatments or sporadic epidemics. Note that in most cases multiple causes lead to death of a single person, see AIHW [89] for a discussion of this topic. Whilst not analysed in this thesis, multiple death causes are interesting insofar as dependencies amongst various causes can be examined with respect to joint occurrence. Considering a setting based on extended CreditRisk$^+$, the number of deaths of each policyholder is then assumed to be Poisson distributed with stochastic intensity, given risk factors. Thus, serving as an approximation for the true case with single deaths, each person can die multiple times within a period. But, with proper parameter scaling, approximations to the true case with single deaths are very good and final loss distributions are accurate due to *Poisson approximation*, as well as related results, see Barbour, Holst and Janson [9] or Vellaisamy and Chaudhuri [124] and the references therein. Introducing a Poisson mixture distribution for the number of deaths allows derivation of the portfolio loss distribution via iterated Panjer's recursion, as mentioned above. Extended CreditRisk$^+$ even allows for dependent risk factors which makes the model, as well as estimation more involved, see Section 19.1.

Given suitable mortality data, in Chapter 12 we provide several methods to estimate model parameters including *matching of moments*, a *maximum a posteriori approach* and *maximum likelihood*. Death and population data are usually freely available on governmental websites or at statistic bureaus. When using maximum a posteriori and maximum likelihood procedures for our high-dimensional models, standard deterministic numerical optimisation routines are not capable of finding solutions. Thus, we suggest the use of *Markov chain Monte Carlo (MCMC)* methods to derive estimates where we choose the random walk Metropolis–Hastings within Gibbs algorithm. It gives reliable results, is easy to implement and provides an approximation for the posterior distribution of parameters in a Bayesian sense. The usage of MCMC in a real world example is illustrated in Section 15. There, we estimate model parameters for Australian death data which results in a setting with 362 model parameters to be estimated. Results are listed in Section 19.3.

A great advantage of our model is that it automatically incorporates many *different sources of risks*, such as trends, statistical volatility risk and parameter risk, see Chapter 13. These risks are reflected in reduced mortality rates and contribute to the risk of longevity. Effects originating from selection risk within individual companies, as well as structural differences amongst different lines of business[11] are not directly addressed in this thesis as we could not find suitable publicly available portfolio data. Whenever portfolio data are available, Remark 13.1 illustrates an approach towards the incorporation of portfolio data and individual information into our model.

Moreover, our setting with common risk factors allows *scenario analysis* in the sense that we can check impacts on annuity portfolios of unexpectedly higher- or lower-than-expected death rates due to certain underlying causes as outlined in Chapter 14.

In Chapter 16 we illustrate further applications of our model including *mortality* and *population forecasts*. In particular, we compare our model with a one-factor setting to the traditional Lee–Carter model, see Lee and Carter [78], Brouhns, Denuit and Vermunt [17] or Kainhofer, Predota and Schmock [69, Section 4.5.1], and conclude that they both give roughly the same results. We also derive expected future life time for Australians in the year 2013 and observe interesting, unexpected results, as given in Section 19.3.

Chapter 17 briefly illustrates *validation* and *model selection techniques*. Model validation approaches are based on our assumed dependence and independence structures. All tests

[11] Often, clients with a particular risk profile are attracted by specific insurance products.

suggest that the model suitably fits Australian data.

In a nutshell, the model proposed here offers a wide range of applications and has advantages over some other approaches, including the following:

(a) The model provides a flexible risk management tool to derive loss distributions of annuity and life insurance portfolios over one period with a special focus on longevity risk as required by many supervisory authorities. In particular, common stochastic risk factors introduce dependence amongst policyholders.

(b) There exists a numerically stable algorithm for our model to derive loss distributions exactly up to a desired cumulative probability given the input data and the chosen granularity associated with stochastic rounding, see Schmock [111, Section 6.2.2]. Risk measures such as value at risk and expected shortfall can then be easily calculated. All in all, the model ensures high accuracy and fast execution times, simultaneously.

(c) Various sources of longevity risk can be incorporated in the model, including trends, statistical volatility risk and estimation risk.

(d) The concept of common stochastic risk factors allows scenario analysis to show implications of changes in health treatments or other unexpected shifts in death rates.

(e) Further applications of the model include stochastic modelling of population forecasts and life tables which is a big advantage in contrast to point estimates.

Part I

Advanced Conditional Risk Measurement

Chapter 2

Motivating Examples

In the introduction we give some indication why conditional and dynamic risk measurement might be very useful and which shortcomings we want to address. To make these motivations more explicit, we give some introductory examples below and suggest a few desirable properties.

A desired, often-cited property for dynamic risk measures is time-consistency as given in Definition 10.6. But, especially in a continuous-time setting, just very few[12] conditional risk measures satisfy this property, see Kupper and Schachermayer [77, Subsection 1.2]. Therefore, we are aiming for alternative desirable properties of conditional and dynamic risk measures as given in the motivation below.

Given a filtration $(\mathcal{F}_t)_{t\geq 0}$, consider an $(\mathcal{F}_t)_{t\geq 0}$-adapted martingale $(M_t)_{t\geq 0}$, as well as a conditional risk measure $\rho[\cdot\,|\,\mathcal{F}_s]$ with $s \geq 0$. Considering a martingale as a fair game where expectations of the outcome stay constant over time, it is desirable that risk increases with time, i.e., the further we look into the future the higher the risk. Therefore, we want to have an prospective increase in uncertainty, i.e.,

$$\rho[M_{t_1}\,|\,\mathcal{F}_s] = \rho[\mathbb{E}[M_{t_2}\,|\,\mathcal{F}_{t_1}]\,|\,\mathcal{F}_s] \leq \rho[M_{t_2}\,|\,\mathcal{F}_s] \quad \text{a.s.,} \tag{2.1}$$

for all $t_1, t_2 \geq 0$ with $s \leq t_1 \leq t_2$.

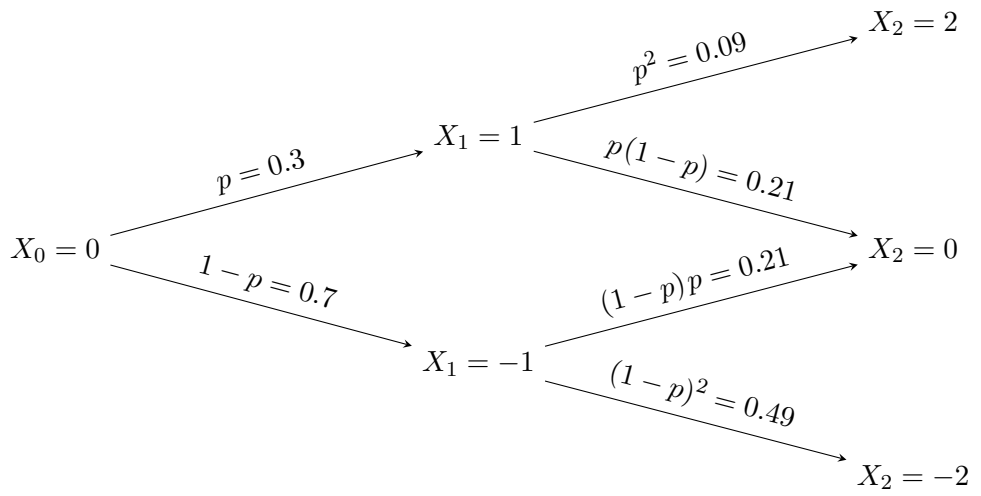
Another desirable property of conditional risk measures in a dynamic setting can easily be motivated. Given a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ with $T > 0$, consider a dynamic risk measure $(\rho_t[\cdot\,|\,\mathcal{F}_t]_t)_{t\in[0,T]}$. Then, for $\mathcal{F}_T$-measurable $X\colon \Omega \to \mathbb{R}$, it is desirable that this dynamic risk measure satisfies the supermartingale property

$$\mathbb{E}[\rho_t[X\,|\,\mathcal{F}_t]\,|\,\mathcal{F}_s] \leq \rho_s[X\,|\,\mathcal{F}_s] \quad \text{a.s., for all } s,t \in [0,T] \text{ with } 0 \leq s \leq t \leq T\,. \tag{2.2}$$

Formulated in words, this means that in expectation risk should decrease the closer we come to maturity as the amount of information increases. Why the converse is not desirable is captured by the following simple example.

Example 2.3 (A binomial lattice)**.** Consider the following two-period binomial lattice (X_0, X_1, X_2) where, starting from zero, we move one unit higher with probability $p = 0.3$ and one unit lower with probability $1 - p = 0.7$, independently of the previous move. Note that losses are assumed to be positive in this thesis.

[12] Conditional expectation, $\mathcal{G}$-measurable upper envelopes, as well as conditional entropic risk measures satisfy time-consistency in a continuous-time setting.



Then, it is immediate that the classical, unconditional lower quantile of X_2 at a 90 percent level gives zero, i.e., indicates no risk at inception. But on the other hand, starting from $X_1 = 1$, the lower quantile of X_2 at a 90 percent level gives two. Analogously, starting from $X_1 = -1$, the lower quantile of X_2 at a 90 percent level gives zero. Thus, in expectation, the lower quantile of X_2 given X_1 gives 0.6 implying that, suddenly, risk is apparent. Hence, it dominates the classical lower quantile and violates (2.2). This example illustrates that a failing supermartingale property can potentially lead to non-identifiable risk at inception.

Remark 2.4 (Motivation I). In a dynamic setting, conditional risk measures should satisfy consistency properties. As just very few conditional risk measures satisfy time-consistency, we are aiming for alternatives such as the supermartingale property, see (2.2), or prospective increase in uncertainty, see (2.1). As we are going to see, both properties are satisfied by conditional expected shortfall and weighted conditional expected shortfall, see Chapter 5 and Chapter 6, respectively.

Recalling Definitions 1.1 and 1.3, we proceed with an example where value at risk is a function of a common stochastic risk factor and, therefore, random. Defining conditional risk measures will get more involved and will require additional techniques once we condition on arbitrary sub-σ-algebras.

Example 2.5 (Different quantiles in scenario analysis, see Section 15.2)**.** Let us consider our annuity model as introduced in Definition 11.11. It is based on the credit risk model extended CreditRisk$^+$, see Schmock [111, Section 6], and basically aggregates losses in credit, life insurance or annuity portfolios. Dependence is introduced via common stochastic risk factors. In the context of life insurance or annuities we identify these stochastic risk factors with different death causes. As outlined in Section 15.2, it is then straight-forward to analyse different mortality scenarios via estimating risk factor realisations and corresponding aggregated losses. High risk factor realisations correspond to increased mortality whilst low risk factor realisations correspond to decreased mortality, reflecting longevity risk. In particular, depending on risk factor realisations, aggregated losses can increase or decrease. More specifically, in Section 15.2 we test the scenario in which deaths due to neoplasms are reduced for all age groups and genders by 25 percent in 2013 which results in an estimated risk factor realisation of $\lambda_2(2013 - t_0) = 0.7991$ with $t_0 = 1986$. As an illustration, we go one step further and test what happens to lower quantiles, i.e., value at risk, see Footnote 10, of

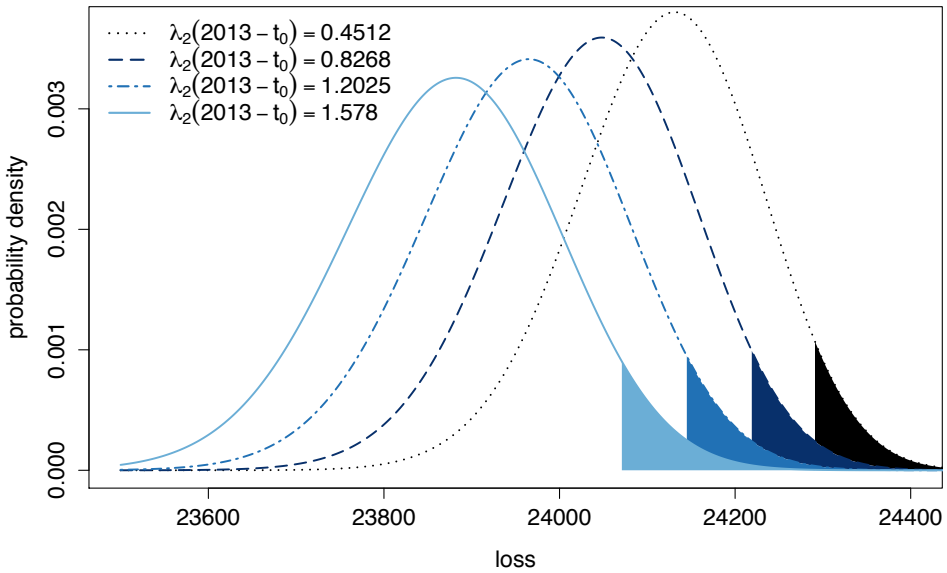


Figure 2.1: Densities and 95 percent quantiles of portfolio loss L^{scen} for different risk factor realisations $\lambda_2(2013 - t_0)$.

the aggregated loss L^{scen} if we vary risk factor realisations. Figure 2.1 shows the density and quantiles at level 95 percent for losses L^{scen} based on different risk factor realisations $\lambda_2(2013 - t_0)$. It nicely illustrates how distributions and lower quantiles shift, given different risk factor realisations. Thus, in this case, we may view lower quantiles as a function of the risk factor for neoplasms, i.e., as a conditional generalisation given this risk factor. In Figure 2.2 we plot lower quantiles at levels 90, 95 and 99 percent as a function of the risk factor for neoplasms $\Lambda_2(2013 - t_0) = \lambda_2(2013 - t_0)$. Since lower risk factor realisations lead to fewer deaths and thus to increased annuity payments from an insurer's perspective, lower quantiles are decreasing in $\lambda_2(2013 - t_0)$.

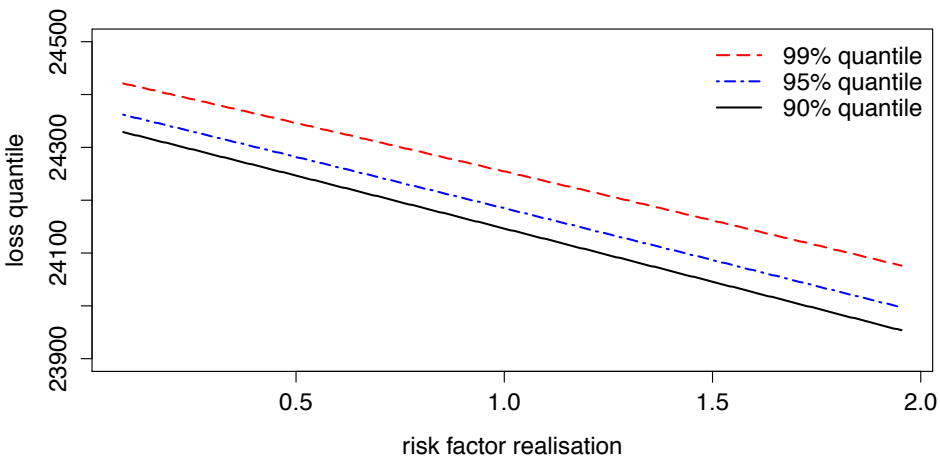


Figure 2.2: Quantiles of portfolio loss L^{scen} at levels 90, 95 and 99 percent as functions of risk factor realisations $\lambda_2(2013 - t_0)$.

Remark 2.6 (Motivation II). We want to give sound definitions of various classes of conditional risk measures which can be calculated explicitly. Moreover, we desire as general definitions as possible with none or just weak integrability conditions.[13]

Example 2.7 (Conditional risk measurement for time series)**.** In the paper of McNeil and Frey [84] an estimation procedure for conditional quantiles and conditional expected shortfall of a heteroscedastic financial return series is presented. There, volatility is modelled via a GARCH model with tails of the innovation distribution being estimated via extreme value theory. Stochasticity of volatility leads to a conditional setting. Via backtesting they show that the approach using conditional risk measures leads to better, more risk sensitive results. We use a similar setting in Example 8.1 where several conditional risk measures are derived explicitly.

Remark 2.8 (Motivation III). Conditional approaches can lead to improved, more risk sensitive risk measurement. This can be of particular interest in the presence of heteroscedasticity within time series.

[13] In the definition of several conditional risk measures, authors often require the existence of regular conditional probabilities which need not necessarily exist, cf. Stoyanov [116, Section 2.4].

Chapter 3

Upper Envelope and Conditional Quantiles

In this chapter we start with the well-known definition of $\mathcal{G}$-measurable upper envelopes and recall several properties, including time-consistency. They are then used in the definition of conditional lower quantiles with stochastic level δ. Conditional lower quantiles with deterministic level, cf. Cheridito and Stadje [24] or McNeil and Frey [84, Section 2], are a a well-understood generalisation of classical lower quantiles. There also exists a pointwise definition of conditional lower quantiles, cf. Acciaio and Goldammer [2], which requires the existence of regular conditional probabilities. Our definition is based on the notion of an essential infimum and is well-defined for all $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$ without any integrability conditions. We then proof notable properties within our generalised setting. Invoking the papers mentioned above, some of these properties are well-known for the less general definition whilst others—like existence of a version of conditional lower quantiles with measurable paths, conditionally comonotonic additivity, Fatou properties or continuity from below—could not be found in previous literature. Some counterexamples then give indications which properties are not satisfied by conditional lower quantiles.

3.1 Upper envelope

We start with the introduction of *$\mathcal{G}$-measurable upper envelopes* for a given sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ which is a coherent and time-consistent conditional risk measure for itself and will occur in the definition of conditional lower quantiles and conditional expected shortfall. The notion of a $\mathcal{G}$-measurable upper envelope is also used in Goldammer and Schmock [57] and is also known as *conditional worst-case risk measure*, see Föllmer [50].

Definition 3.1 ($\mathcal{G}$-measurable upper envelope)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider an $\mathcal{F}$-measurable random variable $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$. Define $X^{\mathcal{G}}$ as the *$\mathcal{G}$-measurable upper envelope* of X, i.e., as the essential infimum of all $\mathcal{G}$-measurable random variables $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathbb{P}(X \leq Z) = 1$.

Example 3.2 (Example 2.3 continued)**.** Recalling Example 2.3 and Definition 3.1, we immediately get that the upper envelope of X_2 given information $\sigma(X_0)$, i.e., at inception,

yields

$$X_2^{\sigma(X_0)} = 2\,.$$

Conversely, the upper of X_2 given information $\sigma(X_1)$, i.e., after one step, yields

$$X_2^{\sigma(X_1)} = \begin{cases} 2 & \text{on } \{X_1 = 1\}\,, \\ 0 & \text{on } \{X_1 = -1\}\,. \end{cases}$$

This nicely illustrates the general pattern that the more information you have the smaller the risk gets, see Remarks 3.3(d).

Remarks 3.3 (Some properties). Given Definition 3.1, for $\mathcal{F}$-measurable random variables $X, Y\colon \Omega \to \mathbb{R} \cup \{\infty\}$ we have the following:

(a) Conditional properties of the $\mathcal{G}$-measurable upper envelope can be found in Lemma 3.18 for the case $\delta = 1$.

(b) Note that $X^{\mathcal{G}}$ is $\mathcal{G}$-measurable and satisfies $X^{\mathcal{G}} \geq X$ a.s. $\mathcal{G}$-measurability of $X^{\mathcal{G}}$ follows by [49, Theorem A.32(a)]. To show the inequality, let $\Phi_{\mathcal{G}}(X)$ be the set of all $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ which are greater than or equal to X a.s. For $Z, Z' \in \Phi_{\mathcal{G}}(X)$ we have $\min\{Z, Z'\} \in \Phi_{\mathcal{G}}(X)$. Thus, by [49, Theorem A.33(b)], we can represent $X^{\mathcal{G}}$ as the a.s. pointwise limit of a decreasing sequence $(Z_n)_{n\in\mathbb{N}}$ in $\Phi_{\mathcal{G}}(X)$. Then, for $n \in \mathbb{N}$, we have $Z_n \geq X$ a.s. which implies $X^{\mathcal{G}} \geq X$ a.s.

(c) Obviously, $X \leq Y$ a.s. implies $X^{\mathcal{G}} \leq Y^{\mathcal{G}}$ a.s.

(d) For sub-σ-algebras $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ we have $X^{\mathcal{G}} \leq X^{\mathcal{H}}$ a.s. since $\Phi_{\mathcal{H}}(X) \subset \Phi_{\mathcal{G}}(X)$, with the notation used above. Thus, if X^- is σ-integrable given $\mathcal{H}$, then $\mathbb{E}[X^{\mathcal{G}}\,|\,\mathcal{H}] \leq X^{\mathcal{H}}$ a.s., i.e., the upper envelope satisfies the supermartingale property[14] in the sense of partial ordering as in Bochner [15].

(e) If X is $\mathcal{G}$-measurable, then $X^{\mathcal{G}} = X$ a.s.

(f) We have $(X+Y)^{\mathcal{G}} \leq X^{\mathcal{G}} + Y^{\mathcal{G}}$ a.s. This follows as $X \leq X^{\mathcal{G}}$ a.s. and $Y \leq Y^{\mathcal{G}}$ a.s. imply $X + Y \leq X^{\mathcal{G}} + Y^{\mathcal{G}}$ a.s. Therefore, $X^{\mathcal{G}} + Y^{\mathcal{G}} \in \Phi_{\mathcal{G}}(X+Y)$.

(g) If $X, Y \geq 0$ a.s., then $(XY)^{\mathcal{G}} \leq X^{\mathcal{G}}Y^{\mathcal{G}}$ a.s. To show this, note that $X \leq X^{\mathcal{G}}$ a.s. and $Y \leq Y^{\mathcal{G}}$ a.s. imply $XY \leq X^{\mathcal{G}}Y^{\mathcal{G}}$ a.s. Therefore, $X^{\mathcal{G}}Y^{\mathcal{G}} \in \Phi_{\mathcal{G}}(XY)$.

(h) For a $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ we have the identity $(X+Z)^{\mathcal{G}} = X^{\mathcal{G}} + Z$ a.s. First note that $(X+Z)^{\mathcal{G}} \leq X^{\mathcal{G}} + Z^{\mathcal{G}} = X^{\mathcal{G}} + Z$ a.s., by (f) and (e). On the other hand, $(X+Z)^{\mathcal{G}} - Z \in \Phi_{\mathcal{G}}(X)$ which implies $X^{\mathcal{G}} \leq (X+Z)^{\mathcal{G}} - Z$ a.s.

(i) For a $\mathcal{G}$-measurable $Z\colon \Omega \to [0,\infty]$ we have $(XZ)^{\mathcal{G}} = X^{\mathcal{G}}Z$ a.s. On the one hand $(XZ)^{\mathcal{G}} \leq X^{\mathcal{G}}Z^{\mathcal{G}} = X^{\mathcal{G}}Z$ a.s., by (e) and (g). On the other hand, define

$$Z^* := \begin{cases} \frac{(XZ)^{\mathcal{G}}}{Z} & \text{on } \{Z > 0\}\,, \\ X^{\mathcal{G}} & \text{on } \{Z = 0\}\,. \end{cases}$$

Then, $Z^* \in \Phi_{\mathcal{G}}(X)$ as $X \leq Z^*$ a.s. which implies $X^{\mathcal{G}} \leq (XZ)^{\mathcal{G}}/Z$ on $\{Z > 0\}$ a.s.[15] On the set $\{Z = 0\}$ the relationship holds a.s. by (e), as $0^{\mathcal{G}} = 0$ a.s.

[14] See Lemma 5.23(n), Corollary 5.27 and Lemma 6.5(m) for the supermartingale property of conditional expected shortfall and weighted conditional expected shortfall.

[15] Throughout this thesis, a property is said to hold on some set H a.s. if it holds on $H \setminus N$, for some $\mathcal{F}$-$\mathbb{P}$-null-set N.

(j) Given another sub-σ-algebra $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, suppose that X^- is σ-integrable with respect to $\mathcal{G}$. Then, $\mathbb{E}[X\,|\,\mathcal{G}]^{\mathcal{H}} \leq X^{\mathcal{H}}$ a.s. because, for every $\mathcal{H}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$, $X \leq Z$ a.s. implies $\mathbb{E}[X\,|\,\mathcal{G}] \leq \mathbb{E}[Z\,|\,\mathcal{G}] = Z$ a.s. Lemma 5.23(o) gives the more general result, termed as supermartingale property, for conditional expected shortfall.

Lemma 3.4 (Coherence)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider Definition 3.1 and set $\mathcal{L}_{\mathcal{G},\mathrm{env}}(\mathbb{P}) := L^0(\Omega, \mathcal{F}, \mathbb{P}, \overline{\mathbb{R}}^+)$. Then, the $\mathcal{G}$-measurable upper envelope is a coherent conditional risk measure on $\mathcal{L}_{\mathcal{G},\mathrm{env}}(\mathbb{P})$.*

Proof. The conditional properties monotonicity, normalisation, subadditivity, translation invariance and positive homogeneity follow by Remark 3.3(c), (e) (f), (h) and (i), respectively. □

Furthermore, we can easily obtain the tower property for upper envelopes. This is a desirable property since this immediately implies time-consistency in a dynamic setting. In particular, the $\mathcal{G}$-measurable upper envelope preserves this property even in a continuous-time framework.

Lemma 3.5 (Tower property)**.** *Given sub-σ-algebras $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, consider an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$. Then, the upper envelope satisfies $(X^{\mathcal{G}})^{\mathcal{H}} = X^{\mathcal{H}}$ a.s.*

Proof. As $X^{\mathcal{G}} \geq X$ a.s. by Remarks 3.3(b), we have $(X^{\mathcal{G}})^{\mathcal{H}} \geq X^{\mathcal{H}}$ a.s. Conversely, invoking Remarks 3.3(d), $X^{\mathcal{G}} \leq X^{\mathcal{H}}$ a.s. and thus, using the notation of Remarks 3.3(b), $X^{\mathcal{H}} \in \Phi_{\mathcal{H}}(X^{\mathcal{G}})$ which gives $(X^{\mathcal{G}})^{\mathcal{H}} \leq X^{\mathcal{H}}$ a.s. □

Remark 3.6 (Time-consistency). Given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P})$, with $I \subset [0, \infty)$, consider $\mathcal{F}$-measurable $X, Y\colon \Omega \to \mathbb{R} \cup \{\infty\}$. Then, the upper envelope is time-consistent, i.e., for any two stopping times $\sigma, \tau\colon \Omega \to I$ with the property $\sigma(\omega) \leq \tau(\omega)$ for all $\omega \in \Omega$ we have $X^{\mathcal{F}_\tau} \leq Y^{\mathcal{F}_\tau}$ a.s. implies $X^{\mathcal{F}_\sigma} \leq Y^{\mathcal{F}_\sigma}$ a.s. This immediately follows since $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$, cf. [72, Lemma 2.15], and by using Remarks 3.3(c), as well as Lemma 3.5, i.e., conditional monotonicity and tower property of the upper envelope. Thus, the upper envelope is also rejection and acceptance consistent, as well as middle and weakly rejection and acceptance consistent, see Definition 10.6. Also note that the tower property is exactly recursiveness as defined in Delbaen [31, Section 6], i.e., $(X^{\mathcal{F}_\sigma})^{\mathcal{F}_\tau} = X^{\mathcal{F}_\tau}$ a.s.

3.2 Definition of conditional lower quantiles

The following definition introduces *conditional lower quantiles* which are straight-forward conditional generalisations of lower quantiles. Cheridito and Stadje [24] as well as McNeil and Frey [84, Section 2] provide a similar conditional concept with deterministic level $\delta \in [0, 1]$ which is closely related to conditional lower quantiles as in our approach. In Jouini and Napp [68] conditional quantiles are introduced under the existence of regular conditional probability measures, but not analysed in detail. Here, we avoid a definition of conditional quantiles based on regular conditional probabilities as their existence heavily depends on the structure of $(\Omega, \mathcal{G})$. In Stoyanov [116, Section 2.4] a simple example is given where no regular conditional probability measure exists. Sufficient conditions for the existence of regular conditional probabilities can be found in Parthasarathy [90, Chapter V.8] and Kallenberg [71, Theorem 6.3]. In statistics, a slightly less general concept of conditional quantiles is used in quantile regression, cf. Koenker [74].

Definition 3.7 (Conditional lower quantiles)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let level $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable. Then, for an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$, the *conditional lower δ-quantile* $q_{\mathcal{G},\delta}(X)$ of X given $\mathcal{G}$ is defined as the essential infimum of all $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathbb{P}(X \leq Z \,|\, \mathcal{G}) \geq \delta$ a.s. on $\{\delta > 0\}$, as well as $\operatorname{ess\,inf}_{\delta' \in (0,1)} q_{\mathcal{G},\delta'}(X)$ on $\{\delta = 0\}$.

Remark 3.8. If we skipped the separate definition of conditional lower quantiles on $\{\delta = 0\}$, then we would get $q_{\mathcal{G},\delta}(X) 1_{\{\delta=0\}} = -\infty 1_{\{\delta=0\}}$ a.s. for all random variables X which leads to inconsistencies in several properties, including normalisation.

Remark 3.9 (Conditional upper quantiles)*.* In a similar fashion as for conditional lower quantiles, we may define *conditional upper δ-quantile* $q^{\mathcal{G},\delta}(X)$ of X given $\mathcal{G}$ as the essential supremum of all $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathbb{P}(X < Z \,|\, \mathcal{G}) \leq \delta$ a.s. on $\{\delta < 1\}$[16], as well as $\operatorname{ess\,sup}_{\delta' \in (0,1)} q^{\mathcal{G},\delta'}(X)$ on $\{\delta = 1\}$. Properties of this risk measure are noted in Remark 3.26.

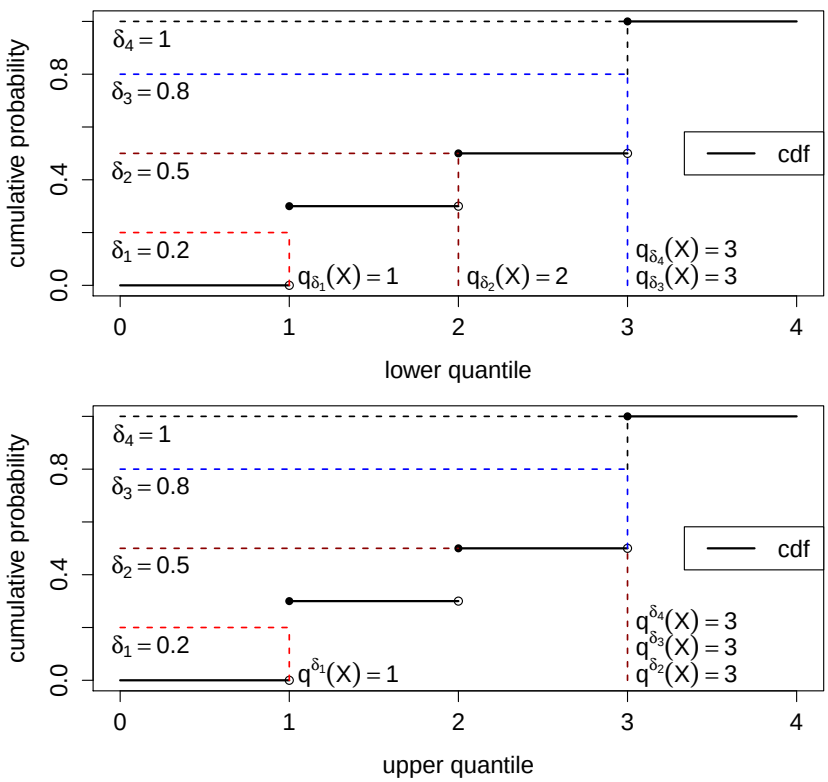


Figure 3.1: Comparison of unconditional lower quantiles (top) and unconditional upper quantiles (bottom) at levels $\delta \in \{0.2, 0.5, 0.8, 1\}$ for the distribution function (cdf) of a discrete random variable X.

Remark 3.10 (Link between lower and upper conditional quantiles)*.* Given the notions of Definition 3.7 and Remark 3.9, note that $q_{\mathcal{G},\delta}(X) = -q^{\mathcal{G},1-\delta}(-X)$ a.s. which follows by Föllmer and Schied [49, Theorem A.34].

[16] Equivalently we can define $q^{\mathcal{G},\delta}(X)$ as the essential supremum of all $\mathcal{G}$-measurable random variables $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathbb{P}(X \leq Z \,|\, \mathcal{G}) \leq \delta$ a.s. or as the essential infimum of all $\mathcal{G}$-measurable random variables $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathbb{P}(X \leq Z \,|\, \mathcal{G}) > \delta$ a.s., on $\{\delta < 1\}$.

Remarks 3.11. (Some remarks on conditional lower quantiles).

(a) Setting $\mathcal{G}$ trivial[17], Definition 3.7 becomes the standard definition lower quantiles or value at risk, cf. McNeil, Frey and Embrechts [85, Section 2.2.2], for all levels $\delta \in (0,1]$ given by $F^{\leftarrow}(\delta) = q_\delta(X) := \inf\{x \in \mathbb{R} \cup \{\infty\} \,|\, F(x) \geq \delta\}$ where F denotes the distribution function of X. At $\delta = 0$ our definition for lower quantiles becomes $F^{\leftarrow}(0) = q_0(X) := \inf_{\delta' \in (0,1)} F^{\leftarrow}(\delta')$ but note that this is not the standard approach in the literature. Basic properties and interesting counterexamples of the generalised inverse can be found in Embrechts and Hofert [39]. For further interesting properties of upper quantiles see the lecture notes of Schmock [111, Section 7.1]. Correspondingly, we define upper quantiles by $F^{\rightarrow}(\delta) = q^\delta(X) := \sup\{x \in \mathbb{R} \cup \{-\infty\} \,|\, \mathbb{P}(X < x) \leq \delta\}$, for $\delta \in [0,1)$, and $F^{\rightarrow}(1) = q_1(X) := \sup_{\delta' \in (0,1)} F^{\rightarrow}(\delta')$. Then, $[0,1] \ni \delta \to F^{\rightarrow}(\delta)$ is right-continuous. In Figure 3.1 some lower and upper quantiles of a discrete cumulative distribution are given. We can observe that they only differ on constant segments of the function and that they coincide on $\delta = 1$ due to the extra definition in that point.

(b) Note that $q_{\mathcal{G},1}(X) = X^{\mathcal{G}}$ a.s. This identity in particular implies that the $\mathcal{G}$-measurable upper envelope also satisfies properties (d) to (h) from Lemma 3.18.

(c) We have $q_{\mathcal{G},\delta}(X) 1_{\{\delta > 0\}} > -\infty$ a.s.

(d) In Bellini, Müller and Roszza [11] generalised quantiles are analysed and used as risk measures. Expectiles arise as special cases of such generalised quantiles.

Example 3.12 (Illustrative example)**.** Given $n = 10$, let $X_1, \ldots, X_n \colon \Omega \to \{0,1\}$ be i.i.d. Bernoulli random variables on $\mathcal{F}$ with $\mathbb{P}(X_1 = 1) = 0.2$, indicating defaults of companies $1, \ldots, n$. Then, clearly, $X := \sum_{i=1}^n X_i$ follows a binomial distribution with parameters $(10, 0.2)$ and gives the total number of defaults of companies $1, \ldots, n$. Moreover, given an event $G \in \mathcal{F}$ with $\mathbb{P}(G) = 0.8$, consider the sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ generated by G, i.e., $\mathcal{G} = \sigma(G)$, where we assume that $X_1, \ldots, X_n$ are conditionally independent given $\mathcal{G}$. The event G represents a joint upgrade in credit ratings of all companies—for example a resolved financial crisis—such that $\mathbb{P}(X_i = 1 \,|\, G) = 0.1$ for all $i \in \{1, \ldots, n\}$. Henceforth, G^{c} represents a joint downgrade in credit ratings—for example due to the outburst of a financial crisis—such that $\mathbb{P}(X_i = 1 \,|\, G^{\mathrm{c}}) = 0.6$ for all $i \in \{1, \ldots, n\}$. Consequently, the lower quantile, equal to value at risk, at level $\delta := 0.95$ of the number of defaults of companies $1, \ldots, n$ is given by

$$q_\delta(X) = \min\{k \in \{0, \ldots, n\} \,|\, \mathbb{P}(X \leq k) \geq \delta\} = 4 \tag{3.13}$$

while the conditional lower quantile at level δ given $\mathcal{G}$, see Section 3.2, is given by

$$q_{\mathcal{G},\delta}(X) = \min\{k \in \{0, \ldots, n\} \,|\, \mathbb{P}(X \leq k \,|\, G) \geq \delta\} = \begin{cases} 3 & \text{on } G\,, \\ 8 & \text{on } G^{\mathrm{c}}\,. \end{cases}$$

Thus, the last expression gives quantiles of the number of defaults given a positive or negative credit event. Comparing this result to the classical, unconditional case in (3.13) illustrates that a positive credit event reduces risk, measured in terms of conditional lower quantiles, whilst a negative increases risk.

[17] $\mathcal{G}$ is called trivial if $\mathbb{P}(G) \in \{0,1\}$ for all $G \in \mathcal{G}$.

3.3 Properties of conditional lower quantiles

In the following, we list several properties of conditional lower quantiles and see that this concept is a natural conditional generalisation of value at risk.

Lemma 3.14. *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$ be $\mathcal{F}$-measurable, as well as let $\delta, \delta'\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable with $\delta \leq \delta'$ a.s. Then, it holds that $\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X) \,|\, \mathcal{G}) \geq \delta$ and $q_{\mathcal{G},\delta}(X) \leq q_{\mathcal{G},\delta'}(X)$, both a.s.*

Proof. The second part is immediate by definition of conditional lower quantiles. For the first argument let $\Phi_{\mathcal{G},\delta}(X)$ be the set of all $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ satisfying $\mathbb{P}(X \leq Z \,|\, \mathcal{G}) \geq \delta$ a.s. For a $Z \in \Phi_{\mathcal{G},\delta}(X)$ we have

$$\mathbb{E}\big[1_{\{X \leq Z\} \cap G}\big] \geq \mathbb{E}[\delta 1_G] \,, \quad G \in \mathcal{G} \,.$$

For $Z, \tilde{Z} \in \Phi_{\mathcal{G},\delta}(X)$, define $Y := \min\{Z, \tilde{Z}\}$ and note that $\{Z \geq \tilde{Z}\} \in \mathcal{G}$. Then,

$$\mathbb{E}\big[1_{\{X \leq Y\} \cap G}\big] = \mathbb{E}\big[1_{\{X \leq Y\}} 1_{\{Z \geq \tilde{Z}\} \cap G}\big] + \mathbb{E}\big[1_{\{X \leq Y\}} 1_{\{Z < \tilde{Z}\} \cap G}\big] \,, \quad G \in \mathcal{G} \,.$$

Taking conditional expectations, we obtain

$$\mathbb{E}\big[1_{\{X \leq Y\} \cap G}\big] \geq \mathbb{E}\big[\delta 1_{\{Z \geq \tilde{Z}\} \cap G}\big] + \mathbb{E}\big[\delta 1_{\{Z < \tilde{Z}\} \cap G}\big] = \mathbb{E}[\delta 1_G] \,, \quad G \in \mathcal{G} \,.$$

Therefore, $Y \in \Phi_{\mathcal{G},\delta}(X)$. Thus, [49, Theorem A.33(b)] can be applied which states that the essential infimum $q_{\mathcal{G},\delta}(X)$ can be represented as an a.s. pointwise limit of a decreasing sequence $(Z_n)_{n \in \mathbb{N}}$ in $\Phi_{\mathcal{G},\delta}(X)$. Thus, for every $n \in \mathbb{N}$,

$$\mathbb{E}\big[1_{\{X \leq Z_n\} \cap G}\big] \geq \mathbb{E}[\delta 1_G] \,, \quad G \in \mathcal{G} \,.$$

As $n \to \infty$, note that $1_{\{X \leq Z_n\}} \searrow 1_{\{X \leq Z\}}$ since $\{X \leq q_{\mathcal{G},\delta}(X)\} = \bigcap_{n \in \mathbb{N}} \{Z_n \geq X\}$, and thus

$$\mathbb{E}\big[1_{\{X \leq Z_n\} \cap G}\big] \searrow \mathbb{E}\big[1_{\{X \leq q_{\mathcal{G},\delta}(X)\} \cap G}\big] \geq \mathbb{E}[\delta 1_G] \,, \quad G \in \mathcal{G} \,,$$

which implies $\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X) \,|\, \mathcal{G}) \geq \delta$ a.s. by definition of conditional expectation. □

Lemma 3.15 (Nice version of conditional lower quantiles)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, define the interval $I := [0, \infty)$. Consider an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R} \cup \{\infty\}$ and let $(\delta_t)_{t \in I}$ be a $(\mathcal{G})_{t \in I}$-adapted*[18] *$[0,1]$-valued process with increasing and left-continuous paths.*[19] *Then, there exists a version of $(q_{\mathcal{G},\delta_t}(X))_{t \in I}$ with increasing and left-continuous paths with corresponding topology on $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.*[20]

Remark 3.16. In particular, this increasing and left-continuous version of conditional lower quantiles is progressively measurable on $(\mathcal{G})_{t \in I}$ and thus $\mathcal{B}(I) \otimes \mathcal{F}$-measurable.

[18] Adapted to the constant filtration $(\mathcal{G})_{t \in I}$.

[19] When speaking about pathwise properties in this thesis, we always assume all paths (and not just almost all paths) to satisfy a certain property in order to avoid problems with joint measurability as, for example, given in [131, Example 7.24].

[20] The classical notions for left-continuity do not apply at infinity. See Embrechts and Hofert [39] for some further remarks on this issue in the context of generalised inverse functions.

Proof of Lemma 3.15. First re-define the process $(q_{\mathcal{G},\delta_t}(X))_{t\in I}$ for every rational $t \in I$ within the corresponding equivalence class of a.s. equal random variables. Define $I_{\mathbb{Q}} := I \cap \mathbb{Q}$ and, for $u, v \in I_{\mathbb{Q}}$ with $u < v$, set

$$\mathcal{N}_{u,v} := \{\omega \in \Omega \,|\, q_{\mathcal{G},\delta_u}(X)(\omega) > q_{\mathcal{G},\delta_v}(X)(\omega)\}\,.$$

By Lemma 3.14, $\mathcal{N}_{u,v}$ is a null set in $\mathcal{G}$ and thus

$$\mathcal{N} := \bigcup_{u\in I_{\mathbb{Q}}} \bigcup_{\substack{v\in I_{\mathbb{Q}}\\ u<v}} \mathcal{N}_{u,v}$$

is a null set with $\mathcal{N} \in \mathcal{G}$ as well. On the null set $\mathcal{N}$ the process $(q_{\mathcal{G},\delta_t}(X))_{t\in I}$ can be defined arbitrarily, e.g. set to zero. For all $t \in I \setminus I_{\mathbb{Q}}$ and $\omega \in \Omega \setminus \mathcal{N}$, define

$$q_{\mathcal{G},\delta_t}(X)(\omega) := \lim_{I_{\mathbb{Q}}\ni u\to t-} q_{\mathcal{G},\delta_u}(X)(\omega) \tag{3.17}$$

where the limit exists since $I_{\mathbb{Q}}$ is dense in I and since $q_{\mathcal{G},\delta_u}(X)(\omega)$ is increasing in u, for all $u \in I_{\mathbb{Q}}$ and $\omega \in \Omega$. Thus, the process we gain has increasing paths which are left-continuous with corresponding topology on $\overline{\mathbb{R}}$. Therefore, this version of the process is progressively measurable on the constant filtration $(\mathcal{G})_{t\in I}$, cf. Karatzas and Shreve [72, Proposition 1.13].

For all $t \in I \setminus I_{\mathbb{Q}}$, we now need to show that, based on (3.17), $q_{\mathcal{G},\delta_t}(X)$ still is a conditional lower quantile. By Lemma 3.14 and the monotone convergence theorem for conditional expectation, see [65, Theorem 1.19(1)], we have, for every $t \in I \setminus I_{\mathbb{Q}}$,

$$\mathbb{P}(X \le q_{\mathcal{G},\delta_t}(X)\,|\,\mathcal{G}) = \lim_{I_{\mathbb{Q}}\ni u\to t-} \mathbb{P}(X \le q_{\mathcal{G},\delta_u}(X)\,|\,\mathcal{G}) \ge \lim_{I_{\mathbb{Q}}\ni u\to t-} \delta_u = \delta_t \quad \text{a.s.}$$

If we now choose a $\mathcal{G}$-measurable random variable Z with properties $Z \le q_{\mathcal{G},\delta_t}(X)$ a.s. and $\mathbb{P}(Z < q_{\mathcal{G},\delta_t}(X)) > 0$, then there exists an $\varepsilon > 0$ such that $B := \{q_{\mathcal{G},\delta}(X) - Z > \varepsilon\}$ satisfies $\mathbb{P}(B) > 0$. Due to the left-continuity of $(q_{\mathcal{G},\delta_t}(X))_{t\in I}$ there exists a $u^* \in I_{\mathbb{Q}}$ with $u^* < t$ such that $\mathbb{P}(B \cap \{q_{\mathcal{G},\delta_t}(X) - q_{\mathcal{G},\delta_{u^*}}(X) < \varepsilon\}) > 0$. Thus, $\mathbb{P}(Z < q_{\mathcal{G},\delta_{u^*}}(X)) > 0$ which implies that Z cannot be a conditional lower δ_{u^*}-quantile of X and thus not a conditional lower δ_{u^*}-quantile of X by Lemma 3.14. Hence, $q_{\mathcal{G},\delta_t}(X)$ defined as in (3.17) is a conditional δ_t-quantile. □

Lemma 3.18 (Properties of conditional lower quantiles)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider $\mathcal{F}$-measurable $X, Y\colon \Omega \to \mathbb{R}$*[21] *and let level $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable. Then, conditional lower quantiles of Definition 3.7 have the following conditional properties*[22]*:*

(a) Conditional normalisation:

$$q_{\mathcal{G},\delta}(0) = 0 \quad a.s.$$

(b) Conditional positive homogeneity: If $Z\colon \Omega \to [0,\infty)$ is $\mathcal{G}$-measurable, then

$$q_{\mathcal{G},\delta}(X\,Z) = q_{\mathcal{G},\delta}(X)\,Z \quad a.s.$$

(c) Conditional translation invariance: If $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ is $\mathcal{G}$-measurable, then

$$q_{\mathcal{G},\delta}(X + Z) = q_{\mathcal{G},\delta}(X) + Z \quad a.s.$$

[21] It can be shown easily that Items (b), (c), (d), (e), (g) with $X \le Y$ a.s., (i), (j) and (k) also hold for $\mathcal{F}$-measurable $X, Y\colon \Omega \to \mathbb{R} \cup \{\infty\}$.

[22] To guarantee consistency, we choose the conventions $\infty \cdot 0 = (-\infty) \cdot 0 = 0 \cdot \infty = 0 \cdot (-\infty) := 0$, as well as $\infty + (-\infty) = -\infty + \infty := \infty$.

(d) Comonotonic additivity: If X and Y are comonotonic, see Definition 10.11, then

$$q_{\mathcal{G},\delta}(X+Y) = q_{\mathcal{G},\delta}(X) + q_{\mathcal{G},\delta}(Y) \quad a.s.$$

(e) Conditionally comonotonic additivity: If X and Y are conditionally comonotonic with respect to $\mathcal{G}$, see Definition 10.11, then

$$q_{\mathcal{G},\delta}(X+Y) = q_{\mathcal{G},\delta}(X) + q_{\mathcal{G},\delta}(Y) \quad a.s.$$

(f) Independence: If X is independent of $\mathcal{G}$, then

$$q_{\mathcal{G},\delta}(X) = F^{\leftarrow}(\delta) \quad a.s.,$$

where $F^{\leftarrow}\colon [0,1] \to \overline{\mathbb{R}}$ denotes the lower quantile function of the distribution function F of X, i.e., $F^{\leftarrow}(y) = \inf\{x \in \overline{\mathbb{R}} \,|\, \mathbb{P}(X \leq x) \geq y\}$ for every $y \in (0,1]$, as well as $F^{\leftarrow}(0) := \inf_{y\in(0,1)} F^{\leftarrow}(y)$.

(g) Monotonicity: If $X \leq_{\mathrm{st}(\mathcal{G})} Y$, see Definition 10.14, then

$$q_{\mathcal{G},\delta}(X) \leq q_{\mathcal{G},\delta}(Y) \quad a.s.$$

In particular, $X \leq_{\mathrm{st}(\mathcal{G})} Y$ is satisfied if $X \leq Y$ a.s.

(h) Determined by conditional law: If $\mathbb{E}[f(X)\,|\,\mathcal{G}] = \mathbb{E}[f(Y)\,|\,\mathcal{G}]$ a.s. for every bounded and continuous function $f\colon \mathbb{R} \to \mathbb{R}$, then

$$q_{\mathcal{G},\delta}(X) = q_{\mathcal{G},\delta}(Y) \quad a.s.$$

(i) Regularity: If $A \in \mathcal{G}$, then $X1_A = Y1_A$ a.s. implies

$$q_{\mathcal{G},\delta}(X)1_A = q_{\mathcal{G},\delta}(Y)1_A \quad a.s.$$

(j) Let $f\colon \mathbb{R}\cup\{\infty\} \to \mathbb{R}\cup\{\infty\}$ be increasing, then

$$q_{\mathcal{G},\delta}(f(X)) \leq f(q_{\mathcal{G},\delta}(X)) \quad a.s.,$$

with equality if f is strictly increasing.

(k) Middle and weak time-consistency: Let $\mathcal{H}$ be a sub-σ-algebra with $\mathcal{H} \subset \mathcal{G}$ and assume that δ and $Z\colon \Omega \to \mathbb{R}$ are $\mathcal{H}$-measurable. Then, $q_{\mathcal{G},\delta}(X) \leq Z$ a.s. implies $q_{\mathcal{H},\delta}(X) \leq Z$ a.s. as well as $q_{\mathcal{G},\delta}(X) \geq Z$ a.s. implies $q_{\mathcal{H},\delta}(X) \geq Z$ a.s., i.e., conditional lower quantiles are middle, thus weakly, acceptance and rejection consistent.

(l) Conditional Fatou I: Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables converging to X in conditional probability[23], i.e., $\lim_{n\to\infty} \mathbb{P}(|X - X_n| \geq \varepsilon\,|\,\mathcal{G}) = 0$ a.s. for every $\varepsilon > 0$. Then,[24]

$$\liminf_{n\to\infty} q_{\mathcal{G},\delta}(X_n) \geq q_{\mathcal{G},\delta}(X) \quad a.s.$$

[23] In particular, almost sure convergence implies convergence in conditional probability by applying conditional bounded convergence, cf. [65, Theorem 1.20]. Also not that if $(X_n)_{n\in\mathbb{N}}$ converges to X in probability, we may always find a subsequence which converges to X in conditional probability, again by conditional bounded convergence.

[24] Note that due to countability of natural numbers $\liminf_{n\to\infty} q_{\mathcal{G},\delta}(X_n)$ is $\mathcal{G}$-measurable and that it is not necessary to use an essential limit inferior.

(m) *Conditional Fatou II: Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of real-valued random variables. Then,* $X := \liminf_{n\to\infty} X_n$ *satisfies*

$$\liminf_{n\to\infty} q_{\mathcal{G},\delta}(X_n) \geq q_{\mathcal{G},\delta}(X) \quad a.s. \tag{3.19}$$

(n) *Continuity from below:*[25] *Let* $(X_n)_{n\in\mathbb{N}}$ *be an increasing sequence of real-valued random variables converging to* X *from below, i.e.,* $X_n \nearrow X$ *a.s. as* $n \to \infty$*. Then,*

$$\lim_{n\to\infty} q_{\mathcal{G},\delta}(X_n) = q_{\mathcal{G},\delta}(X) \quad a.s.$$

Remarks 3.20. (Counterexamples for conditional lower quantiles).

(a) Not a supermartingale. Let X be Bernoulli distributed with parameter $p = \frac{1}{2}$ and define σ-algebras $\mathcal{F} = \mathcal{G} = \sigma(X)$ and $\mathcal{H} = \{\varnothing, \Omega\}$. Then,

$$q_{\mathcal{G},\delta}(X) = X\,, \quad \delta \in [0,1]\,,$$

as well as

$$q_{\mathcal{H},\delta}(X) = \begin{cases} 0 & \text{for } \delta \in \left[0, \frac{1}{2}\right), \\ 1 & \text{for } \delta \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

Thus, $\mathbb{E}[q_{\mathcal{G},\delta}(X)|\mathcal{H}] = \frac{1}{2}$ for all $\delta \in [0,1]$ which implies that the supermartingale property, as given in Lemma 5.23(n) for conditional expected shortfall, does not hold in general for conditional lower quantiles.

(b) Not time-consistent. A counterexample, to show that time-consistency does not hold for a dynamic version of value at risk which is closely related to our concept of conditional lower quantiles, is given in Cheridito and Stadje [24, Example 3.1]. Alternatively, we can prove this by showing that $q_{\mathcal{G},\delta}(X)$ is not recursive, i.e., we give an example where $q_{\mathcal{H},\delta}(q_{\mathcal{G},\delta}(X)) < q_{\mathcal{H},\delta}(X)$. See Delbaen [31, Section 6] for equivalence between time-consistency and recursiveness. Let $\Omega := \{\omega_1, \omega_2, \omega_3, \omega_4\}$ where $\mathcal{F}$ is given by the power set of Ω and where $\mathbb{P}(\{\omega_i\}) := \frac{1}{4}$ for all $i \in \{1,2,3,4\}$. Moreover, let $\mathcal{H} := \{\varnothing, \Omega\}$, as well as $\mathcal{G} := \{\varnothing, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\}$ and let the random variable X be given by

$$X(\omega) = 1_{\{\omega_1\}}(\omega) - 1_{\{\omega_4\}}(\omega)\,, \quad \omega \in \Omega\,.$$

As easily can be seen, conditional lower quantiles satisfy

$$q_{\mathcal{G},\delta}(X)(\omega) = -1_{\{\omega_3,\omega_4\}}(\omega)\,, \quad \delta \in \left[0, \frac{1}{2}\right] \text{ and } \omega \in \Omega\,,$$

as well as

$$q_{\mathcal{H},\delta}(X)(\omega) = 0\,, \quad \delta \in \left(\frac{1}{4}, \frac{3}{4}\right] \text{ and } \omega \in \Omega\,. \tag{3.21}$$

On the other hand,

$$q_{\mathcal{H},\delta}(q_{\mathcal{G},\delta}(X))(\omega) = -1\,, \quad \delta \in \left[0, \frac{1}{2}\right] \text{ and } \omega \in \Omega\,,$$

[25] Modulo some mild technicalities, it is known that the Fatou property as given in (m) and continuity from below are equivalent for conditional convex risk measures, cf. Acciaio and Goldammer [2, Appendix A].

which implies $q_{\mathcal{H},\delta}(q_{\mathcal{G},\delta}(X)) < q_{\mathcal{H},\delta}(X)$ for all $\delta \in \left(\frac{1}{4}, \frac{1}{2}\right]$. This gives a counterexample to rejection consistency and, therefore, to time-consistency of conditional lower quantiles. Moreover, note that

$$\mathbb{E}[X\,|\,\mathcal{G}]\,(\omega) = \frac{1}{2}\left(1_{\{\omega_1,\omega_2\}}(\omega) - 1_{\{\omega_3,\omega_4\}}(\omega)\right), \quad \omega \in \Omega\,,$$

implying that the conditional lower quantile of $\mathbb{E}[X\,|\,\mathcal{G}]$ is given by

$$q_{\mathcal{G},\delta}(\mathbb{E}[X\,|\,\mathcal{G}])(\omega) = \frac{1}{2}\,, \quad \delta \in \left(\frac{1}{2}, 1\right] \text{ and } \omega \in \Omega\,.$$

Thus, by Equation (3.21), we have $q_{\mathcal{G},\delta}(\mathbb{E}[X\,|\,\mathcal{G}]) > q_{\mathcal{G},\delta}(X)$ for $\delta \in \left(\frac{1}{2}, \frac{3}{4}\right]$ which shows that conditional lower quantiles do not satisfy the supermartingale property and the uncertainty decrease in projections as given in Lemma 5.23(n) and (o), in general.

(c) Fatou's lemma for conditional lower quantiles is limited. Consider probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ denotes the Lebesgue–Borel measure on Borel-σ-algebra $\mathcal{B}([0,1])$ and let $\mathcal{G} = \mathcal{F} = \mathcal{B}([0,1])$. Then, define intervals

$$A_n := \left[\frac{k}{2^m}, \frac{k+1}{2^m}\right], \quad n \in \mathbb{N}\,,$$

where $m := \lfloor \log_2(n) \rfloor$ and $k := n - 2^m$. These intervals repeatedly jump through $[0,1]$ and get smaller such that $\lim_{n\to\infty} \mathbb{P}(A_n) \leq \lim_{n\to\infty} \frac{2}{n} = 0$. Then, for all $n \in \mathbb{N}$, define $X_n := 1_{B_n}$ with $B_n := A_n^{\mathrm{c}}$ and set $X := 1$. Clearly, X_n converges to X in probability as $n \to \infty$ since $\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(A_n) \leq \frac{2}{n}$ for all $\varepsilon \in (0,1]$. On the other hand, for every $\mathcal{G}$-measurable $\delta\colon \Omega \to [0,1]$, Lemma 3.18(c) implies

$$\liminf_{n\to\infty} q_{\mathcal{G},\delta}(X_n) = \liminf_{n\to\infty} 1_{B_n} = 0 < X = q_{\mathcal{G},\delta}(X) \quad \text{a.s.}$$

Hence, Fatou's lemma for conditional lower quantiles does not hold under the assumption of convergence in probability.

(d) Not continuous from above. Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ denotes the Lebesgue–Borel measure on Borel-σ-algebra $\mathcal{B}([0,1])$ and let $\mathcal{G} := \{\varnothing, \Omega\}$. Set $\delta = \frac{1}{2}$ and consider a Bernoulli random variable $X(\omega) := 1_{[\delta,1]}(\omega)$ for all $\omega \in \Omega$. Then, for every $n \in \mathbb{N}$ and $\omega \in \Omega$, define $X_n(\omega) := 1_{[\delta_n,1]}(\omega)$ with $\delta_n = \frac{1}{2} - \frac{1}{n+1}$. Note that $X_n \searrow X$ a.s. as $n \to \infty$. But, on the other hand, $q_{\mathcal{G},\delta}(X) = 0$, as well as $q_{\mathcal{G},\delta}(X_n) = 1$ for all $n \in \mathbb{N}$ which implies

$$\lim_{n\to\infty} q_{\mathcal{G},\delta}(X_n) \neq q_{\mathcal{G},\delta}(X)\,.$$

Proof of Lemma 3.18. As a convention, we assume that $\delta > 0$ a.s., since on $\{\delta = 0\}$ all results follow by passing to the essential infimum and taking the cases of $\pm\infty$ into account.

(a) Using the same notation as in the proof of Lemma 3.14, we have $q_{\mathcal{G},\delta}(0) \in \Phi_{\mathcal{G},\delta}(0)$ and, for every $Z \leq 0$ a.s. with $\mathbb{P}(Z < 0) > 0$, we have $q_{\mathcal{G},\delta}(Z) \notin \Phi_{\mathcal{G},\delta}(0)$ which gives the result.

(c) Using the notation as in the proof of Lemma 3.14, we get $q_{\mathcal{G},\delta}(X) + Z \in \Phi_{\mathcal{G},\delta}(X+Z)$ as $\mathbb{P}(X + Z \leq q_{\mathcal{G},\delta}(X) + Z\,|\,\mathcal{G}) = \mathbb{P}(X \leq q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) \geq \delta$ a.s. and as $q_{\mathcal{G},\delta}(X) + Z$ is $\mathcal{G}$-measurable. Therefore,

$$q_{\mathcal{G},\delta}(X+Z) \leq q_{\mathcal{G},\delta}(X) + Z \quad \text{a.s.}$$

Conversely, we have $q_{\mathcal{G},\delta}(X+Z)-Z \in \Phi_{\mathcal{G},\delta}(X)$ as $q_{\mathcal{G},\delta}(X+Z)-Z$ is $\mathcal{G}$-measurable and as $\mathbb{P}(X \le (q_{\mathcal{G},\delta}(X+Z)-Z)\,|\,\mathcal{G}) = \mathbb{P}(X+Z \le q_{\mathcal{G},\delta}(X+Z)\,|\,\mathcal{G}) \ge \delta$ a.s. Therefore,

$$q_{\mathcal{G},\delta}(X) \le q_{\mathcal{G},\delta}(X+Z)-Z \quad \text{a.s.}$$

Altogether, $q_{\mathcal{G},\delta}(X+Z) = q_{\mathcal{G},\delta}(X)+Z$ a.s. On the set $\{\delta=1\}$ the result immediately follows from Remarks 3.3(h).

(b) On $\{Z=0\}$ the result follows by (a). Otherwise, similarly as for (c) we have $q_{\mathcal{G},\delta}(XZ) = q_{\mathcal{G},\delta}(X)Z$ a.s., for $Z \ge 0$ a.s. On $\{\delta=1\}$ we can use Remarks 3.3(i).

(d) As comonotonicity implies conditional comonotonicity with respect to $\mathcal{G}$, as shown in Lemma 10.13, the result follows by (e).

(e) Note that if one side of the equation equals infinity on some set with positive measure, so does the other. Thus, we may assume both sides to be finite a.s. First, let $(q_{\mathcal{G},t}(X))_{t\in[0,1]}$ be the version of the process of conditional lower quantiles with left-continuous and increasing paths as of Lemma 3.15. By extending the probability space, if necessary, we may assume the existence of a random variable U on $(\Omega,\mathcal{F},\mathbb{P})$ which is independent of $\mathcal{G}$ and uniformly distributed on $[0,1]$ meaning that $\mathbb{P}(U \le t) = t$ for all $t \in [0,1]$. Then, by Lemma 10.3(b) and since X and Y are conditionally comonotonic, we get, for all $x,y \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(X \le x, Y \le y\,|\,\mathcal{G}) &= \min\{\mathbb{P}(X \le x\,|\,\mathcal{G}), \mathbb{P}(Y \le y\,|\,\mathcal{G})\}\\ &= \mathbb{P}\big(U \le \min\{\mathbb{P}(X \le x\,|\,\mathcal{G}), \mathbb{P}(Y \le y\,|\,\mathcal{G})\}\,\big|\,\mathcal{G}\big)\\ &= \mathbb{P}\big(U \le \mathbb{P}(X \le x\,|\,\mathcal{G}), U \le \mathbb{P}(Y \le y\,|\,\mathcal{G})\,\big|\,\mathcal{G}\big) \quad \text{a.s.}\end{aligned}$$

Consequently, if we use the result obtained in (5.16) and the corresponding notation, we get, for all $x,y \in \mathbb{R}$,

$$\mathbb{P}(X \le x, Y \le y\,|\,\mathcal{G}) = \mathbb{P}(q_{\mathcal{G},U}(X) \le x, q_{\mathcal{G},U}(Y) \le y\,|\,\mathcal{G}) \quad \text{a.s.} \tag{3.22}$$

Then, fix a $\mathcal{G}$-measurable random variable $Z\colon \Omega \to \mathbb{R}$ and note that

$$\{X+Y \le Z\} = \bigcup_{\substack{p,q,r\in\mathbb{Q}\\ p+q-r\le 0}} \big(\{X \le p\} \cap \{Y \le q\} \cap \{Z \le r\}\big)\,,$$

which, due to the countability of $\mathbb{Q}$, can be rearranged as

$$\{X+Y \le Z\} = \bigcup_{n\in\mathbb{N}} \big(\{X \le p_n\} \cap \{Y \le q_n\} \cap \{Z \le r_n\}\big) =: \bigcup_{n\in\mathbb{N}} A_n(X,Y)\,,$$

where $p_n, q_n, r_n \in \mathbb{Q}$ with $p_n+q_n-r_n \le 0$ for all $n \in \mathbb{N}$. Now, define $B_1(X,Y) := A_1(X,Y)$ and $B_n(X,Y) := A_n(X,Y) \setminus B_{n-1}(X,Y)$ for all $n \in \mathbb{N}$ which is a family of disjoint events such that $\{X+Y \le Z\} = \bigcup_{n\in\mathbb{N}} B_n(X,Y)$. Correspondingly, we get

$$\{q_{\mathcal{G},U}(X)+q_{\mathcal{G},U}(Y) \le Z\} = \bigcup_{n\in\mathbb{N}} B_n(q_{\mathcal{G},U}(X), q_{\mathcal{G},U}(Y))\,.$$

Then, by induction, by (3.22), as well as by the definition of generalised conditional expectation, we get

$$\mathbb{P}(B_n(X,Y)\,|\,\mathcal{G}) = \mathbb{P}\big(B_n(q_{\mathcal{G},U}(X), q_{\mathcal{G},U}(Y))\,\big|\,\mathcal{G}\big) \quad \text{a.s., for all } n \in \mathbb{N}\,.$$

Finally, conditional monotone convergence as given in [65, Theorem 1.19(1)] implies

$$\begin{aligned}\mathbb{P}(X+Y\le Z\,|\,\mathcal{G}) &= \sum_{n\in\mathbb{N}}\mathbb{P}(B_n(X,Y)\,|\,\mathcal{G}) = \sum_{n\in\mathbb{N}}\mathbb{P}\big(B_n(q_{\mathcal{G},U}(X),q_{\mathcal{G},U}(Y))\,\big|\,\mathcal{G}\big)\\ &= \mathbb{P}(q_{\mathcal{G},U}(X)+q_{\mathcal{G},U}(Y)\le Z\,|\,\mathcal{G}) \quad \text{a.s.},\end{aligned} \tag{3.23}$$

for all $\mathcal{G}$-measurable $Z\colon\Omega\to\mathbb{R}$.

Next, we want to show that $Z^* := q_{\mathcal{G},\delta}(X)+q_{\mathcal{G},\delta}(Y)$ is the conditional lower quantile of $X+Y$ with respect to $\mathcal{G}$. Therefore, by (5.16),

$$\begin{aligned}\{q_{\mathcal{G},U}(X)+q_{\mathcal{G},U}(Y)\le Z^*\} &\supset \{q_{\mathcal{G},U}(X)\le q_{\mathcal{G},\delta}(X)\}\cap\{q_{\mathcal{G},U}(Y)\le q_{\mathcal{G},\delta}(Y)\}\\ &= \big\{U\le\min\{\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}),\mathbb{P}(Y\le q_{\mathcal{G},\delta}(Y)\,|\,\mathcal{G})\}\big\}\\ &\supset \{U\le\delta\} \quad \text{a.s.}\end{aligned}$$

which, by (3.23) and by using Lemma 10.3(b) again, implies

$$\mathbb{P}(X+Y\le Z^*\,|\,\mathcal{G})\ge\delta \quad \text{a.s.}$$

Thus, $q_{\mathcal{G},\delta}(X+Y)\le Z^*$ a.s. On the other hand, let $\varepsilon\ge 0$ a.s. be $\mathcal{G}$-measurable such that for $A:=\{\varepsilon>0\}$ we have $\mathbb{P}(A)>0$. Then, by (5.16) again, up to a null set, we have

$$\begin{aligned}&\{q_{\mathcal{G},U}(X)+q_{\mathcal{G},U}(Y)>Z^*-2\varepsilon\}\\ &\qquad\supset \{q_{\mathcal{G},U}(X)>q_{\mathcal{G},\delta}(X)-\varepsilon\}\cap\{q_{\mathcal{G},U}(Y)\le q_{\mathcal{G},\delta}(Y)-\varepsilon\}\\ &\qquad= \big\{U>\max\{\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)-\varepsilon\,|\,\mathcal{G}),\mathbb{P}(Y\le q_{\mathcal{G},\delta}(Y)-\varepsilon\,|\,\mathcal{G})\}\big\}\,.\end{aligned}$$

Since $\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)-\varepsilon\,|\,\mathcal{G})<\delta$ a.s. and $\mathbb{P}(Y\le q_{\mathcal{G},\delta}(Y)-\varepsilon\,|\,\mathcal{G})<\delta$ a.s. on A, we get, using Lemma 10.3(b),

$$\mathbb{P}(X+Y>Z^*-2\varepsilon\,|\,\mathcal{G})>\delta \quad \text{a.s. on } A\,,$$

i.e., $Z^*-2\varepsilon$ cannot be a conditional δ-quantile with respect to $\mathcal{G}$. Hence, $q_{\mathcal{G},\delta}(X+Y)=Z^*$ a.s. which gives the result.

(f) Note that due to Lemma 10.3(b) we can conclude $\mathbb{P}(X\le Z\,|\,\mathcal{G})=F(Z)$, for every $\mathcal{G}$-measurable random variable $Z\colon\Omega\to\mathbb{R}$, where F denotes the distribution function of X. Since $F(x)\ge y$ is equivalent to $F^{\leftarrow}(y)\le x$ for all $x\in\overline{\mathbb{R}}$ and $y\in[0,1]$, cf. McNeil, Frey and Embrechts [85, Appendix A.1.1], we get

$$\begin{aligned}q_{\mathcal{G},\delta}(X) &= \operatorname{ess\,inf}\{Z\colon\Omega\to\mathbb{R}\cup\{\infty\}\text{ is }\mathcal{G}\text{-measurable}\,|\,Z\ge F^{\leftarrow}(\delta)\}\\ &= F^{\leftarrow}(\delta) \quad \text{a.s.}\end{aligned}$$

(g) By Lemma 3.14 and Lemma 10.15 we have

$$\mathbb{P}(X\le q_{\mathcal{G},\delta}(Y)\,|\,\mathcal{G})\ge\mathbb{P}(Y\le q_{\mathcal{G},\delta}(Y)\,|\,\mathcal{G})\ge\delta \quad \text{a.s.},$$

which, similar as in Lemma 3.14, yields the result.

(h) The result immediately follows by Lemma 10.15(a).

(i) This follows from conditional positive homogeneity in (b).

(j) Note that $1_{\{X\le Z\}}\subset 1_{\{f(X)\le f(Z)\}}$. Thus,

$$\mathbb{P}(f(X)\le f(q_{\mathcal{G},\delta}(X))\,|\,\mathcal{G})\ge\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})\ge\delta \quad \text{a.s.},$$

implying that $f(q_{\mathcal{G},\delta}(X)) \geq q_{\mathcal{G},\delta}(f(X))$ a.s. For the case where f is strictly increasing, let Z^* be $\mathcal{G}$-measurable with $Z^* \leq f(q_{\mathcal{G},\delta}(X))$ a.s. and $\mathbb{P}(Z^* < f(q_{\mathcal{G},\delta}(X))) > 0$. Then, $f^{-1}(Z^*) \leq q_{\mathcal{G},\delta}(X)$ a.s. where the inequality is strict with strictly positive probability. Thus,

$$\mathbb{P}\big(\mathbb{P}(X \leq f^{-1}(Z^*)\,|\,\mathcal{G}) \geq \delta\big) = \mathbb{P}\big(\mathbb{P}(f(X) \leq Z^*\,|\,\mathcal{G}) \geq \delta\big) < 1$$

since $q_{\mathcal{G},\delta}(X)$ is the conditional δ-quantile of X which implies that Z^* cannot be a conditional δ-quantile of $f(X)$. Hence, $f(q_{\mathcal{G},\delta}(X)) \leq q_{\mathcal{G},\delta}(f(X))$ which gives the result.

(k) If $q_{\mathcal{G},\delta}(X) \leq Z$ a.s. for $\mathcal{H}$-measurable δ and Z, then

$$\mathbb{P}(X \leq Z\,|\,\mathcal{H}) \geq \mathbb{P}(X \leq q_{\mathcal{G},\delta}(X)\,|\,\mathcal{H}) = \mathbb{E}\big[\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})\,\big|\,\mathcal{H}\big] \geq \delta \quad \text{a.s.},$$

by Lemma 3.14 and properties of conditional expectation, see [65, Chapter I.4]. Thus, we have $q_{\mathcal{H},\delta}(X) \leq Z$ a.s. For the second statement let $q_{\mathcal{G},\delta}(X) \geq Z$ a.s., for $\mathcal{H}$-measurable δ and Z, and assume that there exists an $\mathcal{H}$-measurable set H such that $q_{\mathcal{H},\delta}(X) < Z - \varepsilon$ on H, for a constant $\varepsilon > 0$. Thus, by Lemma 3.14, $\mathbb{P}(X \leq Z - \varepsilon\,|\,\mathcal{H}) \geq \delta$ a.s. on H. On the other hand, as $q_{\mathcal{G},\delta}(X) \geq Z$ a.s. is assumed,

$$\mathbb{P}(X \leq Z - \varepsilon\,|\,\mathcal{H}) \leq \mathbb{E}\big[\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X) - \varepsilon\,|\,\mathcal{G})\,\big|\,\mathcal{H}\big] < \delta \quad \text{a.s. on } H\,.$$

This gives a contradiction to $\mathbb{P}(X \leq Z - \varepsilon\,|\,\mathcal{H}) \geq \delta$ a.s. Hence, H has to be a null set which gives the result.

(l) On $\{q_{\mathcal{G},\delta}(X) = -\infty\}$ the result is immediate. Thus, assume $q_{\mathcal{G},\delta}(X) > -\infty$ a.s. Note that, for all $\mathcal{G}$-measurable $Z_1, Z_2\colon \Omega \to \mathbb{R}$ which satisfy $Z_1 = Z_2 - \varepsilon < Z_2 < q_{\mathcal{G},\delta}(X)$ a.s. and constant $\varepsilon > 0$, we have

$$\mathbb{P}(X_n \leq Z_1\,|\,\mathcal{G}) \leq \mathbb{P}(X \leq Z_2\,|\,\mathcal{G}) + \mathbb{P}(|X - X_n| \geq \varepsilon\,|\,\mathcal{G}) \quad \text{a.s.}$$

Since the last term tends to zero as $n \to \infty$ by assumption, we get, for $\gamma := \mathbb{P}(X \leq Z_2\,|\,\mathcal{G})$,

$$\limsup_{n\to\infty} \mathbb{P}(X_n \leq Z_1\,|\,\mathcal{G}) \leq \gamma < \delta \quad \text{a.s.},$$

where the inequality on the right follows by the minimality of $q_{\mathcal{G},\delta}(X)$. Thus, defining $A_n := \bigcup_{m\geq n}\{\mathbb{P}(X_m \leq Z_1\,|\,\mathcal{G}) > (\delta+\gamma)/2\}$ for $n \in \mathbb{N}$, we have that $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of sets with $\lim_{n\to\infty}\mathbb{P}(A_n) = 0$, as well as $q_{\mathcal{G},\delta}(X_n)1_{A_n^c} \geq Z_1 1_{A_n^c}$ a.s. for all $n \in \mathbb{N}$. Thus, $\liminf_{n\to\infty} q_{\mathcal{G},\delta}(X_n) \geq Z_1$ a.s. Since Z_1 can approach $q_{\mathcal{G},\delta}(X)$ arbitrarily close, the result follows.

(m) Define the sequence $Y_n := X_n \wedge X$ for all $n \in \mathbb{N}$ and note that $\lim_{n\to\infty} Y_n = X$ a.s. Since almost sure convergence implies convergence in conditional probability for all σ-algebras $\mathcal{G}$, see Footnote 23, we can apply (l) to conclude

$$\liminf_{n\to\infty} q_{\mathcal{G},\delta}(Y_n) \geq q_{\mathcal{G},\delta}(X) \quad \text{a.s.}$$

The result then follows by conditional monotonicity of conditional lower quantiles, see (g), since $Y_n \leq X_n$ a.s. for all $n \in \mathbb{N}$.

(n) Note that conditional monotonicity in (g) implies

$$\limsup_{n\to\infty} q_{\mathcal{G},\delta}(X_n) \leq q_{\mathcal{G},\delta}(X) \quad \text{a.s.}$$

Combining this observation with the conditional Fatou property in (l), the result immediately follows since almost sure convergence implies convergence in conditional probability. □

Remark 3.24 ($q_{\mathcal{G},\delta}(X)$ as a conditional risk measure). Given Definition 3.7 and considering levels δ which satisfy $\delta > 0$ a.s., conditional lower quantiles $q_{\mathcal{G},\delta}(\cdot)$ can be seen as conditional risk measures defined on $\mathcal{L}_{\mathcal{G},\text{quant}}(\mathbb{P}) := L^0(\Omega, \mathcal{F}, \mathbb{P})$. However, they are not coherent as subadditivity is even violated in the classical case unless special distributional properties are assumed, see Embrechts, McNeil and Straumann [40].

Remark 3.25. By setting $\mathcal{G}$ trivial, see Footnote 17, all results in Lemma 3.18 correspond to the respective properties of value at risk. In particular, Lemma 3.18(e) corresponds to comonotonic additivity of value at risk as outlined in Pflug [94, Proposition 3(v)].

Remark 3.26 (Properties of conditional upper quantiles). Note that by Remark 3.10 many properties of conditional lower quantiles can immediately be transferred to conditional upper quantiles, modulo some mild adaptions. The first statement of Lemma 3.14 changes to $\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X) \,|\, \mathcal{G}) \leq \delta$ a.s. whereas the second statement remains valid. In Lemma 3.15 we just have to replace 'left-continuous' by 'right-continuous' twice. Properties (a) to (k) of Lemma 3.18 hold one-to-one for conditional upper quantiles whereas for conditional Fatou properties (l) and (m) the conclusions change to $\limsup_{n\to\infty} q_{\mathcal{G},\delta}(X_n) \leq q_{\mathcal{G},\delta}(X)$ a.s. Continuity from below in (n) changes to continuity from above for increasing $X_n \nearrow X$ a.s.

Chapter 4

Conditional Distortion Risk Measures

In this Chapter we define and analyse conditional distortion risk measures which are a natural extension to classical, unconditional distortion risk measures as given, for example, in the work of Dhaene et al. [37, Section 2]. The generalisation of distortion risk measures to a conditional setting has not been studied in the previous literature as far as the authors know. The basic idea is to create a conditional risk measure which takes a weighted average of conditional lower quantiles at various stochastic levels, i.e., a mixture of different levels of risk aversion. Such an approach covers various other conditional risk measures such as conditional expected shortfall. Many properties can immediately be obtained from the corresponding result of conditional lower quantiles. In particular, the Fatou property holds for all conditional distortion risk measures. Considering a distortion process with concave paths gives a conditionally coherent risk measure.

4.1 Definition of conditional distortion risk measures

Motivated by the unconditional case as in Dhaene et al. [37, Section 2], we introduce the class of *conditional distortion risk measures*. Consequently, we create a weighted average of conditional lower quantiles at various levels using a distortion process which is given in Definition 4.1. Distortion processes are a natural extension to distortion functions within the classical approach. As there exists a modification of conditional lower quantiles with left-continuous paths, see Lemma 3.15, the distortion process is assumed to have left-continuous paths as well.

Definition 4.1 (Distortion process)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, a $(\mathcal{G})_{t\in[0,1]}$-adapted process $g\colon [0,1]\times\Omega \to [0,1]$ with increasing, right-continuous paths and boundary conditions $g(0,\cdot) = 0$ a.s., $g(1,\cdot) = 1$ a.s. is called *distortion process*.

Remark 4.2. According to the definition above, each path $[0,1] \ni t \mapsto g(t,\omega)$, with $\omega \in \Omega$, of a distortion process g is in the space $D([0,1])$ equipped with the Skorohod topology, cf. Billingsley [14, Chapter 3]. As paths are right-continuous, a distortion process g is $\mathcal{B}([0,1]) \otimes \mathcal{G}$-measurable with $\mathcal{B}([0,1])$ denoting the Borel σ-algebra on $[0,1]$, cf. Karatzas and Shreve [72, Proposition 1.13].

Definition 4.3 (Suitable subspace)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let g be a distortion process as given in Definition 4.1. By $\mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ we denote the set of all $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$ with

$$\int_{[0,1]} q^-_{\mathcal{G},1-t}(X)\, g(dt,\cdot) < \infty \quad \text{a.s.},$$

where $q^-_{\mathcal{G},1-t}(X) := \max\{0, -q_{\mathcal{G},1-t}(X)\}$ and where $(q_{\mathcal{G},u}(X))_{u\in[0,1]}$ denotes the version of conditional lower quantiles with increasing, left-continuous paths, see Lemma 3.15. By $\mathcal{L}_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ we denote the set of all $X \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ such that

$$\int_{[0,1]} |q_{\mathcal{G},1-t}(X)|\, g(dt,\cdot) < \infty \quad \text{a.s.}$$

Remark 4.4. Integrals in the definitions above and below are pathwise Lebesgue–Stieltjes integral with respect to the measure induced by the paths of the distortion process and, moreover, they are indistinguishable from corresponding stochastic integrals, cf. the textbook of Protter [95, Theorem 17, Section 5, Chapter II]. Of course, tools from general stochastic integration can be used, again see [95, Chapter II].

Definition 4.5 (Conditional distortion risk measures)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let g be a distortion process in the sense of Definition 4.1. Consider an $X \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ and let $(q_{\mathcal{G},u}(X))_{u\in[0,1]}$ denote the version of conditional lower quantiles with increasing and left-continuous paths, see Lemma 3.15. Then, the *conditional g-distortion risk measure* with respect to $\mathcal{G}$ is defined by the pathwise Lebesgue–Stieltjes integral

$$\rho_g[X\,|\,\mathcal{G}] := \int_{[0,1]} q_{\mathcal{G},1-t}(X)\, g(dt,\cdot)\,.$$

Remarks 4.6. (Conditional distortion risk measures).

(a) Conditional distortion risk measures provide a wide and flexible range of useful conditional risk measures. In particular, many different structures of risk aversion can be modelled.

(b) In the unconditional case with trivial $\mathcal{G}$, see Footnote 17, distortion risk measures can be defined via Choquet integrals, cf. Vitali [125], Choquet [27], or Denneberg [33] for theoretical results including the theory of capacities. These theoretical concepts are transferred to risk measurement in the papers of Dhaene et al. [36, Section 5.1] as well as [37, Section 2]. In this thesis we choose a definition using conditional lower quantiles in order to benefit from our previously derived results and to avoid an introduction of a conditional version of Choquet integrals.

(c) As $X \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$, we know that negative parts of the integral are finite while positive parts and, therefore, conditional distortion risk measures may still be infinite.

(d) Fasen and Svejda [46] analyse conditionally consistent multi-period distortion risk measures which, in contrast to conditional distortion risk measures as introduced in this thesis, are based on a different concept.

Remarks 4.7 (Special cases of distortion risk measures). Similar as in the unconditional case, cf. Sereda et al. [112, Section 25.5], conditional lower quantiles and conditional expected shortfall arise as special cases of conditional distortion risk measures. In particular, using

the assumptions of Definition 4.5 and letting $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable, we get the following:

(a) *Conditional lower quantile:* $q_{\mathcal{G},\delta}(X) = \rho_g[X\,|\,\mathcal{G}]$ a.s. with distortion process

$$g(t,\omega) := \begin{cases} 0 & \text{for } 0 \le t < 1-\delta(\omega) \text{ and } \omega \in \Omega\,, \\ 1 & \text{for } 1-\delta(\omega) \le t \le 1 \text{ and } \omega \in \Omega\,. \end{cases}$$

Thus, in general, distortion risk measures are not subadditive.

(b) *Conditional expected shortfall:* Assume that $0 < \delta < 1$ a.s. and consider the quantile representation of conditional expected shortfall as given in Lemma 5.12. Then, we have $\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \rho_g[X\,|\,\mathcal{G}]$ a.s. with distortion process

$$g(t,\omega) := \min\Big\{\frac{t}{1-\delta(\omega)}\,,1\Big\}, \quad t \in [0,1] \text{ and } \omega \in \Omega\,.$$

4.2 Properties of conditional distortion risk measures

In this section we provide a list of various conditional properties of conditional distortion risk measures. In particular, all conditional properties hold in the unconditional case, i.e., for trivial $\mathcal{G}$, see Footnote 17, as well. Most results are easily obtained from respective results of conditional lower quantiles.

Lemma 4.8 (Properties of conditional distortion risk measures)**.** *Recall Definitions 4.1, 4.3 and 4.5. Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and distortion process g and let $X, Y \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$. Then, conditional g-distortion risk measures have the following conditional properties, considering conventions of Footnote 22:*

(a) Conditional normalisation:

$$\rho_g[0\,|\,\mathcal{G}] = 0 \quad \textit{a.s.}$$

(b) Conditional positive homogeneity: If $Z\colon \Omega \to [0,\infty)$ is $\mathcal{G}$-measurable, then

$$\rho_g[X\,Z\,|\,\mathcal{G}] = \rho_g[X\,|\,\mathcal{G}]\,Z \quad \textit{a.s.}$$

(c) Conditional translation invariance: If $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ is $\mathcal{G}$-measurable, then

$$\rho_g[X + Z\,|\,\mathcal{G}] = \rho_g[X\,|\,\mathcal{G}] + Z \quad \textit{a.s.}$$

(d) Comonotonic additivity: If X and Y are comonotonic, see Definition 10.11, then

$$\rho_g[X + Y\,|\,\mathcal{G}] = \rho_g[X\,|\,\mathcal{G}] + \rho_g[Y\,|\,\mathcal{G}] \quad \textit{a.s.}$$

(e) Conditionally comonotonic additivity: If X and Y are conditionally comonotonic with respect to $\mathcal{G}$, see Definition 10.11, then

$$\rho_g[X + Y\,|\,\mathcal{G}] = \rho_g[X\,|\,\mathcal{G}] + \rho_g[Y\,|\,\mathcal{G}] \quad \textit{a.s.}$$

(f) Independence: If X is independent of $\mathcal{G}$, then

$$\rho_g[X\,|\,\mathcal{G}] = \int_{[0,1]} F^{\leftarrow}(1-t)\,g(dt,\cdot) \quad a.s.,$$

where $F^{\leftarrow}\colon [0,1] \to \overline{\mathbb{R}}$ denotes the lower quantile function of the distribution function F of X, i.e., $F^{\leftarrow}(y) = \inf\{x \in \overline{\mathbb{R}}\,|\,\mathbb{P}(X \le x) \ge y\}$ for all $y \in (0,1]$ as well as $F^{\leftarrow}(0) = \inf_{y\in(0,1)} F^{\leftarrow}(y)$.

(g) Conditional monotonicity: If $X \le_{\mathrm{st}(\mathcal{G})} Y$, see Definition 10.14, then

$$\rho_g[X\,|\,\mathcal{G}] \le \rho_g[Y\,|\,\mathcal{G}] \quad a.s.$$

(h) Determined by conditional law: If $\mathbb{E}[f(X)\,|\,\mathcal{G}] = \mathbb{E}[f(Y)\,|\,\mathcal{G}]$ a.s. for every bounded and continuous function $f\colon \mathbb{R} \to \mathbb{R}$, then

$$\rho_g[X\,|\,\mathcal{G}] = \rho_g[Y\,|\,\mathcal{G}] \quad a.s.$$

(i) Regularity: If $A \in \mathcal{G}$, then $X\,1_A = Y\,1_A$ a.s. implies

$$\rho_g[X\,|\,\mathcal{G}]\,1_A = \rho_g[Y\,|\,\mathcal{G}]\,1_A \quad a.s.$$

(j) If $f\colon \mathbb{R} \to \mathbb{R}$ is convex and if $X \in \mathcal{L}_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$, then

$$f(\rho_g[X\,|\,\mathcal{G}]) \le \int_{[0,1]} f(q_{\mathcal{G},1-t}(X))\,g(dt,\cdot) \quad a.s.$$

If in addition f is strictly increasing, then

$$f(\rho_g[X\,|\,\mathcal{G}]) \le \rho_g[f(X)\,|\,\mathcal{G}] \quad a.s.$$

If f is concave, instead, then the reverse inequalities hold where for $\rho_g[f(X)\,|\,\mathcal{G}]$ to exist we require $f(X) \in \mathcal{L}^{-}_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ in our definition.

(k) Conditional Fatou I: Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables bounded from below by some $\mathcal{G}$-measurable $C\colon \Omega \to \mathbb{R}$ converging to X in conditional probability, i.e., $\lim_{n\to\infty} \mathbb{P}(|X - X_n| \ge \varepsilon\,|\,\mathcal{G}) = 0$ a.s. for every $\varepsilon > 0$, see Footnote 23. Then,

$$\liminf_{n\to\infty} \rho_g[X_n\,|\,\mathcal{G}] \ge \rho_g[X\,|\,\mathcal{G}] \quad a.s.$$

(l) Conditional Fatou II: Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables bounded from below by some $\mathcal{G}$-measurable $C\colon \Omega \to \mathbb{R}$. Then, $X := \liminf_{n\to\infty} X_n$ satisfies

$$\liminf_{n\to\infty} \rho_g[X_n\,|\,\mathcal{G}] \ge \rho_g[X\,|\,\mathcal{G}] \quad a.s. \tag{4.9}$$

(m) Continuity from below: Let $(X_n)_{n\in\mathbb{N}}$ be an increasing sequence of real-valued random variables bounded from below by some $\mathcal{G}$-measurable $C\colon \Omega \to \mathbb{R}$ converging to X from below, i.e., $X_n \nearrow X$ a.s., as $n \to \infty$. Then,

$$\lim_{n\to\infty} \rho_g[X_n\,|\,\mathcal{G}] = \rho_g[X\,|\,\mathcal{G}] \quad a.s.$$

(n) Alternative representation with a left-continuous distortion process:

$$\rho_g[X\,|\,\mathcal{G}] = \int_{[0,1]} q_{\mathcal{G},t}(X)\,\bar{g}(dt,\cdot) \quad a.s.$$

where $\bar{g}(t,\cdot) = 1 - g(1-t,\cdot)$ *for all* $t \in [0,1]$.

(o) Derivative representation: If for a.a. $\omega \in \Omega$ *there exists a measurable derivative function* $g'(\cdot,\omega)\colon [0,1] \to [0,\infty)$ *with* $g(t,\omega) = \int_{[0,t]} g'(u,\omega)\,du$ *for all* $t \in [0,1]$*, then*

$$\rho_g[X\,|\,\mathcal{G}] = \int_{[0,1]} q_{\mathcal{G},1-t}(X)\,g'(t,\cdot)\,dt \quad a.s.$$

If in addition the distortion process g *has concave paths, then we have the following:*

(p) Conditional monotonicity under concavity: If either $X \leq_{\mathrm{icx}(\mathcal{G})} Y$ *or* $X \leq_{\mathrm{cx}(\mathcal{G})} Y$*, and if in addition* $X1_{\{\delta=1\}} \leq_{\mathrm{st}(\mathcal{G})} Y1_{\{\delta=1\}}$*, see Definition 10.14, then*

$$\rho_g[X\,|\,\mathcal{G}] \leq \rho_g[Y\,|\,\mathcal{G}] \quad a.s.$$

(q) Subadditivity: If $X + Y \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$*, then*

$$\rho_g[X+Y\,|\,\mathcal{G}] \leq \rho_g[X\,|\,\mathcal{G}] + \rho_g[Y\,|\,\mathcal{G}] \quad a.s.$$

Conversely, if $\mathbb{P}([0,1] \ni t \mapsto g(t,\cdot)$ *is concave*$) < 1$ *and if there exists an* $\mathcal{F}$*-measurable random variable* U *which is uniformly distributed on* $[0,1]$ *and independent of* $\mathcal{G}$[26]*, then there exist* $\mathcal{F}$*-measurable real-valued random variables* X_0 *and* Y_0 *such that, on some set* $C \in \mathcal{F}$ *with* $\mathbb{P}(C) > 0$*, we have*

$$\rho_g[X_0 + Y_0\,|\,\mathcal{G}] > \rho_g[X_0\,|\,\mathcal{G}] + \rho_g[Y_0\,|\,\mathcal{G}] \quad a.s.$$

(r) Conditional convexity: If $Z\colon \Omega \to [0,1]$ *is* $\mathcal{G}$*-measurable with integrability condition* $X\,Z + Y\,(1-Z) \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$*, then*

$$\rho_g[X\,Z + Y\,(1-Z)\,|\,\mathcal{G}] \leq \rho_g[X\,|\,\mathcal{G}]\,Z + \rho_g[Y\,|\,\mathcal{G}]\,(1-Z) \quad a.s.$$

Proof. We choose the approach to directly proof most results using previously derived properties of conditional lower quantiles. Alternatively, most proofs can also be obtained by pathwise reasoning.

Items (a) to (g) follow by the corresponding statement in Lemma 3.18 and by linearity of stochastic or Lebesgue–Stieltjes integrals, respectively. Note that for (d) and (e), using Lemma 3.15 as well as Lemmas 3.18(d) to (e), we have that $(q_{\mathcal{G},t}(X+Y))_{t\in[0,1]}$ and $(q_{\mathcal{G},t}(X) + q_{\mathcal{G},t}(Y))_{t\in[0,1]}$ are indistinguishable. Hence, $X + Y \in \mathcal{L}^-_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ since, for a.a. $\omega \in \Omega$,

$$\big(q_{\mathcal{G},t}(X)(\omega) + q_{\mathcal{G},t}(Y)(\omega)\big)^- \leq 2\max\{q^-_{\mathcal{G},t}(X)(\omega), q^-_{\mathcal{G},t}(Y)(\omega)\}\,, \quad t \in [0,1]\,.$$

Moreover, if one side in the result of (d) or (e) takes the value ∞, so does the other a.s., respectively.

(h) The result follows by Item (b).

[26] Note that such a random variable U always exists on an enlarged probability space.

(i) The result follows by conditional positive homogeneity of conditional lower quantiles, see Lemma 3.18(b).

(j) The first statement follows by pathwise application of Jensen's general inequality on finite spaces, cf. [88, Theorem 1.8.1]. The second statement then follows by Lemma 3.18(j).

(k) By translation invariance from (c), we may assume without loss of generality that X_n is non-negative for every $n \in \mathbb{N}$. Note that due to the pathwise measurability of $[0,1] \ni t \mapsto q_{\mathcal{G},\delta}(X_n)$, for $n \in \mathbb{N}$, the process $[0,1] \ni t \mapsto \liminf_{n\to\infty} q_{\mathcal{G},\delta}(X_n)$ is pathwise measurable as well. Then, a pathwise application of Fatou's lemma as, for example, given in Kallenberg [71, Lemma 1.20] and of Lemma 3.18(l) immediately yield the result.

(l) Define the sequence $Y_n := X_n \wedge X$ for all $n \in \mathbb{N}$ and note that $\lim_{n\to\infty} Y_n = X$ a.s. Since almost sure convergence implies convergence in conditional probability as remarked in Footnote 23, we can apply (k) to conclude

$$\liminf_{n\to\infty} \rho_g[Y_n \,|\, \mathcal{G}] \geq \rho_g[X \,|\, \mathcal{G}] \quad \text{a.s.}$$

The result then follows by conditional monotonicity of conditional lower quantiles, see (g), since $Y_n \leq X_n$ a.s. for all $n \in \mathbb{N}$.

(m) Note that conditional monotonicity in (g) implies

$$\limsup_{n\to\infty} \rho_g[X_n \,|\, \mathcal{G}] \leq \rho_g[X \,|\, \mathcal{G}] \quad \text{a.s.}$$

Combining this observation with the conditional Fatou property in (k), the result immediately follows since almost sure convergence implies convergence in conditional probability.

(n) and (o) These two results follow by a pathwise reasoning with basic properties of Lebesgue–Stieltjes integrals, cf. [20, Chapter 6], where we use the version of conditional lower quantiles $(q_{\mathcal{G},t}(X))_{t\in[0,1]}$ with left-continuous and increasing paths, see Lemma 3.15.

(p) Let $\omega \in \Omega$ be fixed and use the version of conditional lower quantiles $(q_{\mathcal{G},t}(X))_{t\in[0,1]}$ with left-continuous and increasing paths as given in Lemma 3.15. Thus, $(q_{\mathcal{G},t}(X)(\omega))_{t\in[0,1]}$ is a classical lower quantile function in the sense of Remarks 3.11(a) of a distribution function F_ω. To see this, define

$$F_\omega(x) := \sup\{u \in (0,1] \,|\, q_{\mathcal{G},u}(X)(\omega) \leq x\}\,, \quad x \in \overline{\mathbb{R}}\,,$$

with $\sup \varnothing := 0$. Since $q_{\mathcal{G},\delta}(X)(\omega)$ is increasing, F_ω is increasing as well and $F_\omega(-\infty) = 0$ and $F_\omega(\infty) = 1$ which means that F_ω is a distribution function with lower quantile function $q_{\mathcal{G},\cdot}(X)(\omega)$. Thus, we can proceed with pathwise reasoning and use the unconditional case analogously as in [115, Proof of Theorem 2.1]. They note that there exists an increasing, positive and integrable function ϕ_ω such that

$$\bar{g}(t,\omega) = \int_0^t \phi_\omega(u)\, du\,, \quad t \in [0,1)\,,$$

where $\bar{g}$ is given in (n). We can invoke [115, Proof of Theorem 2.1] again to conclude

$$\rho_g[X \,|\, \mathcal{G}](\omega) = \int_{(0,1)} \int_{[u,1)} q_{\mathcal{G},t}(X)(\omega)\, dt\, \phi_\omega(du) + \bar{g}(1-)\, X^{\mathcal{G}} \quad \text{a.s.}$$

Note that the equality above is simply a consequence of properties of Riemann–Stieltjes integrals. Of course, the analogous result is true for Y. Since $X \leq_{\mathrm{icx}(\mathcal{G})} Y$ or, alternatively, $X \leq_{\mathrm{cx}(\mathcal{G})} Y$, Remarks 5.25(d) implies

$$\int_{[u,1)} q_{\mathcal{G},t}(X)\, dt \leq \int_{[u,1)} q_{\mathcal{G},t}(Y)\, dt \quad \text{a.s., for every } u \in (0,1)\,,$$

which then together with the monotonicity of ϕ_ω, i.e., positivity of increments $d\phi_\omega$, yields the result.

(q) With a similar pathwise argumentation as in (p) for fixed $\omega \in \omega$, we can apply the result from the unconditional case, see [33, Theorem 6.3], to conclude that subadditivity holds in the conditional setting if the paths of g are concave. Of course, if the left side takes the value ∞, so does the right side, a.s.

To show the converse statement, we use a similar idea as in [130, Theorem 2.2]. Note that under the stated assumptions, there exist points $0 < \alpha < \beta \leq 1$ and a $t \in (0,1)$ such that for $C := \{g(t\alpha + (1-t)\beta, \cdot) < tg(\alpha, \cdot) + (1-t)g(\beta, \cdot)\}$ we have

$$\mathbb{P}(C) > 0\,. \tag{4.10}$$

Without loss of generality we may assume that $t = \frac{1}{2}$ which can easily be seen by considering the case that g has a discontinuity in $(0,1]$ and by the case that g is continuous on $(0,1]$. Then, set $\eta := \frac{\alpha+\beta}{2}$ and, for U uniformly distributed on $[0,1]$ and independent of $\mathcal{G}$, define

$$X_0 = 1_{\{U \in [0,\eta]\}}$$

and

$$Y_0 = 1_{\{U \in [0,\alpha]\}} + \frac{1}{2} 1_{\{U \in (\eta,\beta]\}}\,.$$

Conditional distributions are, for $x \in \mathbb{R}$, given by $\mathbb{P}(X_0 \leq x \,|\, \mathcal{G}) = (1-\eta)1_{[0,1)}(x) + 1_{[1,\infty)}(x)$, as well as by $\mathbb{P}(Y_0 \leq x \,|\, \mathcal{G}) = (1-\eta)1_{[0,1/2)}(x) + (1-\alpha)1_{[1/2,1)}(x) + 1_{[1,\infty)}(x)$ and finally by $\mathbb{P}(X_0 + Y_0 \leq x \,|\, \mathcal{G}) = (1-\beta)1_{[0,1/2)}(x) + (1-\eta)1_{[1/2,1)}(x) + (1-\alpha)1_{[1,2)}(x) + 1_{[2,\infty)}(x)$, a.s., respectively. Henceforth, by Definition 3.7 and by Lemma 10.3(b), it is straight-forward to calculate conditional lower quantiles of X_0, Y_0 and $X_0 + Y_0$. In particular, these conditional lower quantiles are deterministic step functions. Thus, by computing the Lebesgue–Stieltjes integrals using (n), we get

$$\rho_g[X_0 \,|\, \mathcal{G}] = g(\eta, \cdot)\,, \quad \text{a.s. on } C\,,$$

and

$$\rho_g[Y_0 \,|\, \mathcal{G}] = g(\alpha, \cdot) + \frac{1}{2}\big(g(\eta, \cdot) - g(\alpha, \cdot)\big)\,, \quad \text{a.s. on } C\,,$$

as well as

$$\rho_g[X_0 + Y_0 \,|\, \mathcal{G}] = 2g(\alpha, \cdot) + \big(g(\eta, \cdot) - g(\alpha, \cdot)\big) + \frac{1}{2}\big(g(\beta, \cdot) - g(\eta, \cdot)\big)\,, \quad \text{a.s. on } C\,.$$

Consequently, by (4.10),

$$\rho_g[X_0 + Y_0 \,|\, \mathcal{G}] - \big(\rho_g[X_0 \,|\, \mathcal{G}] + \rho_g[Y_0 \,|\, \mathcal{G}]\big) = \frac{1}{2} g(\beta, \cdot) + \frac{1}{2} g(\beta, \cdot) - g(\eta, \cdot) > 0\,, \quad \text{a.s. on } C\,,$$

which gives the result.

(r) This result follows by the statements in (b) and (q). □

Corollary 4.11 (Coherence)**.** *Conditional distortion risk measures are conditionally coherent if the distortion process has concave paths.*

Remark 4.12 (A little trick). When deriving conditional distortion risk measures, a little trick can sometimes avoid calculation of conditional lower quantiles in the right tail, i.e., conditional lower quantiles at levels close to one. This is particularly interesting for applications in extended CreditRisk$^+$, see Schmock [111, Section 6], where loss distributions are recursively calculated up to a desired cumulative probability, i.e., the right tail of the distribution is not known. First, consider the same assumptions as in Lemma 4.8 where, in addition, X^- is σ-integrable and where we use the version of conditional lower quantiles $(q_{\mathcal{G},t}(X))_{t\in[0,1]}$ with left-continuous and increasing paths as given in Lemma 3.15. Assume that paths of the distortion process g are linear in a surrounding of zero, i.e., there exist $\mathcal{G}$-measurable random variables $Z, U\colon \Omega \to [0,\infty)$, with $0 < U \leq 1$ a.s., such that $g(t,\cdot) = tZ$ a.s. for all $0 \leq t \leq U$. Moreover, assume that $g_Z(t,\omega) := g(t,\omega) - tZ(\omega)$, for $t \in [0,1]$ and $\omega \in \Omega$, has increasing paths for a.a. $\omega \in \Omega$. Then, using the quantile representation of conditional expectation of Lemma 5.21 and a basic property of Lebesgue–Stieltjes integrals, see Carter and van Brunt [20, Theorem 6.1.2], we get

$$\rho_g[X\,|\,\mathcal{G}] = \mathbb{E}[X\,|\,\mathcal{G}]\;Z + \int_{[0,1-U]} q_{\mathcal{G},1-t}(X)\,g_Z(dt,\cdot) \quad \text{a.s.}$$

Alternatively, if $g_Z(t,\omega)$ has decreasing paths for a.a. $\omega \in \Omega$, as it is the case for conditional expected shortfall, then

$$\rho_g[X\,|\,\mathcal{G}] = \mathbb{E}[X\,|\,\mathcal{G}]\;Z - \int_{[0,1-U]} q_{\mathcal{G},1-t}(X)\,(-g_Z)(dt,\cdot) \quad \text{a.s.}$$

In particular, for conditional expected shortfall at $\mathcal{G}$-measurable level $\delta\colon \Omega \to (0,1)$, see Definition 5.3, the above formula simplifies to

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \frac{1}{1-\delta}\left(\mathbb{E}[X\,|\,\mathcal{G}] - \int_{(0,\delta)} q_{\mathcal{G},t}(X)\,dt\right) \quad \text{a.s.}$$

Chapter 5

Conditional Expected Shortfall

In this chapter we give an explicit definition of conditional expected shortfall with stochastic level via an adjusted indicator function, similarly as in Acciaio and Goldammer [2]. Our definition covers all real-valued random variables such that no integrability condition is needed. Notably, we then proof a conditional quantile representation of conditional expected shortfall which immediately gives the link to conditional distortion risk measures. Many properties, most of which are well-known, can then be directly transferred from conditional distortion risk measures. In particular, we prove two dynamic properties—the supermartingale property and prospective increase in uncertainty for submartingales—which give alternatives to time-consistency.

5.1 Definition of conditional expected shortfall

Our next goal is to give an explicit definition of *conditional expected shortfall* using an *adjusted indicator function*. This approach is motivated by the classical, unconditional case as given in Schmock [111, Section 7.2] and can be, for example, found in Acciaio and Goldammer [2]. In particular, it avoids the usage of acceptance sets or limits. It is thus easy to implement in many situations. In a next step, we obtain a conditional quantile representation of conditional expected shortfall which immediately gives the link to conditional distortion risk measures, see Remarks 4.7(b).

Definition 5.1 (Adjusted indicator function)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let level $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable. Then, for an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$ define the *adjusted indicator function* $f_{\mathcal{G},\delta,X}\colon \Omega \to [0,1]$ by

$$f_{\mathcal{G},\delta,X} := 1_{\{X > q_{\mathcal{G},\delta}(X)\}} + \beta_{\mathcal{G},\delta,X} 1_{\{X = q_{\mathcal{G},\delta}(X)\}}$$

where $\beta_{\mathcal{G},\delta,X}\colon \Omega \to [0,1]$ is $\mathcal{G}$-measurable satisfying

$$\beta_{\mathcal{G},\delta,X} := \begin{cases} \frac{\mathbb{P}(X \le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) - \delta}{\mathbb{P}\big(X = q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}\big)} & \text{on the event } \big\{\mathbb{P}\big(X = q_{\mathcal{G},\delta}(X)\,\big|\,\mathcal{G}\big) > 0\big\}, \\ 0 & \text{on the event } \big\{\mathbb{P}\big(X = q_{\mathcal{G},\delta}(X)\,\big|\,\mathcal{G}\big) = 0\big\}. \end{cases}$$

Remarks 5.2. (Adjusted indicator function).
(a) Note that $\beta_{\mathcal{G},\delta,X} \in [0,1]$ a.s. because

$$\mathbb{P}(X < q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) \le \delta \le \mathbb{P}(X \le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) \quad \text{a.s.}$$

Therefore, $f_{\mathcal{G},\delta,X}$ is $[0,1]$-valued a.s.

(b) It holds that $\mathbb{E}[f_{\mathcal{G},\delta,X}\,|\,\mathcal{G}] = 1-\delta$ a.s. because

$$\mathbb{E}[f_{\mathcal{G},\delta,X}\,|\,\mathcal{G}] = \mathbb{P}(X > q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) + \beta_{\mathcal{G},\delta,X}\,\mathbb{P}(X = q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) \quad \text{a.s.}$$

(c) If sub-σ-algebra $\mathcal{G}$ is trivial, see Footnote 17, then we adopt the notation as used in Schmock [111, Section 7.2] and write $f_{\delta,X} := f_{\mathcal{G},\delta,X}$, as well as $\beta_{\delta,X} := \beta_{\mathcal{G},\delta,X}$.

Definition 5.3 (Conditional expected shortfall)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let level $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable. Then, for an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$, *conditional expected shortfall* of X at level δ given $\mathcal{G}$ is defined by

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] := \begin{cases} X^{\mathcal{G}} & \text{on } \{\delta = 1\}\,, \\ \frac{1}{1-\delta}\,\mathbb{E}[f_{\mathcal{G},\delta,X}X\,|\,\mathcal{G}] & \text{on } \{0 < \delta < 1\}\,, \\ \operatorname{ess\,inf}_{\delta' \in (0,1)} \frac{1}{1-\delta'}\,\mathbb{E}[f_{\mathcal{G},\delta',X}X\,|\,\mathcal{G}] & \text{on } \{\delta = 0\}\,, \end{cases}$$

where the adjusted indicator function $f_{\mathcal{G},\delta,X}$ is given by Definition 5.1.

Example 5.4 (Illustrative Example 3.12 continued)**.** If we are interested in expected shortfall and conditional expected shortfall[27] at level $\delta = 0.95$ of the total number of defaults of companies $1,\dots,n$, see Section 5.1, we find

$$\mathrm{ES}_\delta[X] = \frac{1}{1-\delta}\Bigg(\sum_{k=q_\delta(X)+1}^{n} k\,\mathbb{P}(X=k) + \beta_{\delta,X}\,q_\delta(X)\,\mathbb{P}(X = q_\delta(X))\Bigg) = 4.80$$

as well as

$$\begin{aligned} \mathrm{ES}_\delta[X\,|\,\mathcal{G}] &= \frac{1}{1-\delta}\Bigg(\sum_{k=q_{\mathcal{G},\delta}(X)+1}^{n} k\,\mathbb{P}(X=k\,|\,\mathcal{G}) + \beta_{\mathcal{G},\delta,X}\,q_{\mathcal{G},\delta}(X)\,\mathbb{P}(X = q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})\Bigg) \\ &= \begin{cases} 3.29 & \text{on } G\,, \\ 9.05 & \text{on } G^{\mathrm{c}}\,. \end{cases} \end{aligned}$$

Again, this illustrates that depending on the credit rating scenario risk measured in terms of conditional expected shortfall can increase or decrease. Moreover, in this example we observe that classical, unconditional expected shortfall is less than the mean of conditional expected shortfall. This result is called the supermartingale property and is true in general, see Lemma 5.23(n).

Remarks 5.5. (Conditional expected shortfall).

(a) For trivial $\mathcal{G}$, see Footnote 17, the definition above amounts to the standard definition of expected shortfall, cf. Schmock [111, Section 7.2].

(b) For trivial $\mathcal{G}$, see Footnote 17, expected shortfall can be calculated explicitly in many models including collective risk models such as the credit risk model extended CreditRisk$^+$, cf. Schmock [111, Section 7.2]. In particular, using the identity $\mathbb{E}[X\,1_A] = \mathbb{E}[X] - \mathbb{E}[X\,1_{A^{\mathrm{c}}}]$ for a random variable X and a measurable set A, we see that it is not necessary to calculate the right tail of the distribution of X, see Remark 4.12.

[27] Expected shortfall at level δ gives the expected number of defaults exceeding the corresponding δ-quantile divided by $1-\delta$. Its conditional extension is conditional expected shortfall where, in our case, we condition on two different credit rating events.

(c) Acceptance sets and robust representation. Conditional expected shortfall and many other conditional risk measures can be defined using acceptance sets. Under particular continuity conditions, conditional convex risk measures have a robust representation in terms of a minimal penalty function, cf. Acciaio and Penner [3, Chapters 1.2 and 1.3] and the references therein.

Remarks 5.6. (σ-integrability of conditional expected shortfall).

(a) Define the set $M := \{\delta > 0\}$. Then, Definition 5.3 implies that $1_M(f_{\mathcal{G},\delta,X}X)^-$ is σ-integrable with respect to $\mathcal{G}$. To see this, note that $1_M f_{\mathcal{G},\delta,X}X \geq \tilde{X}$ a.s. where $\tilde{X} := 1_M \min\{0, q_{\mathcal{G},\delta}(X)\}$ is $\mathcal{G}$-measurable. Therefore, $\Omega_n := \{|\tilde{X}| \leq n\} \in \mathcal{G}$ for all $n \in \mathbb{N}$, and $\Omega_n \nearrow \Omega$ as $n \to \infty$. Then, $\mathbb{E}[1_{M\cap\Omega_n}(f_{\mathcal{G},\delta,X}X)^-] \leq \mathbb{E}[1_{\Omega_n}|\tilde{X}|] \leq n$ for every $n \in \mathbb{N}$ which implies that $1_M(f_{\mathcal{G},\delta,X}X)^-$ is σ-integrable with respect to $\mathcal{G}$. Hence, conditional expected shortfall is well-defined.

(b) Again, define $M := \{\delta > 0\}$. If $X^{\mathcal{G}} < \infty$ a.s., then Definition 5.3 implies that $1_M f_{\mathcal{G},\delta,X}X$ is σ-integrable with respect to $\mathcal{G}$. To see this, note that $1_M(f_{\mathcal{G},\delta,X}X)^- \leq \tilde{X}$ a.s. where we define $\tilde{X} := -1_M \min\{0, q_{\mathcal{G},\delta}(X)\}$ which is $\mathcal{G}$-measurable. On the other hand, $1_M|f_{\mathcal{G},\delta,X}X| \leq \max\{\tilde{X}, X^{\mathcal{G}}\} := \widehat{X}$ since $(f_{\mathcal{G},\delta,X}X)^+ \leq X^{\mathcal{G}}$. Therefore, for all $n \in \mathbb{N}$, $\Omega_n := \{\widehat{X} \leq n\} \in \mathcal{G}$, as well as $\Omega_n \nearrow \Omega$ as $n \to \infty$. Then, for every $n \in \mathbb{N}$, we have $\mathbb{E}[1_{M\cap\Omega_n}|f_{\mathcal{G},\delta,X}X|] \leq \mathbb{E}[1_{\Omega_n}\widehat{X}] \leq n$ which implies that $1_M f_{\mathcal{G},\delta,X}X$ is σ-integrable with respect to $\mathcal{G}$.

(c) The definition of conditional expected shortfall on $\{\delta = 0\}$ is due to the fact that $\mathbb{E}[f_{\mathcal{G},0,X}X\,|\,\mathcal{G}] = \mathbb{E}[X\,|\,\mathcal{G}]$ need not necessarily exist.

(d) For a non-negative random variable $X \geq 0$ we can write

$$\mathbb{E}[X\,|\,\mathcal{G}] = \operatorname*{ess\,sup}_{n\in\mathbb{N}} \mathbb{E}[\min\{X, n\}\,|\,\mathcal{G}] \quad \text{a.s.}$$

where, for every $n \in \mathbb{N}$, the random variable $\min\{X, n\}$ is σ-integrable with respect to $\mathcal{G}$ as it is even integrable.

Lemma 5.7. *Given sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $X, Y\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and let $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable. Moreover, assume that $Y \geq 0$ a.s. is σ-integrable with respect to $\mathcal{G}$. Define*

$$\mathcal{F}^Y_{\mathcal{G},\delta,X} := \big\{f\colon \Omega \to [0,1] \;\big|\; f \text{ is } \mathcal{F}\text{-measurable and } \mathbb{E}[fY\,|\,\mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}Y\,|\,\mathcal{G}] \text{ a.s.}\big\}$$

where the adjusted indicator function $f_{\mathcal{G},\delta,X}$ is given by Definition 5.1. Then, the following holds:

(a) Conditional optimality of $f_{\mathcal{G},\delta,X}$. If $1_{\{\delta=0\}}X^-$ is σ-integrable with respect to $\mathcal{G}$, then $\mathbb{E}[f_{\mathcal{G},\delta,X}XY\,|\,\mathcal{G}]$ is well-defined with values in $\mathbb{R}\cup\{\infty\}$ and

$$\operatorname*{ess\,sup}_{f\in\mathcal{F}^Y_{\mathcal{G},\delta,X}} \mathbb{E}[fXY\,|\,\mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}XY\,|\,\mathcal{G}] \quad \text{a.s.}$$

(b) If $f^* \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$ satisfies $\mathbb{E}[f^*XY\,|\,\mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}X\,Y\,|\,\mathcal{G}] < \infty$ a.s., then $f^* = f_{\mathcal{G},\delta,X}$ a.s. on the event $\{Y > 0,\, X \neq q_{\mathcal{G},\delta}(X)\}$.

(c) If $0 < \delta < 1$ a.s. as well as if $f_{\mathcal{G},\delta,X} X$ and Y are conditionally uncorrelated[28] *given $\mathcal{G}$, then*

$$\mathbb{E}[f_{\mathcal{G},\delta,X}\, XY \,|\, \mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}\, X \,|\, \mathcal{G}]\, \mathbb{E}[Y \,|\, \mathcal{G}] = (1-\delta)\, \mathrm{ES}_\delta[X \,|\, \mathcal{G}]\, \mathbb{E}[Y \,|\, \mathcal{G}] \quad \textit{a.s.}$$

Remark 5.8. Note that $f_{\mathcal{G},\delta,X} \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$.

Proof of Lemma 5.7. (a) Note that $\mathbb{E}[f_{\mathcal{G},\delta,X} XY \,|\, \mathcal{G}]$ is a well-defined random variable in $\mathbb{R} \cup \{\infty\}$ since $(f_{\mathcal{G},\delta,X} XY)^- = Y\,(f_{\mathcal{G},\delta,X} X)^- \leq Y \min\{0, q_{\mathcal{G},\delta}(X)\}$ a.s., by Remark 5.6, as well as by positivity and σ-integrable of Y.

First, consider the set $M := \{0 < \delta < 1\}$ and let $f \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$ such that $(f\,XY)^-$ is σ-integrable with respect to $\mathcal{G}$. On $\{\mathbb{E}[f_{\mathcal{G},\delta,X} XY \,|\, \mathcal{G}] = \infty\}$ the result of the lemma follows trivially. Therefore, we may assume $\mathbb{E}[f_{\mathcal{G},\delta,X} XY \,|\, \mathcal{G}] < \infty$ a.s. Then, for every $G \in \mathcal{G}$,

$$1_M\, \mathbb{E}[(f - f_{\mathcal{G},\delta,X}) Y\, q_{\mathcal{G},\delta}(X) 1_G \,|\, \mathcal{G}] = 1_{G \cap M}\, q_{\mathcal{G},\delta}(X)\, \mathbb{E}[(f - f_{\mathcal{G},\delta,X}) Y \,|\, \mathcal{G}] = 0 \quad \text{a.s.},$$

since $f_{\mathcal{G},\delta,X} \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$ by Remark 5.8. This observation implies, a.s.,

$$\begin{aligned}
&1_M \big(\mathbb{E}[f\,XY 1_G \,|\, \mathcal{G}] - \mathbb{E}[f_{\mathcal{G},\delta,X} XY 1_G \,|\, \mathcal{G}] \big) \\
&\qquad = 1_M\, \mathbb{E}[(f - f_{\mathcal{G},\delta,X})(X - q_{\mathcal{G},\delta}(X)) Y 1_G \,|\, \mathcal{G}] \\
&\qquad = 1_M\, \mathbb{E}[(f - f_{\mathcal{G},\delta,X})(X - q_{\mathcal{G},\delta}(X)) Y 1_{\{X > q_\delta(X)\}} 1_G \,|\, \mathcal{G}] \\
&\qquad\quad + 1_M\, \mathbb{E}[(f - f_{\mathcal{G},\delta,X})(X - q_{\mathcal{G},\delta}(X)) Y 1_{\{X < q_\delta(X)\}} 1_G \,|\, \mathcal{G}] \leq 0\,,
\end{aligned}$$

i.e., the supremum is identical with $\mathbb{E}[f_{\mathcal{G},\delta,X} XY \,|\, \mathcal{G}]$.

On $N := \{\delta = 0\}$ we have $f_{\mathcal{G},\delta,X} = 1$ a.s. Therefore, $1_N f = 1_N$ a.s., for all $f \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$ which gives the result. On $N' := \{\delta = 1\}$ we have $\beta_{\mathcal{G},\delta,X} = 0$ which implies $f_{\mathcal{G},\delta,X} 1_{N'} = 0$ a.s. and $f 1_{N'} = 0$ a.s., for all $f \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$, which again immediately implies the result.

(b) Assume that there exists an $f^* \in \mathcal{F}^Y_{\mathcal{G},\delta,X}$ with

$$\mathbb{E}[f^* XY \,|\, \mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}\, XY \,|\, \mathcal{G}] < \infty \quad \text{a.s.}$$

Then, by the above calculations, $\mathbb{P}(f^* < f_{\mathcal{G},\delta,X}, X > q_{\mathcal{G},\delta}(X), Y > 0 \,|\, \mathcal{G}) = 0$, as well as $\mathbb{P}(f^* > f_{\mathcal{G},\delta,X}, X < q_{\mathcal{G},\delta}(X) Y > 0) \,|\, \mathcal{G}) = 0$, both a.s. Therefore, $f^* = f_{\mathcal{G},\delta,X}$ on the set $\{X \neq q_{\mathcal{G},\delta}(X), Y > 0\}$ a.s.

(c) This result immediately follows from the definition of conditional correlation and the definition of conditional expected shortfall. □

Definition 5.9. Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and a $\mathcal{G}$-measurable $\delta\colon \Omega \to (0,1)$, let $\mathcal{F}_{\mathcal{G},\delta}$ denote the set of all conditional probability densities given $\mathcal{G}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ bounded by $\frac{1}{1-\delta}$, i.e.,

$$\mathcal{F}_{\mathcal{G},\delta,X} := \left\{ f\colon \Omega \to [0,\infty) \,\middle|\, f \text{ is } \mathcal{F}\text{-measurable}, \mathbb{E}[f \,|\, \mathcal{G}] = 1 \text{ a.s. and } f \leq \frac{1}{1-\delta} \text{ a.s.} \right\}.$$

Furthermore, for an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$, define

$$\mathcal{F}_{\mathcal{G},\delta} := \left\{ f \in \mathcal{F}_{\mathcal{G},\delta} \,\middle|\, \mathbb{E}[X^+ f \,|\, \mathcal{G}] < \infty \text{ a.s. or } \mathbb{E}[X^- f \,|\, \mathcal{G}] < \infty \text{ a.s.} \right\}.$$

[28] Two random variables $\xi, \zeta\colon \Omega \to \mathbb{R}$ are conditionally uncorrelated given $\mathcal{G}$, where $\mathcal{G}$ is a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, if $\mathbb{E}[\xi\zeta \,|\, \mathcal{G}] = \mathbb{E}[\xi \,|\, \mathcal{G}]\, \mathbb{E}[\zeta \,|\, \mathcal{G}]$ a.s. given that all conditional expectations exist. In particular, ξ and ζ are conditionally uncorrelated given $\mathcal{G}$ if they are conditionally independent given $\mathcal{G}$ subject to existence of $\mathbb{E}[\xi\zeta \,|\, \mathcal{G}]$, $\mathbb{E}[\xi \,|\, \mathcal{G}]$ and $\mathbb{E}[\zeta \,|\, \mathcal{G}]$.

Remark 5.10. The definitions above are motivated by the classical, unconditional case as in Schmock [111, Section 7.2]. Note that, for sub-σ-algebras $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, we have $\mathcal{F}_{\mathcal{G},\delta} \subset \mathcal{F}_{\mathcal{H},\delta}$. Furthermore, $\mathcal{F}_{\mathcal{G},\delta'} \subset \mathcal{F}_{\mathcal{G},\delta}$ for all $\mathcal{G}$-measurable $\delta, \delta' \colon \Omega \to (0,1)$ with $\delta' \leq \delta$ a.s.

Remark 5.11 (Conditional expected proportional shortfall). Analogously as in the classical case as given in Belzunce et al. [12], we can now define a scale invariant conditional risk measure[29] based on conditional expected shortfall and conditional value at risk called conditional expected proportional shortfall. Therefore, consider a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ as well as an $\mathcal{F}$-measurable $X \colon \Omega \to \mathbb{R}$. Define

$$D^+(X,\mathcal{G}) = \{\delta \,|\, \delta \text{ is } \mathcal{G}\text{-measurable with } 0 \leq \delta \leq 1 \text{ a.s. and } q_{\mathcal{G},\delta}(X) > 0 \text{ a.s.}\}\,.$$

Then, *conditional expected proportional shortfall* of X at level $\delta \in D^+(X,\mathcal{G})$ is given by

$$\mathrm{CEPS}_\delta[X\,|\,\mathcal{G}] := (1-\delta)\left(\frac{\mathrm{ES}_\delta[X\,|\,\mathcal{G}]}{q_{\mathcal{G},\delta}(X)} - 1\right).$$

This risk measure is scale invariant as, for $\mathcal{G}$-measurable $Z \colon \Omega \to (0,\infty)$ and for $\delta \in D^+(X,\mathcal{G})$, Lemma 3.18(b) and Lemma 5.23(b) imply

$$\mathrm{CEPS}_\delta[X\,Z\,|\,\mathcal{G}] = \mathrm{CEPS}_\delta[X\,|\,\mathcal{G}] \quad \text{a.s.}$$

Furthermore, recall Definition 10.11 and note that conditional expected proportional shortfall is comonotonically subadditive for non-negative random variables with continuous distributions because, for $\mathcal{F}$-measurable $X, Y \colon \Omega \to \mathbb{R}$ and $\delta \in D^+(X,\mathcal{G}) \cap D^+(Y,\mathcal{G})$, we get

$$\begin{aligned}\mathrm{CEPS}_\delta[X+Y\,|\,\mathcal{G}] &= (1-\delta)\left(\frac{\mathrm{ES}_\delta[X\,|\,\mathcal{G}] - q_{\mathcal{G},\delta}(X)}{q_{\mathcal{G},\delta}(X)+q_{\mathcal{G},\delta}(Y)} + \frac{\mathrm{ES}_\delta[Y\,|\,\mathcal{G}] - q_{\mathcal{G},\delta}(Y)}{q_{\mathcal{G},\delta}(X)+q_{\mathcal{G},\delta}(Y)}\right)\\ &\leq \mathrm{CEPS}_\delta[X\,|\,\mathcal{G}] + \mathrm{CEPS}_\delta[Y\,|\,\mathcal{G}] \quad \text{a.s.},\end{aligned}$$

by Lemma 3.18(d) and Lemma 5.23(f). Further properties are left to the reader. Note that for trivial $\mathcal{G}$, see Footnote 17, it is possible to derive asymptotic results as $\delta \nearrow 1$ of the term $\mathrm{ES}_\delta[X\,|\,\mathcal{G}]\,/q_{\mathcal{G},\delta}(X)$[30]. Depending on the tail behaviour of the distribution of X, different limits arise, cf. McNeil, Frey and Embrechts [85] and Embrechts et al. [43].

5.2 Properties of conditional expected shortfall

In the lecture notes of Schmock [111, Lemma 7.20], as well as in various other references, different properties of expected shortfall are noted. As we could not find a comprehensive list of properties for conditional expected shortfall in the representation we used, we will give them in the following lemmas.

We start with a result which shows that conditional expected shortfall is a distortion risk measure with concave distortion process

$$g(t,\omega) := \min\left\{\frac{t}{1-\delta(\omega)}\,,1\right\}, \quad t \in [0,1] \text{ and } \omega \in \Omega\,,$$

as anticipated in Example 4.7(b).

[29] We refer to it as a conditional risk measure even though it is not a conditional risk measure in the narrower sense, according to Definition 1.1, as conditional translation invariance and conditional monotonicity are violated.

[30] This term is also known as shortfall to quantile ratio.

Lemma 5.12 (Conditional quantile representation)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $\delta\colon \Omega \to (0,1)$ be $\mathcal{G}$-measurable. For an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$ let $(q_{\mathcal{G},t}(X))_{t\in[0,1]}$ be the version of conditional lower quantiles with left-continuous and increasing paths as given in Lemma 3.15. Then, conditional expected shortfall of Definition 5.3 satisfies*

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \frac{1}{1-\delta}\int_{[\delta,1)} q_{\mathcal{G},t}(X)\,dt = \frac{1}{1-\delta}\int_{(0,1-\delta]} q_{\mathcal{G},1-t}(X)\,dt \quad a.s. \tag{5.13}$$

Corollary 5.14 (Nice version)**.** *Set $I := [0,\infty)$. If $(\delta_t)_{t\in I}$ is a $(\mathcal{G})_{t\in I}$-adapted $(0,1)$-valued process with continuous paths, Lemma 5.12 and Definition 5.3 imply that there exists a version of $(\mathrm{ES}_{\delta_t}[X\,|\,\mathcal{G}])_{t\in I}$ which has continuous paths. This observation remains valid if we replace continuous by right- or left-continuous in the assumption and conclusion.*

Proof of Lemma 5.12. Note that integrals in (5.13) are well-defined and that the second equality is obvious. By enlarging the probability space, if necessary, we may assume the existence of a random variable U on $(\Omega,\mathcal{F},\mathbb{P})$ which is independent of $\mathcal{G}$ and uniformly distributed on $[0,1]$, meaning that $\mathbb{P}(U \le t) = t$, for all $t \in [0,1]$. Note that the random variable $\Omega \ni \omega \mapsto q_{\mathcal{G},U(\omega)}(X)(\omega)$ is $\mathcal{F}$-measurable since it can be written as a composition of the measurable mappings $[0,1]\times\Omega \ni (t,\omega) \mapsto q_{\mathcal{G},t}(X)(\omega)$ and $\Omega \ni \omega \mapsto (U(\omega),\omega)$. For every $x \in \mathbb{R}$ and $t \in (0,1)$, we have

$$1_{\{q_{\mathcal{G},t}(X)\le x\}} = 1_{\{\mathbb{P}(X\le x\,|\,\mathcal{G})\ge t\}} \quad \text{a.s.}, \tag{5.15}$$

since, by Definition 3.7 of conditional lower quantiles,

$$\mathbb{P}\big(\{q_{\mathcal{G},t}(X) \le x\} \setminus \{\mathbb{P}(X \le x\,|\,\mathcal{G}) \ge t\}\big) = 0\,,$$

as well as

$$\mathbb{P}\big(\{\mathbb{P}(X \le x\,|\,\mathcal{G}) \ge t\} \setminus \{q_{\mathcal{G},t}(X) \le x\}\big) = 0\,.$$

Note that with a similar argumentation as in Lemma 3.15 and using the dominated convergence theorem for conditional expectation, cf. [65, Theorem 1.20], there exists a version of $\mathbb{R} \ni x \mapsto \mathbb{P}(X \le x\,|\,\mathcal{G})$ which has increasing and right-continuous paths. Taking this version of conditional probabilities implies that the functions in (5.15) are decreasing and left-continuous in t, as well as increasing and right-continuous in x. Thus, we do not face problems with null sets in the subsequent argumentation, see Remark 5.20 for a motivating example.

Based on the previous equations we get, for all $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R}$,

$$1_{\{q_{\mathcal{G},U}(X)\le Z\}} = 1_{\{U\le \mathbb{P}(X\le Z\,|\,\mathcal{G})\}} \quad \text{a.s.}, \tag{5.16}$$

implying

$$\mathbb{P}(q_{\mathcal{G},U}(X) \le Z\,|\,\mathcal{G}) = \mathbb{P}(U \le \mathbb{P}(X \le Z\,|\,\mathcal{G})\,|\,\mathcal{G}) = \mathbb{P}(X \le Z\,|\,\mathcal{G}) \quad \text{a.s.} \tag{5.17}$$

Note that the the second equality above follows by a basic property of conditional expectation, see Lemma 10.3(b). Define $\delta' := \mathbb{P}(X \le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})$. Then, $\delta' \ge \delta$ and $q_{\mathcal{G},t}(X) = q_{\mathcal{G},\delta}(X)$ for every $t \in [\delta,\delta']$, both a.s. Therefore, again by Lemma 10.3(b) and by conditional bounded convergence as, for example, given in [65, Theorem 1.20] on $q_{\mathcal{G},U}(X) = \lim_{n\to\infty}\sum_{j=1}^{k_n} q_{\mathcal{G},\alpha_{j-1,n}}(X)1_{U\in[\alpha_{j-1,n},\alpha_{j,n})}$ for proper partitions of the unit interval with mesh tending to zero $0 = \alpha_{0,n} < \alpha_{1,n} < \cdots < \alpha_{k_n,n} = 1$ with $n \in \mathbb{N}$, we have

$$\begin{aligned}\int_{[\delta,1)} q_{\mathcal{G},t}(X)\,dt &= \int_{(\delta',1)} q_{\mathcal{G},t}(X)\,dt + \int_{[\delta,\delta']} q_{\mathcal{G},t}(X)\,dt \\ &= \mathbb{E}\big[q_{\mathcal{G},U}(X)1_{\{U>\delta'\}}\,\big|\,\mathcal{G}\big] + q_{\mathcal{G},\delta}(X)(\delta'-\delta) \quad \text{a.s.}\end{aligned} \tag{5.18}$$

Moreover, by (5.16) we find that $\{U > \delta'\} = \{q_{\mathcal{G},U}(X) > q_{\mathcal{G},\delta}(X)\}$ a.s. Thus, using Fubini's theorem for generalised conditional expectation, see Lemma 10.4,

$$\begin{aligned}\mathbb{E}\big[q_{\mathcal{G},U}(X)\,1_{\{U>\delta'\}}\,\big|\,\mathcal{G}\big] &= \mathbb{E}\bigg[\int_{(0,\infty)} 1_{\{q_{\mathcal{G},U}(X)>t\}}\,dt\,1_{\{q_{\mathcal{G},U}(X)>q_{\mathcal{G},\delta}(X)\}}\,\bigg|\,\mathcal{G}\bigg] \\ &= \int_{(0,\infty)} \mathbb{E}\big[1_{\{q_{\mathcal{G},U}(X)>\max\{t,q_{\mathcal{G},\delta}(X)\}}\,\big|\,\mathcal{G}\big]\,dt \quad \text{a.s.}\end{aligned}$$

Note that Lemma 10.4 can be applied since, with a similar argumentation as in Lemma 3.15 and using conditional dominated convergence, cf. [65, Theorem 1.20], there exists a version of

$$\mathbb{R} \ni t \mapsto \mathbb{E}\big[1_{\{q_{\mathcal{G},U}(X)>\max\{t,q_{\mathcal{G},\delta}(X)\}}\,\big|\,\mathcal{G}\big]$$

which has increasing paths. Applying (5.17) gives

$$\mathbb{E}\big[q_{\mathcal{G},U}(X)\,1_{\{U>\delta'\}}\,\big|\,\mathcal{G}\big] = \mathbb{E}\big[X\,1_{\{X>q_{\mathcal{G},\delta}(X)\}}\,\big|\,\mathcal{G}\big] \quad \text{a.s.}$$

Substituting this result into (5.18) implies

$$\begin{aligned}\int_{[\delta,1)} q_{\mathcal{G},t}(X)\,dt &= \int_{(\delta',1)} q_{\mathcal{G},t}(X)\,dt + \int_{[\delta,\delta']} q_{\mathcal{G},t}(X)\,dt \\ &= \mathbb{E}\big[X1_{\{X>q_{\mathcal{G},\delta}(X)\}}\,\big|\,\mathcal{G}\big] + \mathbb{E}\big[q_{\mathcal{G},\delta}(X)\big(1_{\{X\le q_{\mathcal{G},\delta}(X)\}} - \delta\big)\,\big|\,\mathcal{G}\big] \quad \text{a.s.}\end{aligned} \tag{5.19}$$

Lemma 5.2(a) then gives the result. Note that $(q_{\mathcal{G},U}(X)1_{\{U>\delta'\}})^-$ is σ-integrable with respect to $\mathcal{G}$. To see this, recall that $(q_{\mathcal{G},U}(X)1_{\{U>\delta'\}})^- \le X'$ a.s. where $X' := -\min\{0, q_{\mathcal{G},\delta'}(X)\}$ is $\mathcal{G}$-measurable. Thus, $\Omega_n := \{|X'| \le n\} \in \mathcal{G}$ for all $n \in \mathbb{N}$, and $\Omega_n \nearrow \Omega$ as $n \to \infty$, since $\delta' > 0$ a.s. Then, for every $n \in \mathbb{N}$, we have $\mathbb{E}[1_{\Omega_n}(q_{\mathcal{G},U}(X)1_{\{U>\delta'\}})^-] \le \mathbb{E}[1_{\Omega_n}|X'|] \le n$, i.e., $(q_{\mathcal{G},U}(X)1_{\{U>\delta'\}})^-$ is σ-integrable with respect to $\mathcal{G}$. Hence, $\mathbb{E}[q_{\mathcal{G},U}(X)1_{\{U>\delta'\}}\,|\,\mathcal{G}]$ is well-defined. Continuous paths follow by definition. □

Remark 5.20 (Null sets can be cruel). Consider $(\Omega, \mathcal{F}, \mathbb{P}) := ([0,1], \mathcal{B}([0,1]), \lambda)$ where λ denotes the Lebesgue–Borel measure on Borel-σ-algebra $\mathcal{B}([0,1])$. Define two measurable functions $f, g\colon [0,1] \times \Omega \to \mathbb{R}$ by $f_t(\omega) = 0$ and $g_t(\omega) = 1_{\{\omega\}}(t)$, for $\omega \in \Omega$ and $t \in [0,1]$. Then, $f_t = g_t$ a.s. for every $t \in [0,1]$. Defining the identity map $U(\omega) := \omega$ for all $\omega \in \Omega$ gives $f_U = 0$ and $g_U = 1$, both a.s. Thus, $f_U < g_U$ a.s. This example illustrates that it is crucial to take special versions of the functions used in (5.15) in order to avoid such phenomena involving null sets.

Corollary 5.21. *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $X\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable such that X^- is σ-integrable with respect to $\mathcal{G}$. Moreover, consider the version of $(q_{\mathcal{G},\delta}(X))_{\delta\in(0,1)}$ with left-continuous, increasing paths, as given in Lemma 3.15, as well as let $(\mathbb{P}(X \le t\,|\,\mathcal{G}))_{t\in\mathbb{R}}$ be the version of conditional probabilities with right-continuous, increasing paths. Then,*

$$\mathbb{E}[X\,|\,\mathcal{G}] = \int_{(0,1)} q_{\mathcal{G},t}(X)\,dt = \int_0^\infty \mathbb{P}(X > t\,|\,\mathcal{G})\,dt - \int_{-\infty}^0 \mathbb{P}(X \le t\,|\,\mathcal{G})\,dt \quad \textit{a.s.}$$

In particular, in this case, $\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \mathbb{E}[X\,|\,\mathcal{G}] = \int_{(0,1)} q_{\mathcal{G},t}(X)\,dt$ *a.s. on* $\{\delta = 0\}$.

Proof. The first equality immediately follows using the same argumentation as in the proof of Lemma 5.12 and Equation (5.19) with $\delta = 0$. The last statement then follows by (5.26).

For the second equality note that, for every $n \in \mathbb{N}$, we have

$$\mathbb{E}[\min\{X,n\}\,|\,\mathcal{G}] = \mathbb{E}\left[\int_0^\infty 1_{\{\min\{X,n\}>t\}}\,dt \,\middle|\, \mathcal{G}\right] - \mathbb{E}\left[\int_0^\infty 1_{\{X\le t\}}\,dt \,\middle|\, \mathcal{G}\right] \quad \text{a.s.}$$

Then, using the conditional Fubini theorem of Lemma 10.4, we can conclude

$$\mathbb{E}[\min\{X,n\}\,|\,\mathcal{G}] = \int_0^\infty \mathbb{P}(\min\{X,n\} > t\,|\,\mathcal{G})\,dt - \int_{-\infty}^0 \mathbb{P}(X \le t\,|\,\mathcal{G})\,dt \quad \text{a.s.}$$

Letting $n \to \infty$ gives the result by applying monotone convergence and conditional monotone convergence, see [65, Theorem 1.20]. □

Remark 5.22. If $\mathcal{G}$ is trivial, see Footnote 17, then the previous result corresponds to the well known fact that, for a real-valued random variable X with $\mathbb{E}[X^-] < \infty$, we may write

$$\mathbb{E}[X] = \int_{(0,1)} F^{\leftarrow}(t)\,dt$$

where $F^{\leftarrow}$ denotes the lower quantile function of X, see Remarks 3.11(a).

Many properties of conditional expected shortfall immediately follow by corresponding results of conditional distortion risk measures. Alternative proofs for some of these results are provided in Appendix 10.2.

Lemma 5.23 (Properties of conditional expected shortfall)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $X, Y\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and let $\delta\colon \Omega \to [0,1]$ be $\mathcal{G}$-measurable. Then, conditional expected shortfall of Definition 5.3 has the following conditional properties, considering conventions of Footnote 22:*

(a) Conditional normalisation:

$$\mathrm{ES}_\delta[0\,|\,\mathcal{G}] = 0 \quad \textit{a.s.}$$

(b) Conditional positive homogeneity: If $Z\colon \Omega \to [0,\infty)$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_\delta[X\,Z\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}]\;Z \quad \textit{a.s.}$$

(c) Conditional translation invariance: If $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_\delta[X + Z\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}] + Z \quad \textit{a.s.}$$

(d) Subadditivity:

$$\mathrm{ES}_\delta[X + Y\,|\,\mathcal{G}] \le \mathrm{ES}_\delta[X\,|\,\mathcal{G}] + \mathrm{ES}_\delta[Y\,|\,\mathcal{G}] \quad \textit{a.s.}$$

(e) Conditional convexity: If $Z\colon \Omega \to [0,1]$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_\delta[X\,Z + Y\,(1-Z)\,|\,\mathcal{G}] \le \mathrm{ES}_\delta[X\,|\,\mathcal{G}]\;Z + \mathrm{ES}_\delta[Y\,|\,\mathcal{G}]\;(1-Z) \quad \textit{a.s.}$$

(f) Comonotonic additivity: If X and Y are continuously distributed and if they are comonotonic, see Definition 10.11, then

$$\mathrm{ES}_\delta[X + Y\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}] + \mathrm{ES}_\delta[Y\,|\,\mathcal{G}] \quad \textit{a.s.}$$

(g) *Conditionally comonotonic additivity: If X and Y are conditionally comonotonic with respect to $\mathcal{G}$, see Definition 10.11, then*

$$\mathrm{ES}_\delta[X+Y\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}] + \mathrm{ES}_\delta[Y\,|\,\mathcal{G}] \quad a.s.$$

(h) *Independence: If X is independent of $\mathcal{G}$, then*

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X] \quad a.s.,$$

where $\mathrm{ES}_t[X]$ denotes the unconditional expected shortfall at deterministic level $t \in [0,1]$, cf. Schmock [111, Section 7.2], i.e., $\mathrm{ES}_t[X] = \mathrm{ES}_t[X\,|\,\mathcal{H}]$ a.s. where $\mathcal{H}$ is a trivial σ-algebra, see Footnote 17.

(i) *Conditional monotonicity: If either $X \leq_{\mathrm{icx}(\mathcal{G})} Y$ or $X \leq_{\mathrm{cx}(\mathcal{G})} Y$, and if in addition $X1_{\{\delta=1\}} \leq_{\mathrm{st}(\mathcal{G})} Y1_{\{\delta=1\}}$, see Definition 10.14, then*

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] \leq \mathrm{ES}_\delta[Y\,|\,\mathcal{G}] \quad a.s.$$

In particular, the assumptions are satisfied if $X \leq Y$ a.s.

(j) *Monotonicity in the adjusted indicator function: Let $X\colon \Omega \to [0,\infty)$ be positive and consider an $\mathcal{F}$-measurable function $f\colon \Omega \to [0,\infty)$. Then, if $f \leq_{\mathrm{icx}(\mathcal{G})} f_{\mathcal{G},\delta,X}$, see Definitions 5.3 and 10.14,*

$$\mathbb{E}[f\,X\,|\,\mathcal{G}] \leq \mathrm{ES}_\delta[X\,|\,\mathcal{G}] \quad a.s.$$

(k) *Determined by conditional law: If $\mathbb{E}[f(X)\,|\,\mathcal{G}] = \mathbb{E}[f(Y)\,|\,\mathcal{G}]$ a.s. for every bounded and continuous function $f\colon \mathbb{R} \to \mathbb{R}$, then*

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \mathrm{ES}_\delta[Y\,|\,\mathcal{G}] \quad a.s.$$

(l) *Regularity: If $A \in \mathcal{G}$, then $X1_A = Y1_A$ a.s. implies*

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}]\,1_A = \mathrm{ES}_\delta[Y\,|\,\mathcal{G}]\,1_A \quad a.s.$$

(m) *If $f\colon \mathbb{R} \to \mathbb{R}$ is strictly increasing, as well as convex and if $|\,\mathrm{ES}_\delta[X\,|\,\mathcal{G}]\,| < \infty$ a.s., then*

$$f(\mathrm{ES}_\delta[X\,|\,\mathcal{G}]) \leq \mathrm{ES}_\delta[f(X)\,|\,\mathcal{G}] \quad a.s.$$

If f is concave, then the reverse inequality holds.

(n) *Supermartingale: Let $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ be two σ-algebras and let δ be $\mathcal{H}$-measurable. If $(X\,f_{\mathcal{G},\delta,X})^-$ or if $(\mathrm{ES}_\delta[X\,|\,\mathcal{G}])^-$ is σ-integrable with respect to $\mathcal{H}$, then*

$$\mathbb{E}[\mathrm{ES}_\delta[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \leq \mathrm{ES}_\delta[X\,|\,\mathcal{H}] \quad a.s.$$

(o) *Uncertainty decrease of projections: Let $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ be two σ-algebras and let δ be $\mathcal{H}$-measurable. If X is σ-integrable with respect to $\mathcal{H}$, then*

$$\mathrm{ES}_\delta[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \leq \mathrm{ES}_\delta[X\,|\,\mathcal{H}] \quad a.s.$$

(p) *Conditional Fatou I: Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of real-valued random variables bounded from below by some* $\mathcal{G}$*-measurable* $C\colon \Omega \to \mathbb{R}$ *converging to* X *in conditional probability, i.e.,* $\lim_{n\to\infty} \mathbb{P}(|X - X_n| \geq \varepsilon \,|\, \mathcal{G}) = 0$ *a.s. for every* $\varepsilon > 0$*, see Footnote 23. Then,*

$$\liminf_{n\to\infty} \mathrm{ES}_\delta[X_n \,|\, \mathcal{G}] \geq \mathrm{ES}_\delta[X \,|\, \mathcal{G}] \quad \text{a.s.}$$

(q) *Conditional Fatou II: Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random variables bounded from below by some* $\mathcal{G}$*-measurable* $C\colon \Omega \to \mathbb{R}$*. Then,* $X := \liminf_{n\to\infty} X_n$ *satisfies*

$$\liminf_{n\to\infty} \mathrm{ES}_\delta[X_n \,|\, \mathcal{G}] \geq \mathrm{ES}_\delta[X \,|\, \mathcal{G}] \quad \text{a.s.} \tag{5.24}$$

(r) *Continuity from below: Let* $(X_n)_{n\in\mathbb{N}}$ *be an increasing sequence of real-valued random variables bounded from below by some* $\mathcal{G}$*-measurable* $C\colon \Omega \to \mathbb{R}$ *converging to* X *from below, i.e.,* $X_n \nearrow X$ *a.s. as* $n \to \infty$*. Then,*

$$\lim_{n\to\infty} \mathrm{ES}_\delta[X_n \,|\, \mathcal{G}] = \mathrm{ES}_\delta[X \,|\, \mathcal{G}] \quad \text{a.s.}$$

(s) *Bounds: Define* $\mathbb{E}[X^+ \,|\, \mathcal{G}]\,/0 = \infty$*, then*

$$q_{\mathcal{G},\delta}(X) \leq \mathrm{ES}_\delta[X \,|\, \mathcal{G}] \leq \min\left\{\frac{\mathbb{E}[X^+ \,|\, \mathcal{G}]}{1-\delta}, X^{\mathcal{G}}\right\} \quad \text{a.s.}$$

(t) *Scenario representation: Using the notation of Lemma 5.7 and Definition 5.9, the following holds:*

(1) $\mathrm{ES}_\delta[X \,|\, \mathcal{G}] = \frac{1}{1-\delta} \operatorname{ess\,sup}_{f\in\mathcal{F}^1_{\mathcal{G},\delta,X}} \mathbb{E}[f\,X \,|\, \mathcal{G}]$ *on* $\{0 < \delta < 1\}$ *a.s.*

(2) $\mathrm{ES}_\delta[X \,|\, \mathcal{G}] = \operatorname{ess\,sup}_{f\in\mathcal{F}_{\mathcal{G},\delta,X}} \mathbb{E}[f\,X \,|\, \mathcal{G}]$ *on* $\{0 < \delta < 1\}$ *a.s.*

(3) $\mathrm{ES}_\delta[X \,|\, \mathcal{G}] = \operatorname{ess\,sup}_{f\in\mathcal{F}_{\mathcal{G},\delta}} \mathbb{E}[f\,X \,|\, \mathcal{G}]$ *a.s. if we either have* $\mathbb{E}[X^+ \,|\, \mathcal{G}] < \infty$ *or* $\mathbb{E}[X^- \,|\, \mathcal{G}] < \infty$*, both a.s. on* $\{0 < \delta < 1\}$*.*

(4) $\mathrm{ES}_\delta[X \,|\, \mathcal{G}] = \operatorname{ess\,inf}_{Z\in L^0(\Omega,\mathcal{G},\mathbb{P})}(Z + \frac{1}{1-\delta}\,\mathbb{E}[(X-Z)^+ \,|\, \mathcal{G}])$ *on* $\{0 < \delta < 1\}$ *a.s. where* $L^0(\Omega,\mathcal{G},\mathbb{P})$ *is given in Footnote 3.* $Z \in L^0(\Omega,\mathcal{G},\mathbb{P})$ *attains the infimum if and only if* $Z = q_{\mathcal{G},\delta}(X)$ *on* $\{\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X) \,|\, \mathcal{G}) > \delta\}$ *and* $q_{\mathcal{G},\delta}(X) \leq Z \leq q^{\mathcal{G},\delta}(X)$ *a.s. on* $\{\mathbb{P}(X \leq q_{\mathcal{G},\delta}(X) \,|\, \mathcal{G}) = \delta\}$*.*

Proof. On $\{0 < \delta < 1\}$, Items (a) to (r), except (j) and (n), follow by the corresponding result for conditional distortion risk measures with distortion process

$$g(t,\omega) := \min\left\{\frac{t}{1-\delta(\omega)}, 1\right\}, \quad t \in [0,1] \text{ and } \omega \in \Omega\,.$$

Note that $X, Y \in \mathcal{L}_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$ if $\delta > 0$ a.s. since $q_{\mathcal{G},\delta}(X) > 0$ a.s., see Definition 4.1. By invoking Corollary 5.26, the results then extend to the essential infimum, i.e., on $\{\delta = 0\}$, where it is necessary to define $\infty - \infty := \infty$. On $\{\delta = 1\}$, the results follow by the corresponding results in Lemmas 3.4 and 3.18. Moreover, as we use Remarks 5.25(d) for the proof of Lemma 4.8(p) in the previous section, we have to give an alternative proof for conditional monotonicity of conditional expected shortfall, see (i).

(i) Define $M := \{0 < \delta < 1\}$. Since $X \leq_{\mathrm{icx}(\mathcal{G})} Y$, Lemma 10.15(c) applied to convex functions $h(x,z) := (x-z)^+$ with $x, z \in \mathbb{R}$ implies, for all $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R}$,

$$1_M\Big(Z + \frac{1}{1-\delta}\,\mathbb{E}[(X-Z)^+ \,|\, \mathcal{G}]\Big) \leq 1_M\Big(Z + \frac{1}{1-\delta}\,\mathbb{E}[(X-Z)^+ \,|\, \mathcal{G}]\Big) \quad \text{a.s.}$$

Thus, by Representation 5.23(t4),

$$1_M \operatorname{ES}_\delta[X\,|\,\mathcal{G}] \le 1_M \operatorname{ES}_\delta[Y\,|\,\mathcal{G}] \quad \text{a.s.}$$

The result remains true if the weaker order $X \le_{\mathrm{cx}(\mathcal{G})} Y$ is assumed. The result extends to the essential infimum, i.e., on $\{\delta = 0\}$. On $\{\delta = 1\}$ the result follows by Lemma 3.18(g).

(j) By definition of $f_{\mathcal{G},\delta,X}$, there exists a $(\mathcal{G})_{t\in[0,1]}$-adapted process $(h(t,\cdot))_{t\in\mathbb{R}}$ with increasing paths such that $f_{\mathcal{G},\delta,X} = h(X,\cdot)$ a.s. Note that by a fundamental property of lower inverse functions, see Remarks 3.11(a) and [39], we have $1_{\{h(X,\cdot)\le x\}} = 1_{\{X\le h^{\leftarrow}(x,\cdot)\}}$ for all $x \in \mathbb{R}$ where $\mathbb{R} \ni x \mapsto h^{\leftarrow}(x,\omega)$ denotes the pathwise lower inverse of $h(\cdot,\omega)$ for all $\omega \in \Omega$. Thus, for all $x, y \in \mathbb{R}$, we get

$$\begin{aligned}\mathbb{P}(f_{\mathcal{G},\delta,X} \le x, X \le y\,|\,\mathcal{G}) &= \mathbb{P}(X \le \min\{h^{\leftarrow}(x,\cdot), y\}\,|\,\mathcal{G}) \\ &= \min\big\{\mathbb{P}(f_{\mathcal{G},\delta,X} \le x\,|\,\mathcal{G}), \mathbb{P}(X \le y,\,|\,\mathcal{G})\big\} \quad \text{a.s.},\end{aligned}$$

which states that $f_{\mathcal{G},\delta,X}$ and X are conditionally comonotonic given $\mathcal{G}$. Thus, an application of Lemma 10.15(d) gives the result.

(n) First, define $M := \{0 < \delta < 1\}$. By Representation (t3) of Lemma 5.23 and the tower property of conditional expectation, see [65, Theorem 1.22],

$$1_M \operatorname{ES}_\delta[X\,|\,\mathcal{H}] = 1_M \operatorname*{ess\,sup}_{f\in\mathcal{F}_{\mathcal{H},\delta}} \mathbb{E}[f\,X\,|\,\mathcal{H}] = 1_M \operatorname*{ess\,sup}_{f\in\mathcal{F}_{\mathcal{H},\delta}} \mathbb{E}[\mathbb{E}[f\,X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \quad \text{a.s.}$$

By Remark 5.10 and by defining $f_{\mathcal{G},\delta,X}/(1-\delta) =: \tilde{f}_{\mathcal{G},\delta,X} \in \mathcal{F}_{\mathcal{G},\delta}$ with $f_{\mathcal{G},\delta,X}$ defined as in Definition 5.1, we get

$$\begin{aligned}1_M \operatorname*{ess\,sup}_{f\in\mathcal{F}_{\mathcal{H},\delta}} \mathbb{E}[\mathbb{E}[f\,X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \ge 1_M \operatorname*{ess\,sup}_{f\in\mathcal{F}_{\mathcal{G},\delta}} \mathbb{E}[\mathbb{E}[f\,X\,|\,\mathcal{G}]\,|\,\mathcal{H}] &\ge 1_M\,\mathbb{E}\big[\mathbb{E}\big[\tilde{f}_{\mathcal{G},\delta,X}X\,\big|\,\mathcal{G}\big]\,\big|\,\mathcal{H}\big] \\ &= 1_M\,\mathbb{E}[\operatorname{ES}_\delta[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \quad \text{a.s.},\end{aligned}$$

which proves the result on M. The result of course extends to the essential infimum, i.e. on $\{\delta = 0\}$. By Remarks 3.3(d), note that $X^{\mathcal{H}} \ge X^{\mathcal{G}}$ a.s. on $\{\delta = 1\}$ and thus

$$\mathbb{E}\big[X^{\mathcal{G}}\,\big|\,\mathcal{H}\big] - X^{\mathcal{H}} = \mathbb{E}\big[X^{\mathcal{G}} - X^{\mathcal{H}}\,\big|\,\mathcal{H}\big] \ge 0 \quad \text{a.s.},$$

which gives the result.

(o) First, consider the result on the set $M := \{0 < \delta < 1\}$. Let $\mathcal{F}^1_{\mathcal{H},\delta,X}$ be defined as in Lemma 5.7 and define $\mathcal{F}^1_{\mathcal{H},\delta,X}(\mathcal{G})$ as the set of all $f \in \mathcal{F}^1_{\mathcal{H},\delta,X}$ which are $\mathcal{G}$-measurable. Then, for every $\mathcal{F}$-measurable $f\colon \Omega \to \mathbb{R}$, we get

$$1_M\,\mathbb{E}[f\,\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] = 1_M\,\mathbb{E}[\mathbb{E}[f\,|\,\mathcal{G}]\,\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \quad \text{a.s.}$$

Thus, using Lemma 5.7(a),

$$\begin{aligned}1_M \operatorname{ES}_\delta[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] &= \frac{1_M}{1-\delta} \operatorname*{ess\,sup}_{f\in\mathcal{F}^1_{\mathcal{H},\delta,X}} \mathbb{E}[f\,\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \le \frac{1_M}{1-\delta} \operatorname*{ess\,sup}_{f\in\mathcal{F}^1_{\mathcal{H},\delta,X}(\mathcal{G})} \mathbb{E}[f\,X\,|\,\mathcal{H}] \\ &\le \frac{1_M}{1-\delta} \sup_{f\in\mathcal{F}^1_{\mathcal{H},\delta,X}} \mathbb{E}[f\,X\,|\,\mathcal{H}] = 1_M \operatorname{ES}_\delta[X\,|\,\mathcal{H}] \quad \text{a.s.}\end{aligned}$$

On $\{\delta = 1\}$ the result follows by Remarks 3.3(j). Finally, on $\{\delta = 0\}$, we have

$$1_{\{\delta=0\}} \operatorname{ES}_\delta[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \le 1_{\{\delta=0\}} \operatorname{ES}_{\delta'}[X\,|\,\mathcal{H}] \quad \text{a.s., for all } \delta' \in (0,1)\,,$$

implying that

$$\begin{aligned}1_{\{\delta=0\}}\,\mathrm{ES}_\delta[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] &= 1_{\{\delta=0\}}\operatorname*{ess\,inf}_{\delta'\in(0,1)}\mathrm{ES}_{\delta'}[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}]\\ &\le 1_{\{\delta=0\}}\operatorname*{ess\,inf}_{\delta'\in(0,1)}\mathrm{ES}_{\delta'}[X\,|\,\mathcal{H}] = 1_{\{\delta=0\}}\,\mathrm{ES}_\delta[X\,|\,\mathcal{H}] \quad \text{a.s.}\end{aligned}$$

(s) First, on $\{0<\delta<1\}$, note that $X\le X^+$ implies $\mathrm{ES}_\delta[X\,|\,\mathcal{G}]\le \mathrm{ES}_\delta[X^+\,|\,\mathcal{G}]$ a.s. by (i). Furthermore, $f_{\mathcal{G},\delta,X}\le 1$ a.s. Using these observations for the upper bound, both bounds follow directly from Definition 5.3. On $\{\delta=0\}$ the lower bound follows by taking the essential supremum on both sides whereas the upper bound follows as $X\le X^+$ a.s. On $\{\delta=1\}$ we have $\mathbb{E}[X^+\,|\,\mathcal{G}]\,/(1-\delta)=\infty$ and $q_{\mathcal{G},1}(X)=\mathrm{ES}_1[X\,|\,\mathcal{G}]=X^{\mathcal{G}}$, both a.s.

(t1) The first representation follows directly from Lemma 5.7.

(t2) and (t3) The proof of these results works similarly as the proof of Lemma 5.7. If $\tilde f_{\mathcal{G},\delta,X}:=f_{\mathcal{G},\delta,X}/(1-\delta)\in\mathcal{F}_{\mathcal{G},\delta,X}$, then the essential supremum is an upper bound for $\mathrm{ES}_\delta[X\,|\,\mathcal{G}]$ and (t2) holds on $\{\mathrm{ES}_\delta[X\,|\,\mathcal{G}]=\infty\}$. Thus, we may assume $\mathrm{ES}_\delta[X\,|\,\mathcal{G}]<\infty$ a.s. Then, necessarily, $\mathbb{E}[X^+\,|\,\mathcal{G}]<\infty$ a.s. and hence $\mathcal{F}_{\mathcal{G},\delta,X}=\mathcal{F}_{\mathcal{G},\delta}$ a.s. Consider an $f\in\mathcal{F}_{\mathcal{G},\delta}$ with $\mathbb{E}[f\,X\,|\,\mathcal{G}]>-\infty$ a.s. We have $\mathbb{E}[f-\tilde f_{\mathcal{G},\delta,X}\,|\,\mathcal{G}]=0$ a.s. and hence, for every $G\in\mathcal{G}$,

$$\begin{aligned}\mathbb{E}[f\,X\,1_G\,|\,\mathcal{G}]-\mathbb{E}\big[\tilde f_{\mathcal{G},\delta,X}\,X\,1_G\,\big|\,\mathcal{G}\big] &= \mathbb{E}\big[(f-\tilde f_{\mathcal{G},\delta,X})(X-q_{\mathcal{G},\delta}(X))\,1_G\,\big|\,\mathcal{G}\big]\\ &= \mathbb{E}\big[(f-\tilde f_{\mathcal{G},\delta,X})(X-q_{\mathcal{G},\delta}(X))\,1_{\{X>q_\delta(X)\}}1_G\,\big|\,\mathcal{G}\big]\\ &\quad+\mathbb{E}\big[(f-\tilde f_{\mathcal{G},\delta,X})(X-q_{\mathcal{G},\delta}(X))\,1_{\{X<q_\delta(X)\}}1_G\,\big|\,\mathcal{G}\big]\le 0 \quad \text{a.s.},\end{aligned}$$

implying that the supremum is identical with $\mathbb{E}[\tilde f_{\mathcal{G},\delta,X}X\,|\,\mathcal{G}]$.

(t4) Using the last term in (5.19), we can write

$$\begin{aligned}\mathrm{ES}_\delta[X\,|\,\mathcal{G}] &= \frac{1}{1-\delta}\Big(\mathbb{E}\big[X\,1_{\{X>q_{\mathcal{G},\delta}(X)\}}\,\big|\,\mathcal{G}\big]+\mathbb{E}\big[q_{\mathcal{G},\delta}(X)\big(1_{\{X\le q_{\mathcal{G},\delta}(X)\}}-\delta\big)\,\big|\,\mathcal{G}\big]\Big)\\ &= \frac{1}{1-\delta}\,\mathbb{E}\big[(X-q_{\mathcal{G},\delta}(X))^+\,\big|\,\mathcal{G}\big]+q_{\mathcal{G},\delta}(X) \quad \text{a.s.},\end{aligned}$$

which gives equality in the case $Z=q_{\mathcal{G},\delta}(X)$.

Thus, we have to show that the term $Z+\frac{1}{1-\delta}\,\mathbb{E}[(X-Z)^+\,|\,\mathcal{G}]$ takes an essential infimum for $Z=q_{\mathcal{G},\delta}(X)$. Let $Z\colon\Omega\to\mathbb{R}$ be $\mathcal{G}$-measurable and fixed from now on. Then, for $M:=\{q_{\mathcal{G},\delta}(X)<Z\}$,

$$1_M\,(X-q_{\mathcal{G},\delta}(X))^+\le 1_M\big((Z-q_{\mathcal{G},\delta}(X))\,1_{\{X>q_{\mathcal{G},\delta}(X)\}}+(X-Z)^+\big) \quad \text{a.s.},$$

with strict inequality on the event $\{q_{\mathcal{G},\delta}(X)<X<Z\}$. Adding $q_{\mathcal{G},\delta}(X)(1-\delta)$ to both sides and taking conditional expectations with respect to $\mathcal{G}$, we get

$$\begin{aligned}&1_M\big(q_{\mathcal{G},\delta}(X)(1-\delta)+\mathbb{E}\big[(X-q_{\mathcal{G},\delta}(X))^+\,\big|\,\mathcal{G}\big]\big)\\ &\quad\le 1_M\big(q_{\mathcal{G},\delta}(X)(1-\delta)+(Z-q_{\mathcal{G},\delta}(X))\,\mathbb{P}(X>q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})\big)+\mathbb{E}\big[(X-Z)^+\,\big|\,\mathcal{G}\big]\big)\\ &\quad\le 1_M\big(Z(1-\delta)+\mathbb{E}\big[(X-Z)^+\,\big|\,\mathcal{G}\big]\big) \quad \text{a.s.},\end{aligned}$$

since $\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})\ge\delta$ a.s., by Lemma 3.14. Equality in the equation above holds if and only if $\mathbb{P}(q_{\mathcal{G},\delta}(X)<X<Z)=0$ and $\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})=\delta$ a.s. which, by the definition of conditional lower and upper quantiles, is equivalent to $q_{\mathcal{G},\delta}(X)<Z\le q^{\mathcal{G},\delta}(X)$ a.s. on the set $\{\mathbb{P}(X\le q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G})=\delta\}$.

Conversely, consider the set $M^{\mathrm{c}}:=\{q_{\mathcal{G},\delta}(X)>Z\}$ and note that

$$1_{M^{\mathrm{c}}}\,(X-q_{\mathcal{G},\delta}(X))^+\le 1_M\big((Z-q_{\mathcal{G},\delta}(X))\,1_{\{X\ge q_{\mathcal{G},\delta}(X)\}}+(X-Z)^+\big) \quad \text{a.s.}$$

Just as before, we get

$$\begin{aligned}
&1_{M^c}\big(q_{\mathcal{G},\delta}(X)(1-\delta)+\mathbb{E}\big[(X-q_{\mathcal{G},\delta}(X))^+\,\big|\,\mathcal{G}\big]\big)\\
&\quad\le 1_{M^c}\big(q_{\mathcal{G},\delta}(X)(1-\delta)+(Z-q_{\mathcal{G},\delta}(X))\mathbb{P}(X\ge q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}))+\mathbb{E}\big[(X-Z)^+\,\big|\,\mathcal{G}\big]\big)\\
&\quad\le 1_{M^c}\big(Z(1-\delta)+\mathbb{E}\big[(X-Z)^+\,\big|\,\mathcal{G}\big]\big)\quad\text{a.s.},
\end{aligned}$$

since $\mathbb{P}(X < q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) < \delta$ a.s. because otherwise we would get a contradiction to the minimality of $q_{\mathcal{G},\delta}(X)$. Moreover, $\mathbb{P}(X < q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) < \delta$ implies that the inequality above is strict on M^c which finally gives the result. □

Remarks 5.25. (Conditional expected shortfall).

(a) Whilst many papers assume bounded losses for conditional expected shortfall, the results in Lemma 5.23 apply to all random variables.

(b) Note that properties (a), (b), (c), (d) and (i) in Lemma 5.23 imply that conditional expected shortfall is a conditional coherent risk measure. In the work of Acciaio and Penner [3, Example 1.10] it is also shown that conditional expected shortfall is conditionally coherent.

(c) Lemma 3.15, Corollary 5.14 and Lemma 5.23(s) imply $\mathrm{ES}_\delta[X\,|\,\mathcal{G}] \nearrow X^{\mathcal{G}}$ a.s. given that $\delta \nearrow 1$ a.s. with corresponding topology on $\overline{\mathbb{R}} = \mathbb{R}\cup\{\infty\}$.

(d) Given the same setup as in the lemma above, assume that $X \le_{\mathrm{icx}(\mathcal{G})} Y$ or $X \le_{\mathrm{cx}(\mathcal{G})} Y$, see Definition 10.14. Then Lemmas 5.12, 5.21 and 5.23(i) imply

$$\int_{[\delta,1)} q_{\mathcal{G},t}(X)\,dt \le \int_{[\delta,1)} q_{\mathcal{G},t}(Y)\,dt \quad \text{a.s.}$$

Note that this result is also valid on $\{\delta = 0\}$ and $\{\delta = 1\}$, given that integrals exist for the latter. For trivial $\mathcal{G}$, see Footnote 17, this result is closely related to alternative representations of increasing convex orders and convex orders, see Shaked and Shanthikumar [113, Theorems 3.A.5 and 4.A.3].

(e) Time-consistency. In general, conditional expected shortfall is neither time-consistent nor recursive as Example 8.3 will show. Note that time-consistency and recursiveness are equivalent, cf. Delbaen [31, Section 6]. But, as stated in Lemma 3.5, at level $\delta = 1$ conditional expected shortfall is time-consistent even in a continuous-time setting. Cheridito and Stadje [24, Sections 4–5] show that there exist time-consistent alternatives to conditional expected shortfall in a discrete-time setting. Also assuming a discrete-time setting, Roorda and Schumacher [105, Definitions 8.2 and 8.4] define dynamically consistent tail value at risk and sequentially consistent tail value at risk which give dynamically consistent and sequentially consistent alternatives to conditional expected shortfall. In Acciaio and Penner [3, Example 1.38(2)] it is shown that, in a dynamic setting, stochastic levels $(\delta_t)_{t\in\mathbb{N}}$ which vary over time can lead to middle and weak acceptance consistency of conditional expected shortfall. They also show that, in general, conditional expected shortfall is not even weakly time consistent.

(f) In Detlefsen and Scandolo [35, Proposition 2] it is shown that a conditional risk measure satisfies regularity if it is normalised, translation invariant, monotone and convex, all meant in a conditional sense.

(g) Similarly as in Remarks 3.3(j), the result of (o) takes a look at conditional expected shortfall of conditional expectation. Note that this property does not hold for conditional lower quantiles, see Remarks 3.20(b), and therefore conditional distortion risk measures, in general.

(h) Scenario representations of Lemma 5.23(t1) to (t3) are equivalent to the widely used dual definition of conditional expected shortfall as given, for example, in Acciaio and Penner [3, Example 1.10].

(i) For an economic interpretation of the scenario representation given in Lemma 5.23(t4), assume that based on the given information $\mathcal{G}$ you can choose an amount Z and enter into a special stop-loss insurance contract such that whenever your loss X is above Z you must pay the fair insurance premium $\mathbb{E}[(X-Z)^+ \,|\, \mathcal{G}]$ multiplied by security loading factor $1/(1-\delta)$. In exchange you receive amount $X - Z$ to cover losses above Z. Of course, this deal may deliver positive, as well as negative returns. If Z is chosen too high, then the deductible is high in the case when $X > Z$ happens. Conversely, if Z is too small, the premium is high when $X > Z$ happens. The optimal solution is given by Lemma 5.23(t4), i.e., $q_{\mathcal{G},t}(X) \leq Z \leq q^{\mathcal{G},t}(X)$ a.s.

(j) If X is $\mathcal{G}$-measurable, then $\mathrm{ES}_\delta[X \,|\, \mathcal{G}] = X$ a.s. This immediately follows from Lemma 5.23(c).

(k) In the classical, unconditional case on an atomless probability space, it can be shown that worst conditional expectation—a risk measure closely related to expected shortfall—is the smallest coherent risk measure which is law-determined, satisfies the Fatou property and dominates value at risk, see, for example, Delbaen [30, Theorem 6.10], as well as Artzner et al. [8, Proposition 5.4] for a similar result. In Delbaen [30, Theorem 6.8] it is shown that value at risk itself is the minimum of all coherent risk measures which dominate value at risk and satisfy the Fatou property, i.e., there exists no smallest coherent risk measure with the Fatou property that dominates value at risk.

Corollary 5.26 (Nice version II)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $I = [0, \infty)$. Consider an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$ and let $(\delta_t)_{t\in I}$ be a $[0,1]$-valued process on $(\mathcal{G})_{t\in I}$ with increasing, càdlàg or continuous paths. Then, there exists a version of the process $(\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{G}])_{t\in I}$ with increasing, càdlàg or continuous paths, respectively, with corresponding topology on $\overline{\mathbb{R}}$.*

Proof. Let $\omega \in \Omega$. Using Corollary 5.14 or the conditional quantile representation of conditional expected shortfall as given in Lemma 5.12, alternatively, we get the result for all $\{t \in I \,|\, 0 < \delta_t(\omega) < 1\}$. Thus, as $\mathrm{ES}_0[X \,|\, \mathcal{G}] \leq \mathrm{ES}_{\delta_t}[X \,|\, \mathcal{G}] \leq \mathrm{ES}_1[X \,|\, \mathcal{G}]$ a.s., we may find a version of the process $(\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{G}])_{t\in I}$ with increasing paths if $(\delta_t)_{t\in I}$ has increasing paths. Càdlàg or continuous paths, whenever $(\delta_t)_{t\in I}$ approaches zero from the right or one from the left, follow by the definition of conditional expected shortfall as an essential infimum at $\{\delta_t = 0\}$ or by Remarks 5.25(c), respectively, using the corresponding topology on $\overline{\mathbb{R}}$. $\square$

The following corollary gives the *supermartingale property*, see Example 2.3, of conditional expected shortfall. Note that this result holds simultaneously for weighted conditional expected shortfall. Detlefsen and Scandolo [35] provide a general approach towards the supermartingale property in a discrete-time setting. They give sufficient conditions for the supermartingale property to hold and they show that this property is satisfied by the dynamic entropic risk measure.

Corollary 5.27 (Supermartingale property)**.** *Setting $I = [0, \infty)$, consider a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in I}$ as well as an $\mathcal{F}$-measurable $X \colon \Omega \to \mathbb{R}$ such that X is σ-integrable with respect to $\mathcal{F}_0$. Moreover, let $(\delta_t)_{t \in I}$ be a $[0, 1]$-valued $\mathbb{F}$-adapted process with decreasing paths. Then, the process $(\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_t])_{t \in I}$ is a supermartingale.*

Proof. Note that the process $(\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_t])_{t \in I}$ is $\mathbb{F}$-adapted. Using the tower property of general conditional expectation, see [65, Theorem 1.22], $(\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_t])^-$ is σ-integrable with respect to $\mathcal{F}_s$, for every $s, t \in I$ with $s \leq t$. From Lemma 5.23(n) and Corollary 5.26 we obtain, for all $s, t \in I$ with $s \leq t$,

$$\mathbb{E}[\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_t] \,|\, \mathcal{F}_s] \leq \mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_s] \leq \mathrm{ES}_{\delta_s}[X \,|\, \mathcal{F}_s] \quad \text{a.s.},$$

which gives the supermartingale property. □

Remark 5.28. Note that the result of Corollary 5.27 in particular includes the case when $(\delta_t)_{t \in I}$ has constant paths. Moreover, if the filtration $(\mathcal{F}_t)_{t \in I}$ is right-continuous, if $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets and if the mapping $t \to \mathbb{E}[\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_t]]$ is right-continuous, then there exists a càdlàg version of $(\mathrm{ES}_{\delta_t}[X \,|\, \mathcal{F}_t])_{t \in I}$, cf. Klenke [73, Theorem 21.24].

Not surprisingly, taking a look at conditional expected shortfall of submartingales gives the following result.

Corollary 5.29 (Prospective increase in uncertainty for submartingales)**.** *Given $I := [0, \infty)$ and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in I}$, let $(M_t)_{t \in I}$ be an $\mathbb{F}$-submartingale. Then,*

$$\mathrm{ES}_{\delta}[M_{t_1} \,|\, \mathcal{F}_s] \leq \mathrm{ES}_{\delta}[M_{t_2} \,|\, \mathcal{F}_s] \quad \textit{a.s.},$$

for all $s, t_1, t_2 \in I$ with $s \leq t_1 \leq t_2$ and $\mathcal{F}_s$-measurable $\delta \colon \Omega \to [0, 1]$.

Remark 5.30 (Fair game). In particular, the result above holds for $\mathbb{F}$-martingales which implies that risk associated with a fair game, measured in terms of conditional expected shortfall, increases over time.

Proof of Corollary 5.29. The result immediately follows by Lemma 5.23(i) and (o). □

Chapter 6

Weighted Conditional Expected Shortfall

In this chapter we define weighted conditional expected shortfall which has not been previously discussed in the literature and which is based on the classical, unconditional approach as in Cherny and Madan [25, Example 2.6]. Since it is a weighted average of conditional expected shortfall at different stochastic levels, all properties from the previous chapter can be directly transferred to weighted conditional expected shortfall. This class thus provides a flexible family of conditional coherent risk measures satisfying numerous properties. We point out the link to conditional distortion risk measures and discuss beta-weighted conditional expected shortfall, motivated by the classical case as in Cherny and Madan [25, Example 2.9].

6.1 Definition of weighted conditional expected shortfall

Recalling Definition 5.3, we observe that conditional expected shortfall depends solely on a single level δ of risk-aversion. As several levels might be interesting for measuring risk, we go one step further and introduce *weighted conditional expected shortfall* based on the analogous, unconditional approach as in Cherny and Madan [25, Example 2.6], called weighted value at risk. Under mild conditions, weighted conditional expected shortfall arises as a special case of conditional distortion risk measures as shown in Lemma 6.8.

Definition 6.1 (Suitable subspace)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $G\colon [0,1] \times \Omega \to [0,1]$ be a $(\mathcal{G})_{t\in[0,1]}$-adapted *weighting process* which has increasing, right-continuous paths with boundary conditions $G(0,\cdot) = 0$ a.s. and $G(1,\cdot) = 1$ a.s. By $\mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ we denote the set of all $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$ with

$$\int_{[0,1]} \mathrm{ES}^-_t[X\,|\,\mathcal{G}]\; G(dt,\cdot) < \infty \quad \text{a.s.},$$

where $\mathrm{ES}^-_t[X\,|\,\mathcal{G}] := \max\{0, -\mathrm{ES}_t[X\,|\,\mathcal{G}]\}$ and where $(\mathrm{ES}_t[X\,|\,\mathcal{G}])_{t\in[0,1]}$ denotes the version of conditional expected shortfall at level t with continuous paths as given in Lemma 5.12. By $\mathcal{L}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ we denote the set of all $X \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ for which

$$\int_{[0,1]} |\,\mathrm{ES}_t[X\,|\,\mathcal{G}]\,|\; G(dt,\cdot) < \infty \quad \text{a.s.}$$

Definition 6.2 (Weighted conditional expected shortfall)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and a weighting process G with the properties specified in Definition 6.1, let $X \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ and let $(\mathrm{ES}_t[X\,|\,\mathcal{G}])_{t\in[0,1]}$ denote the version of conditional expected shortfall at level t with continuous paths as given in Lemma 5.12. Then, *conditional G-weighted expected shortfall* is defined by the pathwise Lebesgue–Stieltjes integral

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] := \int_{[0,1]} \mathrm{ES}_t[X\,|\,\mathcal{G}]\; G(dt,\cdot)\,.$$

Remark 6.3. As $X \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$, we know that the negative part of the integral is finite while the positive part may still be infinite.

Remark 6.4 (Conditional expected shortfall as a special case). Given $\mathcal{G}$-measurable level $\delta\colon \Omega \to (0,1]$ and setting $G(t,\cdot) := 1_{[\delta,1]}(t)$ for all $t \in [0,1]$ and $\omega \in \Omega$, then weighted conditional expected shortfall simplifies to conditional expected shortfall, i.e.,

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}] \quad \text{a.s.}$$

6.2 Properties of weighted conditional expected shortfall

Lemma 6.5 (Properties of weighted conditional expected shortfall)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and a process G with the properties specified in Definition 6.1, let $X, Y \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$. Then, weighted conditional expected shortfall of Definition 6.2 has the following conditional properties, considering conventions of Footnote 22:*

(a) Conditional normalisation:

$$\mathrm{ES}_G[0\,|\,\mathcal{G}] = 0 \quad a.s.$$

(b) Conditional positive homogeneity: If $Z\colon \Omega \to [0,\infty)$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_G[X\,Z\,|\,\mathcal{G}] = \mathrm{ES}_G[X\,|\,\mathcal{G}]\; Z \quad a.s.$$

(c) Conditional translation invariance: If $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_G[X + Z\,|\,\mathcal{G}] = \mathrm{ES}_G[X\,|\,\mathcal{G}] + Z \quad a.s.$$

(d) Subadditivity: If $X + Y \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$

$$\mathrm{ES}_G[X + Y\,|\,\mathcal{G}] \le \mathrm{ES}_G[X\,|\,\mathcal{G}] + \mathrm{ES}_G[Y\,|\,\mathcal{G}] \quad a.s.$$

(e) Conditional convexity: If $Z\colon \Omega \to [0,1]$ is $\mathcal{G}$-measurable with integrability condition $X\,Z + Y\,(1-Z) \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$, then

$$\mathrm{ES}_G[X\,Z + Y\,(1-Z)\,|\,\mathcal{G}] \le \mathrm{ES}_G[X\,|\,\mathcal{G}]\; Z + \mathrm{ES}_G[Y\,|\,\mathcal{G}]\;(1-Z) \quad a.s.$$

(f) Comonotonic additivity: If X and Y are continuously distributed and if they are comonotonic, see Definition 10.11, then

$$\mathrm{ES}_G[X + Y\,|\,\mathcal{G}] = \mathrm{ES}_G[X\,|\,\mathcal{G}] + \mathrm{ES}_G[Y\,|\,\mathcal{G}] \quad a.s.$$

(g) *Conditionally comonotonic additivity: If X and Y are conditionally comonotonic with respect to $\mathcal{G}$, see Definition 10.11, then*

$$\mathrm{ES}_G[X+Y\,|\,\mathcal{G}] = \mathrm{ES}_G[X\,|\,\mathcal{G}] + \mathrm{ES}_G[Y\,|\,\mathcal{G}] \quad a.s.$$

(h) *Independence: If X is independent of $\mathcal{G}$, then*

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] = \int_{[0,1]} \mathrm{ES}_t[X]\; G(dt,\cdot) \quad a.s.,$$

where $\mathrm{ES}_t[X]$ denotes the unconditional expected shortfall at deterministic level $t \in [0,1]$, cf. Schmock [111, Section 7.2], i.e., $\mathrm{ES}_t[X] = \mathrm{ES}_t[X\,|\,\mathcal{H}]$ a.s. where $\mathcal{H}$ denotes a trivial σ-algebra, see Footnote 17.

(i) *Conditional monotonicity: If either $X \leq_{\mathrm{icx}(\mathcal{G})} Y$ or $X \leq_{\mathrm{cx}(\mathcal{G})} Y$, and if in addition $X1_{\{\delta=1\}} \leq_{\mathrm{st}(\mathcal{G})} Y1_{\{\delta=1\}}$, see Definition 10.14, then*

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] \leq \mathrm{ES}_G[Y\,|\,\mathcal{G}] \quad a.s.$$

In particular, the assumptions are satisfied if $X \leq Y$ a.s.

(j) *Determined by conditional law: If $\mathbb{E}[f(X)|\mathcal{G}] = \mathbb{E}[f(Y)|\mathcal{G}]$ a.s. for every bounded and continuous function $f\colon \mathbb{R} \to \mathbb{R}$, then*

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] = \mathrm{ES}_G[Y\,|\,\mathcal{G}] \quad a.s.$$

(k) *Regularity: If $A \in \mathcal{G}$, then $X1_A = Y1_A$ a.s. implies*

$$\mathrm{ES}_G[X\,|\,\mathcal{G}]\,1_A = \mathrm{ES}_G[Y\,|\,\mathcal{G}]\,1_A \quad a.s.$$

(l) *If $f\colon \mathbb{R} \to \mathbb{R}$ is strictly increasing, as well as convex and if $X \in \mathcal{L}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ a.s., then*

$$f(\mathrm{ES}_G[X\,|\,\mathcal{G}]) \leq \mathrm{ES}_G[f(X)\,|\,\mathcal{G}] \quad a.s.$$

If f is concave, then the reverse inequality holds where for $\mathrm{ES}_G[f(X)|\mathcal{G}]$ to exist we require $f(X) \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$.

(m) *Supermartingale: Let $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ be two sub-σ-algebras and let G be $(\mathcal{H})_{t\in[0,1]}$-adapted. If $\mathrm{ES}_G[X\,|\,\mathcal{G}]$ and $(\mathrm{ES}_t[X\,|\,\mathcal{G}])^-$ for all $t \in [0,1]$ are σ-integrable with respect to $\mathcal{H}$, then*

$$\mathbb{E}[\mathrm{ES}_G[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \leq \mathrm{ES}_G[X\,|\,\mathcal{H}] \quad a.s.$$

(n) *Uncertainty decrease of projections: Let $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ be two σ-algebras and let G be $(\mathcal{H})_{t\in[0,1]}$-adapted. If X^- is σ-integrable with respect to $\mathcal{G}$ and if $\mathbb{E}[X\,|\,\mathcal{G}] \in \mathcal{L}^-_{\mathcal{H},G,\mathrm{wces}}(\mathbb{P})$, then*

$$\mathrm{ES}_G[\mathbb{E}[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] \leq \mathrm{ES}_G[X\,|\,\mathcal{H}] \quad a.s.$$

(o) *Conditional Fatou I: Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables bounded from below by some $\mathcal{G}$-measurable $C\colon \Omega \to \mathbb{R}$ converging to X in conditional probability, i.e., $\lim_{n\to\infty} \mathbb{P}(|X - X_n| \geq \varepsilon\,|\,\mathcal{G}) = 0$ a.s. for every $\varepsilon > 0$, see Footnote 23. Then,*

$$\liminf_{n\to\infty} \mathrm{ES}_G[X_n\,|\,\mathcal{G}] \geq \mathrm{ES}_G[X\,|\,\mathcal{G}] \quad a.s.$$

(p) *Conditional Fatou II: Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random variables bounded from below by some* $\mathcal{G}$*-measurable* $C\colon \Omega \to \mathbb{R}$*. Then,* $X := \liminf_{n\to\infty} X_n$ *satisfies*

$$\liminf_{n\to\infty} \mathrm{ES}_G[X_n\,|\,\mathcal{G}] \geq \mathrm{ES}_G[X\,|\,\mathcal{G}] \quad a.s.$$

(q) *Continuity from below: Let* $(X_n)_{n\in\mathbb{N}}$ *be an increasing sequence of real-valued random variables bounded from below by some* $\mathcal{G}$*-measurable* $C\colon \Omega \to \mathbb{R}$ *converging to* X *from below, i.e.,* $X_n \nearrow X$ *a.s. as* $n \to \infty$*. Then,*

$$\lim_{n\to\infty} \mathrm{ES}_G[X_n\,|\,\mathcal{G}] = \mathrm{ES}_G[X\,|\,\mathcal{G}] \quad a.s.$$

(r) *Bounds: If the left side in the inequality below is well-defined, then*

$$\int_{[0,1]} q_{t,\mathcal{G}}(X)\, G(dt,\cdot) \leq \mathrm{ES}_G[X\,|\,\mathcal{G}] \leq \mathbb{E}[X^+\,|\,\mathcal{G}] \int_{[0,1]} \frac{1}{1-t}\, G(dt,\cdot) \quad a.s.,$$

(s) *Conditional quantile representations:*

(1) $\mathrm{ES}_G[X\,|\,\mathcal{G}] = \int_{[0,1)} \frac{1}{1-t} \int_{[t,1)} q_{s,\mathcal{G}}(X)\, ds\, G(dt,\cdot)$ *on* $\{G(1-,\cdot) = 1\}$ *a.s.*

(2) $\mathrm{ES}_G[X\,|\,\mathcal{G}] = \int_{[0,1)} q_{s,\mathcal{G}}(X) \int_{[0,s]} \frac{G(dt,\cdot)}{1-t}\, ds$ *on* $\{G(1-,\cdot) = 1\}$ *a.s.*

(3) $\mathrm{ES}_G[X\,|\,\mathcal{G}] = \int_{[0,1)} q_{s,\mathcal{G}}\big(X \int_{[0,s]} \frac{G(dt,\cdot)}{1-t}\big)\, ds$ *on* $\{G(1-,\cdot) = 1\}$ *a.s.*

Proof. All properties follow by corresponding results in Lemma 5.23 and pathwise fundamental properties of Lebesgue–Stieltjes integrals such as linearity, monotonicity, monotone convergence, Jensen's inequality or Fatou's lemma, respectively. As $(\mathrm{ES}_\delta[X_t\,|\,\mathcal{G}])_{t\in[0,1]}$ is jointly measurable due to increasing and continuous paths, note that, for Item (m), the conditional Fubini theorem of Lemma 10.4 can be used to deduce

$$\mathbb{E}[\mathrm{ES}_G[X\,|\,\mathcal{G}]\,|\,\mathcal{H}] = \int_{[0,1]} \mathbb{E}[\mathrm{ES}_t[X\,|\,\mathcal{G}]\,|\,\mathcal{H}]\; G(dt,\cdot) \quad \text{a.s.}$$

Applying Lemma 5.23(n), as well as Fubini's theorem, again, yields the corresponding result. Moreover, note that conditional quantile representations in (s) follow by the quantile representation of conditional expected shortfall, see Lemma 5.12, and by interchanging order of integration via Fubini's classical theorem, cf. Kallenberg [71, Theorem 1.27]. □

Corollary 6.6. *Weighted conditional expected shortfall is a coherent conditional risk measure.*

Remark 6.7 (Dynamic results). The results obtained in Corollaries 5.27 and 5.29 hold analogously for weighted conditional expected shortfall by applying Lemma 6.5(m) and (n), respectively.

As previously mentioned, weighted conditional expected shortfall can be written as a conditional distortion risk measure under mild conditions. This result is shown in the following lemma.

Lemma 6.8 (Link to conditional distortion risk measures)**.** *Given a sub-*σ*-algebra* $\mathcal{G} \subset \mathcal{F}$*, consider a process* G *as given in Definition 6.1 with constraint* $G(1-,\cdot) = 1$ *a.s. and let* $X \in \mathcal{L}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$*. Then,* G*-weighted conditional expected shortfall* $\mathrm{ES}_G[X\,|\,\mathcal{G}]$ *is a conditional distortion risk measure* $\rho_g[X\,|\,\mathcal{G}]$ *with pathwise concave distortion process*

$$g(t,\cdot) = \int_{[0,t)} \int_{[0,1-s]} \frac{G(dr,\cdot)}{1-r}\, ds\,, \quad t \in [0,1]\,. \tag{6.9}$$

Proof. Note that due to Lemma 6.5(s2), we may write

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] = \int_{[0,1)} q_{1-s,\mathcal{G}}(X) \int_{[0,1-s]} \frac{G(dr,\cdot)}{1-r}\, ds\,, \quad \text{a.s.}$$

Thus, recalling Lemma 4.8(o) and defining the process g by (6.9), the representation of $\mathrm{ES}_G[X\,|\,\mathcal{G}]$ as a conditional distortion risk measure follows. Obviously, g has continuous and increasing paths since the inner integral is positive. Moreover, $g(0,\cdot) = 0$ a.s. and, by Fubini's theorem,

$$g(1,\cdot) = \int_{[0,1)} \int_{[r,1)} ds\, \frac{G(dr,\cdot)}{1-r} = G(1) - G(0) = 1 \quad \text{a.s.}$$

Thus, Definition 4.5 and Lemma 4.8(o) imply that conditional G-weighted expected shortfall is a distortion risk measure. Note that since

$$(0,1) \ni t \mapsto \frac{\partial}{\partial t} g(t,\omega) = \int_{[0,1-t]} \frac{G(dr,\omega)}{1-r}\,, \quad \omega \in \Omega\,,$$

is decreasing. Thus, all paths of g are concave. □

Lemma 6.10 (Alternative representations)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider a process G as given in Definition 6.1. Let $X \in \mathcal{L}^-_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ and let the sequence of adjusted indicator functions $(f_{\mathcal{G},t,X})_{t\in[0,1]}$ be given by Definition 5.1 where we take the pathwise right-continuous version of it and define*

$$f_{\mathcal{G},G,X} := \int_{[0,1)} \frac{f_{\mathcal{G},t,X}}{1-t}\, G(dt,\cdot)\,. \tag{6.11}$$

Moreover, assume that $f_{\mathcal{G},G,X} X$ is σ-integrable with respect to $\mathcal{G}$ and, for each $t \in [0,1]$ and $\omega \in \Omega$, let $G^{\leftarrow}(t,\omega)$ be the pathwise generalised inverse of G, see Remarks 3.11(a). Then,

$$\mathrm{ES}_G[X\,|\,\mathcal{G}] = \int_{[0,1]} \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X\,|\,\mathcal{G}]\; dt = \mathbb{E}[f_{\mathcal{G},G,X}\, X\,|\,\mathcal{G}] + (1 - G(1-,\cdot))\, X^{\mathcal{G}} \quad \textit{a.s.} \tag{6.12}$$

In particular, if in addition $\mathcal{G}$ is trivial, see Footnote 17, and if $G(1-) = 1$, then we have $\mathrm{ES}_G[X] := \mathrm{ES}_G[X\,|\,\mathcal{G}] = \mathbb{E}[Y]$ where Y has distribution function $\bar{G} \circ F$ with F denoting the distribution function of X and for $\bar{G}$ see Lemma 4.8(n).

Remark 6.13. Given $t \in [0,1]$, note that $G^{\leftarrow}(t,\cdot)$ is $\mathcal{G}$-measurable as $G^{\leftarrow}(t,\omega) \le x$ is equivalent to $G(x,\omega) \ge t$ for all $x \in \mathbb{R}$ and $\omega \in \Omega$. Thus, $\{G^{\leftarrow}(t,\cdot) \le x\} = \{G(x,\cdot) \ge t\} \in \mathcal{G}$ for all $x \in \mathbb{R}$ which gives $\mathcal{G}$-measurability of $G^{\leftarrow}(t,\cdot)$.

Proof of Lemma 6.10. Note that due to Lemma 3.15 there exists a version of the process $[0,1] \ni t \mapsto f_{\mathcal{G},t,X}$ with right-continuous decreasing paths. In particular, these paths are either continuous or a right-continuous indicator function. The first equality in (6.12) follows by the definition of weighted conditional expected shortfall and by a pathwise application of a change-of-variable formula for Lebesgue–Stieltjes integrals, cf. [45]. Then, by using the definition of conditional expected shortfall, we get,

$$\int_{[0,1]} \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X\,|\,\mathcal{G}]\; dt = \int_{[0,G(1-,\cdot))} \frac{\mathbb{E}[f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}\, X\,|\,\mathcal{G}]}{1 - G^{\leftarrow}(t,\cdot)}\, dt + (1 - G(1-,\cdot))\, X^{\mathcal{G}} \quad \text{a.s.}$$

By a pathwise application of a change-of-variable formula for Lebesgue–Stieltjes integrals, cf. [45], we get

$$\int_{[0,G(1-,\cdot))} \frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}}{1-G^{\leftarrow}(t,\cdot)}\,dt = \int_{[0,1)} \frac{f_{\mathcal{G},t,X}}{1-t}\,G(dt,\cdot) = f_{\mathcal{G},G,X} \quad \text{a.s.} \tag{6.14}$$

Thus, due to the σ-integrability of $f_{\mathcal{G},G,X}\,X$, we can apply the conditional Fubini theorem of Lemma 10.4 to conclude

$$\int_{[0,G(1-,\cdot))} \frac{\mathbb{E}[f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}\,X\,|\,\mathcal{G}]}{1-G^{\leftarrow}(t,\cdot)}\,dt = \mathbb{E}[f_{\mathcal{G},G,X}\,X\,|\,\mathcal{G}] \quad \text{a.s.},$$

which gives (6.12).

If $\mathcal{G}$ is trivial, see Footnote 17, then the distortion representation of Lemma 6.8 gives

$$\mathrm{ES}_G[X] = \int_{[0,1)} F^{\leftarrow}(t)\,\bar{g}(dt)\,.$$

Thus, by a change-of-variable formula for Lebesgue–Stieltjes integrals, see Carter and van Brunt [20, Theorem 6.2.1], and since $F^{\leftarrow}(F(x)) = x$ unless $F(x-\varepsilon) = F(x)$ for some $\varepsilon > 0$ and every $x \in \mathbb{R}$, we get

$$\mathrm{ES}_G[X] = \int_{\mathbb{R}} F^{\leftarrow}(F(x))\,\bar{g}(F(dx)) = \int_{\mathbb{R}} x\,\bar{g}(F(dx)) = \mathbb{E}[Y]$$

where Y is a random variable with distribution function $\bar{g} \circ F$. □

Lemma 6.15. *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let G be a process as given in Definition 6.1. Let $X\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and consider $f_{\mathcal{G},G,X}$ as defined in (6.11). Then,*

$$\mathbb{E}[f_{\mathcal{G},G,X}\,|\,\mathcal{G}] = G(1-,\cdot) \quad \textit{a.s.}, \tag{6.16}$$

and $f_{\mathcal{G},G,X}$ and X are conditionally comonotonic with respect to $\mathcal{G}$.

Proof. Using Equation (6.14), as well as the conditional monotone convergence theorem, see, for example, [65, Theorems 1.19(1) and 1.20], we get

$$\begin{aligned}\mathbb{E}[f_{\mathcal{G},G,X}\,|\,\mathcal{G}] &= \mathbb{E}\left[\int_{[0,G(1-,\cdot))} \frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}}{1-G^{\leftarrow}(t,\cdot)}\,dt \,\middle|\, \mathcal{G}\right] \\ &= \lim_{n\to\infty} \mathbb{E}\left[\int_{[0,G(1-,\cdot)-1/n)} \frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}}{1-G^{\leftarrow}(t,\cdot)}\,dt \,\middle|\, \mathcal{G}\right] \quad \text{a.s.}\end{aligned}$$

Since the integral inside the latter conditional expectation is certainly finite, we may use the conditional Fubini theorem of Lemma 10.4 to obtain

$$\begin{aligned}\mathbb{E}[f_{\mathcal{G},G,X}\,|\,\mathcal{G}] &= \lim_{n\to\infty} \int_{[0,G(1-,\cdot)-1/n)} \frac{\mathbb{E}[f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}\,|\,\mathcal{G}]}{1-G^{\leftarrow}(t,\cdot)}\,dt \\ &= \int_{[0,G(1-,\cdot))} \frac{\mathbb{E}[f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}\,|\,\mathcal{G}]}{1-G^{\leftarrow}(t,\cdot)}\,dt \quad \text{a.s.}\end{aligned}$$

Then, since $\mathbb{E}[f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}\,|\,\mathcal{G}] = 1 - G^{\leftarrow}(t,\cdot)$ a.s., (6.16) follows.

By definition of $f_{\mathcal{G},G,X}$, there exists a $(\mathcal{G})_{t\in[0,1]}$-adapted process $(h(t,\cdot))_{t\in\mathbb{R}}$ with increasing paths such that $f_{\mathcal{G},G,X} = h(X,\cdot)$ a.s. Note that by a fundamental property of lower inverse functions, see Remarks 3.11(a) and [39], we have $1_{\{h(X,\cdot)\le x\}} = 1_{\{X\le h^{\leftarrow}(x,\cdot)\}}$ for all $x\in\mathbb{R}$ where $\mathbb{R}\ni x\mapsto h^{\leftarrow}(x,\omega)$ denotes the pathwise lower inverse of $h(\cdot,\omega)$ for all $\omega\in\Omega$. Thus, for all $x,y\in\mathbb{R}$, we get

$$\begin{aligned}\mathbb{P}(f_{\mathcal{G},G,X}\le x, X\le y\,|\,\mathcal{G}) &= \mathbb{P}(X\le \min\{h^{\leftarrow}(x,\cdot),y\}\,|\,\mathcal{G})\\ &= \min\big\{\mathbb{P}(f_{\mathcal{G},G,X}\le x\,|\,\mathcal{G}),\mathbb{P}(X\le y,\,|\,\mathcal{G})\big\}\quad \text{a.s.},\end{aligned}$$

which states that $f_{\mathcal{G},G,X}$ and X are conditionally comonotonic given $\mathcal{G}$. □

Remark 6.17. For a related statement in the context of conditional monotonicity with the additional assumption of the existence of a transition kernel, see Jouini and Napp [68].

Lemma 6.18 (Conditional monotonicity in $f_{\mathcal{G},G,X}$)**.** *Given a sub-σ-algebra $\mathcal{G}\subset\mathcal{F}$, consider a process G as defined in Definition 6.1 with additional constraint $G(1-,\cdot)=1$ a.s. Consider an $X\in\mathcal{L}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$ with $X\ge 0$ a.s. and let $f_{\mathcal{G},G,X}$ be given by Lemma 6.10. Moreover, consider a measurable function $f\colon\Omega\to[0,\infty)$ such that $f\le_{\mathrm{icx}(\mathcal{G})} f_{\mathcal{G},G,X}$ where $\le_{\mathrm{icx}(\mathcal{G})}$ denotes conditional increasing convex order, see Definition 10.14. Then,*

$$\mathbb{E}[f\,X\,|\,\mathcal{G}]\le\mathbb{E}[f_{\mathcal{G},G,X}\,X\,|\,\mathcal{G}]=\mathrm{ES}_G[X\,|\,\mathcal{G}]\quad \textit{a.s.} \tag{6.19}$$

Proof. On $\{\mathrm{ES}_G[X\,|\,\mathcal{G}]=\infty\}$ the result is clear. Thus, we may assume $\mathrm{ES}_G[X\,|\,\mathcal{G}]<\infty$ a.s. Since $G(1-,\cdot)=1$ a.s. by assumption and since X and $f_{\mathcal{G},G,X}$ are conditionally comonotonic with respect to $\mathcal{G}$, see Lemma 6.15, we can apply Lemma 10.15(d) to obtain the result. □

As a particular example of weighted conditional expected shortfall we define *beta-weighted conditional expected shortfall.* This conditional risk measure is motivated by beta-value-at-risk and alpha-value-at-risk, see Cherny and Madan [25, Example 2.9], which arise as special cases.

Definition 6.20 (Beta-weighted conditional expected shortfall)**.** Consider a sub-σ-algebra $\mathcal{G}\subset\mathcal{F}$. For $\mathcal{G}$-measurable α,β with $\alpha>\beta>-1$ a.s., let $B_{\alpha,\beta}$ denote the conditional beta distribution with parameters $\alpha-\beta$ and $\beta+1$, i.e.,

$$B_{\alpha,\beta}(t)=\int_0^t\frac{1}{\mathrm{B}(\alpha-\beta,\beta+1)}x^{\alpha-\beta-1}(1-x)^{\beta}\,dx\,,\quad t\in[0,1]\,,$$

where B denotes the beta function. Then, *beta-weighted conditional expected shortfall* of a random variable $X\in\mathcal{L}_{\mathcal{G},B_{\alpha,\beta},\mathrm{wces}}(\mathbb{P})$ is defined as $\mathrm{ES}_{B_{\alpha,\beta}}[X\,|\,\mathcal{G}]$.

The following lemma is a generalisation to the result which is obtained in Cherny and Madan [25, Example 2.9]. In particular, if sub-σ-algebra $\mathcal{G}$ is trivial, see Footnote 17, then our result coincides with the result given in their paper.

Lemma 6.21. *Given a sub-σ-algebra $\mathcal{G}\subset\mathcal{F}$ and an $\mathcal{F}$-measurable $X\colon\Omega\to\mathbb{R}$ such that X^- is σ-integrable with respect to $\mathcal{G}$, let $\alpha,\beta\colon\Omega\to\mathbb{N}_0$ be $\mathcal{G}$-measurable with $\alpha>\beta$ a.s. and let $(X_i)_{i\in\mathbb{N}}$ be a sequence of $\mathcal{F}$-measurable i.i.d. copies of X. Moreover, let the $\mathcal{F}$-measurable random variable $\xi\colon\Omega\to\mathbb{N}$ be conditionally independent of X and $(X_i)_{i\in\mathbb{N}}$ given $\mathcal{G}$ with conditional distribution $\mathbb{P}(\xi=k\,|\,\mathcal{G})=\frac{1}{\beta}1_{\{\alpha-\beta+1,\dots,\alpha\}}(k)$. Then,*

$$\mathrm{ES}_{B_{\alpha,\beta}}[X\,|\,\mathcal{G}]=\mathbb{E}\left[\frac{1}{\beta}\sum_{i=\alpha-\beta+1}^{\alpha}X_{(i)}\,\middle|\,\mathcal{G}\right]=\mathbb{E}[X_{(\xi)}\,|\,\mathcal{G}]\quad \textit{a.s.},$$

where $B_{\alpha,\beta}$ is defined in Definition 6.20 and where $X_{(1)}, \dots, X_{(\alpha)}$ denote the order statistics of $X_1, \dots, X_\alpha$ satisfying $X_{(1)} \le \dots \le X_{(\alpha)}$ a.s.

Remark 6.22. Fixing $\beta = 1$ results in the so-called *conditional alpha-value-at-risk*, see again [25, Example 2.9] for the classical case. It can be written as

$$\mathrm{ES}_{B_{\alpha,1}}[X\,|\,\mathcal{G}] = \mathbb{E}[\max\{X_1, \dots, X_\alpha\}\,|\,\mathcal{G}] \quad \text{a.s.}$$

Proof of Lemma 6.21. The second equality immediately follows if we use decomposition $1 = \sum_{k=\alpha-\beta+1}^{\alpha} 1_{\{\xi=k\}}$ together with the stated conditional independence assumption of ξ, as well as its uniform distribution. For the first equality, if necessary, enlarge the probability space such that all following random variables exist and are $\mathcal{F}$-measurable. Let $(U_i)_{i\in\mathbb{N}}$ be a vector of $\mathcal{F}$-measurable random variables which are mutually conditionally independent and conditionally independent of ξ given $\mathcal{G}$, as well as uniformly distributed on $[0,1]$. Given $\mathcal{G}$, let $U_{(1)}, \dots, U_{(\alpha)}$ denote the order statistics of $U_1, \dots, U_\alpha$. Similarly as in [25, Example 2.9], we can then conclude

$$\begin{aligned}\frac{d^2}{dt^2}\mathbb{P}(U_{(\xi)} \le t\,|\,\mathcal{G}) &= \frac{1}{\beta}\sum_{k=\alpha-\beta+1}^{\alpha}\frac{d}{dt}\left(\frac{\alpha!}{(\alpha-k)!\,(k-1)!}t^{k-1}(1-t)^{\alpha-k}\right)\\ &= \frac{\alpha!}{(\alpha-\beta-1)!\,\beta!}t^{\alpha-\beta-1}(1-t)^{\beta-1} \quad \text{a.s., for all } t\in[0,1]\,.\end{aligned}$$

Thus, by the fundamental theorem of calculus,

$$\mathbb{P}(U_{(\xi)} \le t\,|\,\mathcal{G}) = \bar{g}_{\alpha,\beta}(t) \quad \text{a.s., for all } t\in[0,1]\,,$$

where

$$\bar{g}_{\alpha,\beta}(t) := \int_{[0,t)}\int_{[0,s]}\frac{B_{\alpha,\beta}(dr,\cdot)}{1-r}\,ds\,, \quad t\in[0,1]\,.$$

Recalling Lemma 10.3 and (5.17), for every $\mathcal{G}$-measurable $Z\colon \Omega\to\mathbb{R}$, we obtain

$$\mathbb{P}(X_{(\xi)} \le Z\,|\,\mathcal{G}) = \mathbb{P}\big(U_{(\xi)} \le \mathbb{P}(X\le Z\,|\,\mathcal{G})\,\big|\,\mathcal{G}\big) = \bar{g}_{\alpha,\beta}(\mathbb{P}(X\le Z\,|\,\mathcal{G})) \quad \text{a.s.}$$

Since $\bar{g}_{\alpha,\beta}$ is pathwise strictly increasing, we get

$$q_{t,\mathcal{G}}(X_{(\xi)}) = q_{\bar{g}^{\leftarrow}_{\alpha,\beta}(t),\mathcal{G}}(X) \quad \text{a.s., for all } t\in[0,1]\,.$$

Thus, by invoking Lemma 4.8(n) and by using a change-of-variable formula for Lebesgue–Stieltjes integrals, cf. [45], we get

$$\mathbb{E}[X_{(\xi)}\,|\,\mathcal{G}] = \int_{[0,1)} q_{t,\mathcal{G}}(X)\,d\bar{g}_{\alpha,\beta}(t) \quad \text{a.s.}$$

This, together with Lemma 6.8, finally yields $\mathbb{E}[X_{(\xi)}\,|\,\mathcal{G}] = \mathrm{ES}_{B_{\alpha,\beta}}[X\,|\,\mathcal{G}]$ a.s. □

Chapter 7

Contributions to Weighted Conditional Expected Shortfall

In this chapter we introduce contributions to weighted conditional expected shortfall based on the approach chosen in Schmock [111, Section 7.3] where classical, unconditional contributions to expected shortfall are discussed. We show that under mild conditions on the weighting process, weighted conditional expected shortfall contributions are a coherent allocation principle in the sense of Kalkbrener [70]. Besides other properties, we prove a directional derivative representation of weighted conditional expected shortfall contributions which then gives rise to the Euler allocation principle in a conditional setting.

7.1 Definition of contributions to weighted conditional expected shortfall

Based on the concept of contributions to expected shortfall within the classical case as discussed in Schmock [111, Section 7.3], we will now analyse *contributions to weighted conditional expected shortfall.* This conditional concept provides the possibility to derive weighted conditional expected losses caused by a subportfolio or by some individual obligors in a financial or insurance portfolio. For an axiomatic approach to risk capital allocation see Kalkbrener [70]. Contributions to conditional lower quantiles are not introduced in this thesis as they are more involved to define. Already in the unconditional case, following the approach of Tasche [119] we observe that existence of value at risk contributions depends on technical constraints. Therefore, we prioritise the concept of weighted conditional expected shortfall contributions as they can be defined in a very general way.

Definition 7.1 (Suitable subspace)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $G\colon [0,1] \times \Omega \to [0,1]$ be a $(\mathcal{G})_{t\in[0,1]}$-adapted *weighting process* with increasing, right-continuous paths, as well as boundary conditions $G(0,\cdot) = 0$ a.s. and $G(1,\cdot) = 1$ a.s. Moreover, let portfolio loss $L\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable. By $\mathcal{L}^{-}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$ we denote all $\mathcal{F}$-measurable subportfolio losses $X\colon \Omega \to \mathbb{R}$ such that $f_{\mathcal{G},G,L}X^{-}$ is σ-integrable with respect to $\mathcal{G}$ where $f_{\mathcal{G},G,L}$ is defined in (6.11). By $\mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$ we denote the cone of those $X \in \mathcal{L}^{-}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$ such that $f_{\mathcal{G},G,L}X$ is σ-integrable with respect to $\mathcal{G}$.

Definition 7.2 (Weighted conditional expected shortfall contributions)**.** For sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider weighting process G with the properties specified in Definition 7.1. Let

portfolio loss $L\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and let subportfolio loss $X \in \mathcal{L}^-_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$. Then, the *G-weighted conditional expected shortfall contribution* of subportfolio loss X to the total portfolio loss L is defined as

$$\mathrm{ES}_G[X, L\,|\,\mathcal{G}] := \mathbb{E}[f_{\mathcal{G},G,L}\,X\,|\,\mathcal{G}] + (1 - G(1-,\cdot))\,X^{\mathcal{G}}\,.$$

Definition 7.3 (Conditional expected shortfall contributions)**.** For $\mathcal{G}$-measurable level $\delta\colon \Omega \to [0,1]$ set $G(t,\cdot) = 1_{[\delta,1]}(t)$ for all $t \in [0,1]$. Given the definitions above and ignoring boundary condition $G(0,\cdot) = 0$ a.s. for a moment, let $L\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable, as well as let $X \in \mathcal{L}^-_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$. Then, the *conditional expected shortfall contribution* of subportfolio loss X to the total portfolio loss L is defined as

$$\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] := \begin{cases} \mathrm{ES}_G[X, L\,|\,\mathcal{G}] & \text{on } \{\delta > 0\}\,, \\ \operatorname{ess\,inf}_{\delta' \in (0,1)} \frac{1}{1-\delta'}\,\mathbb{E}[f_{\mathcal{G},\delta',L}\,X\,|\,\mathcal{G}] & \text{on } \{\delta = 0\}\,. \end{cases}$$

Consequently, it is easy to see that

$$\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] = \begin{cases} X^{\mathcal{G}} & \text{on } \{\delta = 1\}\,, \\ \frac{1}{1-\delta}\,\mathbb{E}[f_{\mathcal{G},\delta,L}\,X\,|\,\mathcal{G}] & \text{on } \{0 < \delta < 1\}\,, \\ \operatorname{ess\,inf}_{\delta' \in (0,1)} \frac{1}{1-\delta'}\,\mathbb{E}[f_{\mathcal{G},\delta',L}\,X\,|\,\mathcal{G}] & \text{on } \{\delta = 0\}\,. \end{cases}$$

Remarks 7.4. (Notes on conditional expected shortfall contributions).

(a) Recalling the definition of the adjusted indicator function $f_{\mathcal{G},\delta,L}$ in Definition 5.1, conditional expected shortfall contributions give the expected subportfolio loss of all outcomes which contribute to large losses in the total portfolio, i.e., losses exceeding the conditional lower quantile, given $\mathcal{G}$.

(b) Note that results for weighted conditional expected shortfall contributions, derived in the following section, directly apply to conditional expected shortfall contributions $\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}]$ as well.

(c) If $\mathcal{G}$ is trivial, see Footnote 17, then the definition of conditional expected shortfall contributions is identical with the usual definition of expected shortfall contributions $\mathrm{ES}_\delta[X, L]$ as, for example, given in Schmock [111, Section 7.3].

Example 7.5 (Illustrative Example 3.12 continued)**.** Next, we take a closer look at expected shortfall contributions of the subportfolio $Y_1 := X_1 + X_2$ to the total portfolio X, as well as their conditional generalisations. Such a subportfolio may represent a collection of companies in a specific industry or country. Defining $Y_2 := X_3 + \cdots + X_n$ yields, for all $k \in \{0,1,2\}$,

$$\begin{aligned} \mathbb{E}\big[Y_1\,1_{\{X > q_\delta(X)\}}\big] &= \sum_{k=0}^{2} k\,\mathbb{P}(Y_1 = k, X > q_\delta(X)) \\ &= \sum_{k=0}^{2} k\,\mathbb{P}(Y_1 = k)\,\mathbb{P}(Y_2 > q_\delta(X) - k)\,, \end{aligned}$$

which gives the expected shortfall contribution of subportfolio Y_1 to the total portfolio X at level $\delta = 0.95$, denoted by $\mathrm{ES}_\delta[Y_1, X]$,

$$\begin{aligned} \mathrm{ES}_\delta[Y_1, X] &= \frac{1}{1-\delta} \sum_{k=0}^{2} k\,\mathbb{P}(Y_1 = k)\,\big(\mathbb{P}(Y_2 > q_\delta(X) - k) + \beta_{\delta,X}\,\mathbb{P}(Y_2 = q_\delta(X) - k)\big) \\ &= 0.96\,. \end{aligned}$$

Analogously, the contribution of Y_2 to the total portfolio X is $\mathrm{ES}_\delta[Y_2, X] = 3.84$. Then, as

$$\begin{aligned}\mathbb{E}\big[Y_1 1_{\{X>q_\delta(X)\}}\,\big|\,\mathcal{G}\big] &= \sum_{k=0}^{2} k\,\mathbb{P}(Y_1 = k, X > q_{\mathcal{G},\delta}(X)\,|\,\mathcal{G}) \\ &= \sum_{k=0}^{2} k\,\mathbb{P}(Y_1 = k\,|\,\mathcal{G})\mathbb{P}(Y_2 > q_{\mathcal{G},\delta}(X) - k\,|\,\mathcal{G})\,,\end{aligned}$$

we may similarly obtain conditional contributions of the subportfolio Y_1 to the total portfolio X given credit rating events $\mathcal{G} = \sigma(G)$

$$\begin{aligned}&\mathrm{ES}_\delta[Y_1, X\,|\,\mathcal{G}] \\ &= \frac{1}{1-\delta}\sum_{k=0}^{2} k\,\mathbb{P}(Y_1 = k\,|\,\mathcal{G})\big(\mathbb{P}(Y_2 > q_{\mathcal{G},\delta}(X) - k) + \beta_{\mathcal{G},\delta,X}\,\mathbb{P}(Y_2 = q_\delta(X) - k\,|\,\mathcal{G})\big) \\ &= \begin{cases} 0.65 & \text{on } G\,, \\ 1.81 & \text{on } G^{\mathrm{c}}\,. \end{cases}\end{aligned}$$

Finally, conditional contributions of the subportfolio Y_2 to the total portfolio X given credit rating events $\mathcal{G}$ are given by

$$\mathrm{ES}_\delta[Y_2, X\,|\,\mathcal{G}] = \begin{cases} 2.63 & \text{on } G\,, \\ 7.24 & \text{on } G^{\mathrm{c}}\,. \end{cases}$$

As intuition suggests, depending on the given credit rating event contributions of subportfolios change but not entirely in a linear way. Moreover, recalling the results for conditional expected shortfall obtained in Example 5.4, we observe linear aggregation of subportfolios in the classical, as well as conditional case, i.e.,

$$\mathrm{ES}_\delta[X] = \mathrm{ES}_\delta[Y_1, X] + \mathrm{ES}_\delta[Y_2, X]\,,$$

as well as

$$\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \mathrm{ES}_\delta[Y_1, X\,|\,\mathcal{G}] + \mathrm{ES}_\delta[Y_2, X\,|\,\mathcal{G}]\,.$$

This result is true in general as shown in Lemma 7.7(a) and (g). Assuming that linear aggregation should hold, we could now take a naive approach and define contributions $\Pi[Y_1]$ and $\Pi[Y_2]$ of Y_1 and Y_2, respectively, as

$$\Pi[Y_i, X] := \frac{\mathrm{ES}_\delta[Y_i]}{\mathrm{ES}_\delta[Y_1] + \mathrm{ES}_\delta[Y_2]}\,\mathrm{ES}_\delta[X]\,, \quad i \in \{1, 2\},$$

i.e., as the fraction of the subportfolio risk to the aggregated subportfolio risk and scaled by the total portfolio risk. Correspondingly for the conditional case, we could define naive conditional contributions as

$$\Pi[Y_i, X\,|\,\mathcal{G}] := \frac{\mathrm{ES}_\delta[Y_i\,|\,\mathcal{G}]}{\mathrm{ES}_\delta[Y_1\,|\,\mathcal{G}] + \mathrm{ES}_\delta[Y_2\,|\,\mathcal{G}]}\,\mathrm{ES}_\delta[X\,|\,\mathcal{G}]\,, \quad i \in \{1, 2\}.$$

The results are shown in Figure 7.1 where we compare the naive approach, i.e., $\Pi[Y_i, X]$ and $\Pi[Y_i, X\,|\,\mathcal{G}]$, to conditional and unconditional expected shortfall contributions, i.e., $\mathrm{ES}_\delta[Y_i, X]$ and $\mathrm{ES}_\delta[Y_i, X\,|\,\mathcal{G}]$ for $i \in \{1, 2\}$. Obviously, both approaches give the same cumulated risk capital but contributions change. Naive contributions are less skewed, meaning that greater losses get relatively smaller allocations as in the case of expected shortfall contributions.

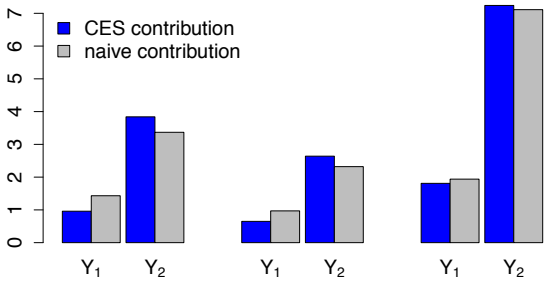


Figure 7.1: Comparison of naive risk contributions (grey) to expected shortfall and conditional expected shortfall contributions (blue), i.e., left the unconditional case with $\mathrm{ES}_\delta[Y_i, X]$ *and* $\Pi[Y_i, X]$*, in the middle the conditional case on* G*, as well as right the conditional case on* G^c *with* $\mathrm{ES}_\delta[Y_i, X\,|\,\mathcal{G}]$ *and* $\Pi[Y_i, X\,|\,\mathcal{G}]$ *for* $i \in \{1, 2\}$*.*

7.2 Properties of contributions to weighted conditional expected shortfall

Lemma 7.6. *Given the assumptions of Definition 7.3 and, in addition,* $X \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$*, note that contributions to weighted conditional expected shortfall can be represented as a weighted average of contributions to conditional expected shortfall, i.e.,*

$$\mathrm{ES}_G[X, L\,|\,\mathcal{G}] = \int_{[0,1]} \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}]\; dt\,.$$

Proof. Note that by definition of weighted conditional expected shortfall contributions and by definition of $f_{\mathcal{G},G,L}$ in (6.11) we have

$$\begin{aligned}\mathrm{ES}_G[X, L\,|\,\mathcal{G}] &= \mathbb{E}[f_{\mathcal{G},G,L}\, X\,|\,\mathcal{G}] + (1 - G(1-,\cdot))X^{\mathcal{G}} \\ &= \mathbb{E}\left[\int_{[0,G(1-,\cdot))} \frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),L}}{1 - G^{\leftarrow}(t,\cdot)}\, X\, dt \,\middle|\, \mathcal{G}\right] + (1 - G(1-,\cdot))X^{\mathcal{G}}\,, \quad \text{a.s.}\end{aligned}$$

Now, recall Equation 6.14 and note that we can find a measurable version of

$$\left(\mathbb{E}\left[\frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),L}}{1 - G^{\leftarrow}(t,\cdot)}\, X \,\middle|\, \mathcal{G}\right]\right)_{t\in[0,G(1-,\cdot))}$$

as there exist versions of $(G^{\leftarrow}(t,\cdot))_{t\in[0,G(1-,\cdot))}$, as well as $(f_{\mathcal{G},G^{\leftarrow}(t,\cdot),L})_{t\in[0,G(1-,\cdot))}$ with increasing paths. Thus, together with σ-integrability of $f_{\mathcal{G},G,L}\, X$, we can apply the conditional Fubini theorem, see Lemma 10.4, to obtain

$$\begin{aligned}\mathrm{ES}_G[X, L\,|\,\mathcal{G}] &= \int_{[0,G(1-,\cdot))} \mathbb{E}\left[\frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),L}}{1 - G^{\leftarrow}(t,\cdot)}\, X \,\middle|\, \mathcal{G}\right] dt + (1 - G(1-,\cdot))\, X^{\mathcal{G}} \\ &= \int_{[0,1]} \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}]\; dt \quad \text{a.s.},\end{aligned}$$

which gives the result. □

Having proved the latter result, some of the following results can be derived by reducing the problem to conditional expected shortfall contributions whilst others have to be proven directly.

Lemma 7.7 (Properties of weighted conditional expected shortfall contributions)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider a process G with the properties specified in Definition 7.1. Let $L\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and let $X, Y \in \mathcal{L}^-_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$. Then, weighted conditional expected shortfall contributions have the following conditional properties:*

(a) Consistency with weighted conditional expected shortfall:

$$\mathrm{ES}_G[L, L\,|\,\mathcal{G}] = \mathrm{ES}_G[L\,|\,\mathcal{G}]\,.$$

(b) Diversification: If $X \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$, then

$$\mathrm{ES}_G[X, L\,|\,\mathcal{G}] \leq \mathrm{ES}_G[X\,|\,\mathcal{G}] \quad a.s.$$

For conditional expected shortfall contributions it suffices if $X \in \mathcal{L}^-_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$.

(c) Invariance of portfolio scale: If $Z\colon \Omega \to [0,\infty)$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_G[X, L\,Z\,|\,\mathcal{G}] = \mathrm{ES}_G[X, L\,|\,\mathcal{G}] \quad a.s.$$

(d) Portfolio translation invariance: If $Z\colon \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_G[X, L + Z\,|\,\mathcal{G}] = \mathrm{ES}_G[X, L\,|\,\mathcal{G}] \quad a.s.$$

(e) Conditional positive homogeneity: If $Z\colon \Omega \to [0,\infty)$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_G[X\,Z, L\,|\,\mathcal{G}] = \mathrm{ES}_G[X, L\,|\,\mathcal{G}]\;Z \quad a.s.$$

(f) Conditional translation invariance: If $Z\colon \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable, then

$$\mathrm{ES}_G[X + Z, L\,|\,\mathcal{G}] = \mathrm{ES}_G[X, L\,|\,\mathcal{G}] + Z \quad a.s.$$

(g) Conditional linearity: If $Z_1, Z_2\colon \Omega \to [0,\infty)$ are $\mathcal{G}$-measurable and if in addition $X\,Z_1 + Y\,Z_2 \in \mathcal{L}^-_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$, then

$$\mathrm{ES}_G[X\,Z_1 + Y\,Z_2, L\,|\,\mathcal{G}] \leq \mathrm{ES}_G[X, L\,|\,\mathcal{G}]\;Z_1 + \mathrm{ES}_G[Y, L\,|\,\mathcal{G}]\;Z_2 \quad a.s.$$

If $G(1-,\cdot) = 1$ a.s., then we have equality above. If $X, Y \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$, the result holds for all $\mathcal{G}$-measurable Z_1 and Z_2.

(h) Conditional monotonicity: If $X \leq Y$ a.s., then

$$\mathrm{ES}_G[X, L\,|\,\mathcal{G}] \leq \mathrm{ES}_G[Y, L\,|\,\mathcal{G}] \quad a.s.$$

(i) Conditionally uncorrelated: If X and $f_{\mathcal{G},G,L}$ are conditionally uncorrelated given $\mathcal{G}$, see Footnote 28, then

$$\mathrm{ES}_G[X, L\,|\,\mathcal{G}] = G(1-,\cdot)\,\mathbb{E}[X\,|\,\mathcal{G}] + (1 - G(1-,\cdot))X^{\mathcal{G}} \quad a.s.$$

(j) Subportfolio continuity: If $G(1-,\cdot) = 1$ a.s. and if $X, Y \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$, then

$$\big|\,\mathrm{ES}_G[X, L\,|\,\mathcal{G}] - \mathrm{ES}_G[Y, L\,|\,\mathcal{G}]\,\big| \leq \mathrm{ES}_G\big[|X - Y|, L\,\big|\,\mathcal{G}\big] \leq \mathrm{ES}_G\big[|X - Y|\,\big|\,\mathcal{G}\big] \quad a.s.$$

(k) *Portfolio continuity: Given* $X \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$*, assume that in addition we either have* $\mathbb{P}(L \le q_{\mathcal{G},G^{\leftarrow}(t,\cdot)}(L)\,|\,\mathcal{G}) = G^{\leftarrow}(t,\cdot)$ *a.s. or that* X *is a.s. constant on* $\{L = q_{\mathcal{G},G^{\leftarrow}(t,\cdot)}(L)\}$ *for all* $t \in [0,1]$*. Then, weighted expected shortfall contributions of* X *are continuous at* L*, meaning that for every sequence* $(L_n)_{n\in\mathbb{N}}$ *of real-valued random variables converging to* L *in conditional probability, i.e.,* $\lim_{n\to\infty}\mathbb{P}(|L-L_n| \ge \varepsilon\,|\,\mathcal{G}) = 0$ *for every* $\varepsilon > 0$*, such that* $X \in \mathcal{L}_{\mathcal{G},G,L_n,\mathrm{contr}}(\mathbb{P})$ *for all* $n \in \mathbb{N}$ *it holds that*

$$\mathrm{ES}_G[X,L\,|\,\mathcal{G}] = \lim_{n\to\infty}\mathrm{ES}_G[X,L_n\,|\,\mathcal{G}] \quad \text{in } L^1\,.$$

(l) *Representation of weighted conditional expected shortfall contribution by directional derivative: Let* $G(1-,\cdot) = 1$ *a.s. and suppose that* $X \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$*. If, for all* $t \in [0,1]$*, either* $\mathbb{P}(L \le q_{\mathcal{G},G^{\leftarrow}(t,\cdot)}(L)\,|\,\mathcal{G}) = G^{\leftarrow}(t,\cdot)$ *a.s. or if* X *is a.s. constant on* $\{L = q_{\mathcal{G},G^{\leftarrow}(t,\cdot)}(L)\}$*, then*

$$\mathrm{ES}_G[X,L\,|\,\mathcal{G}] = \lim_{n\to\infty}\frac{\mathrm{ES}_G[L+\varepsilon_n X\,|\,\mathcal{G}] - \mathrm{ES}_G[L\,|\,\mathcal{G}]}{\varepsilon_n} \quad \text{in } L^1\,,$$

for every sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ *of* $\mathcal{G}$*-measurable real-valued random variables tending to zero a.s. with* $\mathbb{P}(\bigcup_{n\in\mathbb{N}}\{\varepsilon_n = 0\}) = 0$ *and* $L, L+\varepsilon_1 X, L+\varepsilon_2 X, \dots \in \mathcal{L}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$*.*

Corollary 7.8 (Conditional coherent allocation)**.** *Considering sub-*σ*-algebra* $\mathcal{G} \subset \mathcal{F}$ *and weighting process* G *such that* $G(1-,\cdot) = 1$ *a.s., weighted conditional expected shortfall contributions satisfy all three conditional coherent allocation axioms*[31] *as follows: For a portfolio* $X = X_1 + \dots + X_n$*,* $n \in \mathbb{N}$*, with* $X \in \mathcal{L}^{-}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$*, as well as* $X_i \in \mathcal{L}^{-}_{\mathcal{G},G,X,\mathrm{contr}}(\mathbb{P})$ *for all* $i \in \{1,\dots,n\}$ *we have conditional linear aggregation*

$$\mathrm{ES}_G[X_1,X\,|\,\mathcal{G}] + \dots + \mathrm{ES}_G[X_n,X\,|\,\mathcal{G}] = \mathrm{ES}_G[X\,|\,\mathcal{G}] \quad \text{a.s.},$$

by Lemma 7.7(a) and (g), as well as diversification as given in Lemma 7.7(b) and portfolio continuity as given in Lemma 7.7(k).

Remarks 7.9. (Portfolio continuity and directional derivative).

(a) The conditions in Lemma 7.7(k) and (l), i.e., $\mathbb{P}(L \le q_{\mathcal{G},G^{\leftarrow}(t,\cdot)}(L)\,|\,\mathcal{G}) = G^{\leftarrow}(t,\cdot)$ a.s. and X being a.s. constant on $\{L = q_{\mathcal{G},G^{\leftarrow}(t,\cdot)}(L)\}$ for all $t \in [0,1]$, simplify to the easier conditions $\mathbb{P}(L \le q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) = \delta$ a.s. and X being a.s. constant on $\{L = q_{\mathcal{G},\delta}(L)\}$ if we want to have the results for conditional expected shortfall contributions $\mathrm{ES}_\delta[X,L\,|\,\mathcal{G}]$. It should also be mentioned that these properties are necessary in certain settings, see Schmock [111, Example 7.32] for a counterexample.

(b) Euler allocation[32]: For $I := \{0,\dots,T\}$, let $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in I},\mathbb{P})$ be a filtered probability space and consider a portfolio with $d \in \mathbb{N}$ discounted assets. $\mathbb{R}^d$-valued discounted asset prices $(X_t)_{t\in I}$ are $(\mathcal{F}_t)_{t\in I}$-adapted and, for $t \in I$, we write $X_t = (X_t^1,\dots,X_t^d)$. Moreover, consider an $\mathbb{R}^d$-valued trading strategy $(H_t)_{t\in I}$ where $H_t = (H_t^1,\dots,H_t^d)$ gives the number of shares held in the corresponding assets over the period $(t-1,t]$ and is $\mathcal{F}_{t-1}$-measurable for $t \in I\setminus\{0\}$. Portfolio loss $(L_t)_{t\in I}$ is then given by $L_0 = 0$ and

$$L_t = \sum_{s=1}^{t}\sum_{i=1}^{d} H_s^i \Delta' X_s^i\,, \quad t \in I\setminus\{0\}\,,$$

[31] These axioms are simply transferred from corresponding classical, unconditional axioms as given in Kalkbrener [70].

[32] For an axiomatic approach to Euler allocation see, for example, Tasche [120].

with $\Delta' X_s^i := (X_{s-1}^i - X_s^i)$. If $\Delta' X_t^i \in \mathcal{L}^-_{\mathcal{G},G,L_t,\mathrm{contr}}(\mathbb{P})$ and $(-X_t^i) \in \mathcal{L}^-_{\mathcal{G},G,L_t-L_{t-1},\mathrm{contr}}(\mathbb{P})$ for $i \in \{1,\dots,d\}$ and $t \in I \setminus \{0\}$, then

$$\mathrm{ES}_G\big[\Delta' X_i^t, L_t \,\big|\, \mathcal{F}_{t-1}\big] = X_{t-1}^i + \mathrm{ES}_G\big[-X_t^i, L_t - L_{t-1} \,\big|\, \mathcal{F}_{t-1}\big] \quad \text{a.s.},$$

by Lemma 7.7(d) and (f). Thus, at time $t \in I \setminus \{0\}$ for asset $i \in \{1,\dots,d\}$, the risk associated with the contribution given $\mathcal{F}_{t-1}$ of loss $\Delta' X_t^i$ to the total accumulated portfolio loss L_t comes down to the contribution of the i-th asset to the portfolio loss $\sum_{i=0}^d H_t^i \Delta' X_t^i$. Moreover, if all necessary assumptions for the derivative representation in Lemma 7.7(l) are satisfied, we get the Euler allocation, for all $i \in \{1,\dots,d\}$ and $t \in I \setminus \{0\}$,

$$\mathrm{ES}_G\big[\Delta' X_i^t, L_t \,\big|\, \mathcal{F}_{t-1}\big] = \frac{\partial\, \mathrm{ES}_G[L_t \,|\, \mathcal{F}_{t-1}]}{\partial H_t^i} \quad \text{a.s.}$$

This result corresponds to derivative representations of shortfall risk measures as outlined in Tasche [119] for the classical case. It shows that weighted conditional expected shortfall contributions measure the change rate of weighted conditional expected shortfall of the total portfolio loss L_t when holdings H_t^i in asset i, with $i \in \{1,\dots,d\}$, are increased or decreased at time $t \in I \setminus \{0\}$.

Proof of Lemma 7.7. (a) The result follows by Definition 7.2 and Lemma 6.10.

(b) Recall the notation of Lemma 5.7. As $f_{\mathcal{G},\delta,L} \in \mathcal{F}^1_{\mathcal{G},\delta,X}$, for all $\mathcal{G}$-measurable levels $\delta\colon \Omega \to [0,1]$, we get

$$\mathbb{E}[f_{\mathcal{G},\delta,L}\, X \,|\, \mathcal{G}] \le \operatorname*{ess\,sup}_{f \in \mathcal{F}^1_{\mathcal{G},\delta,X}} \mathbb{E}[f\, X \,|\, \mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}\, X \,|\, \mathcal{G}] \quad \text{a.s.}$$

Thus, whenever $X \in \mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$, the proof of Lemma 7.6 implies

$$\begin{aligned}
\mathbb{E}[f_{\mathcal{G},G,L} X \,|\, \mathcal{G}] &= \int_{[0,G(1-,\cdot))} \mathbb{E}\Big[\frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),L}}{1-G^{\leftarrow}(t,\cdot)} X \,\Big|\, \mathcal{G}\Big]\, dt \\
&\le \int_{[0,G(1-,\cdot))} \mathbb{E}\Big[\frac{f_{\mathcal{G},G^{\leftarrow}(t,\cdot),X}}{1-G^{\leftarrow}(t,\cdot)} X \,\Big|\, \mathcal{G}\Big]\, dt = \mathbb{E}[f_{\mathcal{G},G,X}\, X \,|\, \mathcal{G}] \quad \text{a.s.}
\end{aligned}$$

(c) and (d) It is easy to see that $f_{\mathcal{G},\delta,LZ} = f_{\mathcal{G},\delta,L}$ a.s. and $f_{\mathcal{G},\delta,L+Z} = f_{\mathcal{G},\delta,L}$ a.s. which imply $f_{\mathcal{G},G,LZ} = f_{\mathcal{G},G,L}$ a.s. and $f_{\mathcal{G},G,L+Z} = f_{\mathcal{G},G,L}$ a.s., respectively, for all $\mathcal{G}$-measurable $\delta\colon \Omega \to [0,1]$. The result then follows by Definition 7.2.

(e), (f) and (g) The statements follow from linearity of conditional expectation, for which we refer to [65, Theorem 1.18], and Remarks 3.3(f)–(i).

(h) The result follows from Lemma 6.10, Definition 7.2 and Lemma 3.18(g), as

$$\mathbb{E}[f_{\mathcal{G},G,L}\, X \,|\, \mathcal{G}] \le \mathbb{E}[f_{\mathcal{G},G,L}\, Y \,|\, \mathcal{G}] \quad \text{a.s.}$$

(i) The result follows from Lemma 6.15.

(j) Setting $V := X - Y$, by Jensen's inequality from Lemma 10.3(a) we have

$$\begin{aligned}
\big|\mathrm{ES}_G[V, L \,|\, \mathcal{G}]\big| &= \big|\mathbb{E}[f_{\mathcal{G},G,L}\, V \,|\, \mathcal{G}] + (1 - G(1-,\cdot))X^{\mathcal{G}}\big| \\
&\le \big|\mathbb{E}[f_{\mathcal{G},G,L}\, V \,|\, \mathcal{G}]\big| + \big|(1 - G(1-,\cdot))X^{\mathcal{G}}\big| \\
&\le \mathbb{E}[|f_{\mathcal{G},G,L}\, V| \,|\, \mathcal{G}] + (1 - G(1-,\cdot))|X|^{\mathcal{G}} = \mathrm{ES}_G[|V|, L \,|\, \mathcal{G}] \quad \text{a.s.}
\end{aligned}$$

This proves the lower inequality. The upper inequality follows from (b).

(k) Since the proof is longer, let us reduce the problem. We will first show the result for conditional expected shortfall contributions $\mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}]$ which then extends to weighted conditional expected shortfall contributions $\mathrm{ES}_G[X, L\,|\,\mathcal{G}]$. To see this, assume that we have shown the result for $\mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}]$ for all $t \in [0,1]$. Then, note that by Corollary 7.6, as well as by Fubini's theorem, cf. Kallenberg [71, Theorem 1.27],

$$\begin{aligned}
&\mathbb{E}\big[\big|\,\mathrm{ES}_G[X, L_n\,|\,\mathcal{G}] - \mathrm{ES}_G[X, L\,|\,\mathcal{G}]\,\big|\big] \\
&\quad\le \mathbb{E}\bigg[\int_{[0,1]} \big|\mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}] - \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L_n\,|\,\mathcal{G}]\big|\,dt\bigg] \\
&\quad= \int_{[0,1]} \mathbb{E}\big[\big|\mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}] - \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L_n\,|\,\mathcal{G}]\big|\big]\,dt\,, \quad n \ge n^*\,,
\end{aligned}$$

where n^* is chosen such that the integrand in the last expression is bounded from above by some constant. Note that we may find such an n^* due to Equation (7.13). Taking limits $n \to \infty$ and applying the dominated convergence theorem as, for example, given in Kallenberg [71, Theorem 1.21], gives

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\big[\big|\,\mathrm{ES}_G[X, L_n\,|\,\mathcal{G}] - \mathrm{ES}_G[X, L\,|\,\mathcal{G}]\,\big|\big] \\
&\qquad\le \int_{[0,1]} \lim_{n\to\infty} \mathbb{E}\big[\big|\mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L\,|\,\mathcal{G}] - \mathrm{ES}_{G^{\leftarrow}(t,\cdot)}[X, L_n\,|\,\mathcal{G}]\big|\big]\,dt = 0\,,
\end{aligned}$$

which shows that it is enough to prove the result for conditional expected shortfall contributions $\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}]$.

Secondly, let $t \in [0,1]$ be fixed and define $\delta := G^{\leftarrow}(t,\cdot)$. On $\{\delta = 1\}$ the result is trivial since the sequence is constant. On $\{\delta = 0\}$ the result follows by passing to the essential infimum once we have proven the result for $\{0 < \delta < 1\}$. Thus, we may assume $0 < \delta < 1$ a.s. Given $\varepsilon > 0$, there exists a constant z such that the bounded random variable $X_\varepsilon := X\,1_{\{|X|\le z\}}$ satisfies $\mathbb{E}[|X - X_\varepsilon|] = \mathbb{E}\big[|X|\,1_{\{|X|>z\}}\big] \le \varepsilon$, by dominated convergence. By subportfolio continuity (j) and by translation invariance (f), it therefore suffices to prove the result for bounded X.

Thirdly, without loss of generality we may assume that $\mathbb{E}[X\,1_{\{L=q_{\mathcal{G},\delta}(L)\}}\,|\,\mathcal{G}] = 0$ a.s. because in case $\mathbb{P}(L = q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) > 0$ with strictly positive probability we could, using translation invariance from (f), switch to $X' := X - a$ where,

$$a := \begin{cases} \frac{\mathbb{E}[X\,1_{\{L=q_{\mathcal{G},\delta}(L)\}}\,|\,\mathcal{G}]}{\mathbb{P}(L=q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G})} & \text{on } \{\mathbb{P}(L = q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) > 0\}\,, \\ 0 & \text{on } \{\mathbb{P}(L = q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) = 0\}\,. \end{cases}$$

This simplifies the adjusted indicator function $f_{\mathcal{G},\delta,L}$ as given by Definition 5.1. Due to linearity (g), we may restrict to those X which are bounded by $1 - \delta$.

Having reduced the problem, let $\varepsilon > 0$. By right-continuity of distribution functions, there exists an $\eta > 0$ such that

$$\mathbb{P}(0 < |L - q_{\mathcal{G},\delta}(L)| < 2\eta) \le \varepsilon\,. \tag{7.10}$$

Since $(L_n)_{n\in\mathbb{N}}$ converges to L in conditional probability, we have convergence in probability by taking expectations and by conditional bounded convergence as, for example, given in [65, Theorem 1.20]. Thus, there exists an $n_\varepsilon \in \mathbb{N}$ such that

$$\mathbb{P}(|L - L_n| \ge \eta) \le \varepsilon\,, \quad n \ge n_\epsilon\,, \tag{7.11}$$

and, by Lemma 3.18(l),

$$\mathbb{P}(q_{\mathcal{G},\delta}(L) - q_{\mathcal{G},\delta}(L_n) > \eta) \leq \varepsilon\,, \quad n \geq n_\epsilon\,. \tag{7.12}$$

Moreover, note that $|1_A - 1_B| = 1_{A\cap B^c} + 1_{A^c\cap B}$ for all $A, B \in \mathcal{F}$.

In the following, by considering three cases, we will show that

$$\mathbb{E}\big[\big|\,\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] - \mathrm{ES}_\delta[X, L_n\,|\,\mathcal{G}]\,\big|\big] \leq 8\varepsilon\,, \quad \text{a.s., for all } n \geq n_\varepsilon\,. \tag{7.13}$$

Since $\varepsilon > 0$ is arbitrary, the equation above implies the desired L^1-convergence result. From now on, let $\varepsilon > 0$ be fixed and let $n \geq n_\varepsilon$, as above, and define

$$M_{R_1,R_2(n)} := \{L\,R_1\,q_{\mathcal{G},\delta}(L), L_n\,R_2\,q_{\mathcal{G},\delta}(L_n)\}\,, \quad R_1, R_2 \in \{<, >, \leq, \geq\}\,. \tag{7.14}$$

Case I: Starting on the set $M := \{q_{\mathcal{G},\delta}(L_n) > q_{\mathcal{G},\delta}(L) + \eta\}$, we do not need the additional assumptions made in the lemma. First, note that in this case

$$\begin{aligned}
1_M(1 - \beta_{\mathcal{G},\delta,X})\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big] &= 1_M\big(\delta - \mathbb{P}(L_n < q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G})\big)\\
&\leq 1_M\big(\mathbb{P}(L < q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) - \mathbb{P}(L_n < q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G})\big)\\
&\leq 1_M\mathbb{P}(M_{\leq,\geq(n)}\,|\,\mathcal{G}) \quad \text{a.s.}
\end{aligned}$$

Taking expectations in the expression above and using (7.11) we get

$$\mathbb{E}\big[1_M(1 - \beta_{\mathcal{G},\delta,X})\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big]\big] \leq \mathbb{P}(M \cap M_{\leq,\geq(n)}) \leq \varepsilon\,. \tag{7.15}$$

By partitioning $\{L_n \geq q_{\mathcal{G},\delta}(L_n)\}$ we get

$$\begin{aligned}
1_M(1 - \delta) &\leq 1_M\,\mathbb{P}(L_n \geq q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G})\\
&= 1_M\big(\mathbb{P}(M_{>,\geq(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{\leq,\geq(n)}\,|\,\mathcal{G})\big) \quad \text{a.s.}
\end{aligned}$$

Hence, taking expectations on both sides gives $\mathbb{P}(M \cap M_{>,\geq(n)}) \geq \mathbb{E}[1_M\,(1 - \delta)] - \varepsilon$. Partitioning $\{L > q_{\mathcal{G},\delta}(L)\}$ yields

$$\begin{aligned}
1_M\,(1 - \delta) &\geq 1_M\,\mathbb{P}(L > q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G})\\
&= 1_M\big(\mathbb{P}(M_{>,\geq(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{>,<(n)}\,|\,\mathcal{G})\big) \quad \text{a.s.}
\end{aligned}$$

Thus, taking expectations on both sides again and using $\mathbb{P}(M \cap M_{>,\geq(n)}) \geq \mathbb{E}[1_M\,(1 - \delta)] - \varepsilon$ yield $\mathbb{P}(M \cap M_{>,<(n)}) \leq \varepsilon$. Finally, using the definition of conditional expected shortfall contributions, since $\mathbb{E}[X\,1_{\{L=q\}}\,|\,\mathcal{G}] = 0$ a.s. and by the boundedness of X by $1 - \delta$ we obtain

$$\begin{aligned}
&1_M\big|\,\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] - \mathrm{ES}_\delta[X, L_n\,|\,\mathcal{G}]\,\big|\\
&\quad\leq 1_M\big((1 - \beta_{\mathcal{G},\delta,X})\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big] + \mathbb{E}\big[\big|1_{\{L_n \geq q_{\mathcal{G},\delta}(L_n)\}} - 1_{\{L > q_{\mathcal{G},\delta}(L)\}}\big|\,\big|\,\mathcal{G}\big]\big)\\
&\quad= 1_M\big((1 - \beta_{\mathcal{G},\delta,X})\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big] + \mathbb{P}(M_{\leq,\geq(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{>,<(n)}\,|\,\mathcal{G})\big) \quad \text{a.s.}
\end{aligned}$$

Thus, by (7.15) and since $\mathbb{P}(M \cap M_{>,<(n)}) \leq \varepsilon$,

$$\mathbb{E}\big[1_M\big|\,\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] - \mathrm{ES}_\delta[X, L_n\,|\,\mathcal{G}]\,\big|\big] \leq 3\varepsilon$$

which ends the first case.

Case II: Secondly, consider the case on $M^{\mathrm{c}} = \{q_{\mathcal{G},\delta}(L_n) \leq q_{\mathcal{G},\delta}(L_n) + \eta\}$ for the two different assumptions given in (k). Note that, by using (7.10) and (7.11), as well as by partitioning $M_{>,\leq(n)}$ we get

$$\begin{aligned}\mathbb{P}(M^{\mathrm{c}} \cap M_{>,\leq(n)}) = &\mathbb{P}\big(M^{\mathrm{c}} \cap \{q_{\mathcal{G},\delta}(L) < L < q_{\mathcal{G},\delta}(L) + 2\eta\} \cap \{L_n \leq q_{\mathcal{G},\delta}(L_n)\}\big) \\ &+ \mathbb{P}\big(M^{\mathrm{c}} \cap \{L \geq q_{\mathcal{G},\delta}(L) + 2\eta\} \cap \{L_n \leq q_{\mathcal{G},\delta}(L_n)\}\big) \leq 2\varepsilon\,.\end{aligned} \tag{7.16}$$

Case II(a): Let the assumption $\mathbb{P}(L \leq q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) = \delta$ a.s. be satisfied. By partitioning $\{L_n \leq q_{\mathcal{G},\delta}(L_n)\}$ we obtain

$$\begin{aligned}1_{M^{\mathrm{c}}}\,\delta &\leq 1_{M^{\mathrm{c}}}\,\mathbb{P}(L_n \leq q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G}) \\ &= 1_{M^{\mathrm{c}}}\,\big(\mathbb{P}(M_{\leq,\leq(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{>,\leq(n)}\,|\,\mathcal{G})\big) \quad \text{a.s.}\end{aligned}$$

Taking expectations on both sides and (7.16) yield $\mathbb{P}(M^{\mathrm{c}} \cap M_{\leq,\leq(n)}) \geq \mathbb{E}[1_{M^{\mathrm{c}}}\,\delta] - 2\varepsilon$. Moreover, partitioning $\{L \leq q_{\mathcal{G},\delta}(L)\}$ yields

$$1_{M^{\mathrm{c}}}\,\delta \geq 1_{M^{\mathrm{c}}}\,\mathbb{P}(L \leq q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) = 1_{M^{\mathrm{c}}}\,\big(\mathbb{P}(M_{\leq,>(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{\leq,\leq(n)}\,|\,\mathcal{G})\big) \quad \text{a.s.}$$

Thus, taking expectations on both sides and using $\mathbb{P}(M^{\mathrm{c}} \cap M_{\leq,\leq(n)}) \geq \mathbb{E}[1_{M^{\mathrm{c}}}\,\delta] - 2\varepsilon$ give $\mathbb{P}(M^{\mathrm{c}} \cap M_{\leq,\leq(n)}) \leq \mathbb{E}[1_{M^{\mathrm{c}}}\,\delta]$, as well as $\mathbb{P}(M^{\mathrm{c}} \cap M_{\leq,>(n)}) \leq 2\varepsilon$. Furthermore, we get

$$\begin{aligned}1_{M^{\mathrm{c}}}\,\beta_{\mathcal{G},\delta,X}\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big] &= 1_{M^{\mathrm{c}}}\big(\mathbb{P}(L_n \leq q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G}) - \delta\big) \\ &= 1_{M^{\mathrm{c}}}\big(\mathbb{P}(M_{>,\leq(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{\leq,\leq(n)}\,|\,\mathcal{G}) - \delta\big) \quad \text{a.s.,}\end{aligned} \tag{7.17}$$

and hence $\mathbb{E}\big[1_M\,\beta_{\mathcal{G},\delta,X}\,\mathbb{E}[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,|\,\mathcal{G}]\big] \leq 2\varepsilon$ by taking expectations on both sides and using (7.16). Finally, using the definition of conditional expected shortfall contributions, since $\mathbb{E}[X\,1_{\{L=q\}}\,|\,\mathcal{G}] = 0$ a.s. and by the boundedness of X by $1-\delta$ it follows

$$\begin{aligned}&1_{M^{\mathrm{c}}}\big|\,\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] - \mathrm{ES}_\delta[X, L_n\,|\,\mathcal{G}]\,\big| \\ &\quad\leq 1_{M^{\mathrm{c}}}\,\big(\beta_{\mathcal{G},\delta,X}\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big] + \mathbb{E}\big[\big|1_{\{L_n > q_{\mathcal{G},\delta}(L_n)\}} - 1_{\{L > q_{\mathcal{G},\delta}(L)\}}\big|\,\big|\,\mathcal{G}\big]\big) \\ &\quad= 1_{M^{\mathrm{c}}}\,\big(\beta_{\mathcal{G},\delta,X}\,\mathbb{E}\big[1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big] + \mathbb{P}(M_{>,\leq(n)}\,|\,\mathcal{G}) + \mathbb{P}(M_{\leq,>(n)}\,|\,\mathcal{G}) \quad \text{a.s.}\end{aligned}$$

Thus, by (7.16) and (7.17), as well as since $\mathbb{P}(M^{\mathrm{c}} \cap M_{\leq,>(n)}) \leq 2\varepsilon$, we get

$$\mathbb{E}\big[1_{M^{\mathrm{c}}}\big|\,\mathrm{ES}_\delta[X, L\,|\,\mathcal{G}] - \mathrm{ES}_\delta[X, L_n\,|\,\mathcal{G}]\,\big|\big] \leq 6\varepsilon\,.$$

This implies (7.13) and thus ends the proof for the case $\mathbb{P}(L \leq q_{\mathcal{G},\delta}(L)\,|\,\mathcal{G}) = \delta$ a.s.

Case II(b): For this case, let X be a.s. constant on the set $\{L = q_{\mathcal{G},\delta}(L)\}$. Therefore, $\mathbb{E}[X\,1_{\{L=q_{\mathcal{G},\delta}(L)\}}\,|\,\mathcal{G}] = 0$ a.s. implies $\mathbb{E}[|X|\,1_{\{L=q_{\mathcal{G},\delta}(L),\,L_n=q_{\mathcal{G},\delta}(L_n)\}}\,|\,\mathcal{G}] = 0$ a.s., as well as $\mathbb{E}[|X|\,1_{\{L=q_{\mathcal{G},\delta}(L),\,L_n>q_{\mathcal{G},\delta}(L_n)\}}\,|\,\mathcal{G}] = 0$ a.s. Thus, by the boundedness of X,

$$\begin{aligned}\frac{\mathbb{E}\big[|X|\,1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big]}{1-\delta} &= \frac{\mathbb{E}\big[|X|\,1_{\{L \neq q_{\mathcal{G},\delta}(L),\,L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big]}{1-\delta} \\ &\leq \mathbb{P}\big(L \neq q_{\mathcal{G},\delta}(L), L_n = q_{\mathcal{G},\delta}(L_n)\,\big|\,\mathcal{G}\big) \\ &\leq \mathbb{P}\big(0 < |L - q_{\mathcal{G},\delta}(L)| < 2\eta\,\big|\,\mathcal{G}\big) + \mathbb{P}\big(|L - q_{\mathcal{G},\delta}(L)| \geq 2\eta, L_n = q_{\mathcal{G},\delta}(L_n)\,\big|\,\mathcal{G}\big) \quad \text{a.s.}\end{aligned}$$

Taking expectations on both sides and using (7.10), (7.11), as well as (7.12) gives

$$\mathbb{E}\left[\frac{\mathbb{E}\big[|X|\,1_{\{L_n = q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big]}{1-\delta}\right] \leq 3\varepsilon. \tag{7.18}$$

Moreover, we get

$$\begin{aligned}\frac{\mathbb{E}\big[|X|1_{M_{\leq,>(n)}}\,\big|\,\mathcal{G}\big]}{1-\delta} &\leq \mathbb{P}(L \leq q_{\mathcal{G},\delta}(L), L_n > q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G}) \\ &= \mathbb{P}(q_{\mathcal{G},\delta}(L) - 2\eta < L \leq q_{\mathcal{G},\delta}(L), L_n > q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G}) \\ &\quad + \mathbb{P}(L \leq q_{\mathcal{G},\delta}(L) - 2\eta, L_n > q_{\mathcal{G},\delta}(L_n)\,|\,\mathcal{G}) \quad \text{a.s.}\end{aligned}$$

Again, taking expectations on both sides and using (7.10), (7.11), as well as (7.12) yields

$$\mathbb{E}\left[\frac{\mathbb{E}\big[|X|1_{M_{\leq,>(n)}}\,\big|\,\mathcal{G}\big]}{1-\delta}\right] \leq 3\varepsilon\,. \tag{7.19}$$

Finally, using the definition of conditional expected shortfall contributions together with $0 \leq \beta_{\mathcal{G},\delta,X} \leq 1$ a.s. and $\mathbb{E}[X1_{\{L=q\}}\,|\,\mathcal{G}] = 0$ a.s., we obtain

$$\begin{aligned}&1_{M^c}\big|\operatorname{ES}_\delta[X,L\,|\,\mathcal{G}] - \operatorname{ES}_\delta[X,L_n\,|\,\mathcal{G}]\,\big| \\ &\quad \leq 1_{M^c}\frac{\mathbb{E}\big[|X|1_{\{L_n=q_{\mathcal{G},\delta}(L_n)\}}\,\big|\,\mathcal{G}\big]}{1-\delta} + \frac{\mathbb{E}\big[|X|1_{M_{>,\leq(n)}}\,\big|\,\mathcal{G}\big]}{1-\delta} + \frac{\mathbb{E}\big[|X|1_{M_{\leq,>(n)}}\,\big|\,\mathcal{G}\big]}{1-\delta} \quad \text{a.s.}\end{aligned}$$

Thus, by the boundedness of X and by (7.16), (7.18), as well as (7.19), we get

$$\mathbb{E}\big[1_{M^c}\big|\operatorname{ES}_\delta[X,L\,|\,\mathcal{G}] - \operatorname{ES}_\delta[X,L_n\,|\,\mathcal{G}]\,\big|\big] \leq 8\varepsilon\,.$$

This implies (7.13) and thus ends the proof of the second case.

(l) Let $n \in \mathbb{N}$ and start on the set $M := \{\varepsilon_n > 0\}$. By consistency (a), diversification (b) and linearity (g), we get

$$\begin{aligned}1_M \operatorname{ES}_G[L+\varepsilon_n X\,|\,\mathcal{G}] &\geq 1_M \operatorname{ES}_G[L+\varepsilon_n X, L\,|\,\mathcal{G}] \\ &= 1_M\big(\operatorname{ES}_G[L\,|\,\mathcal{G}] + \varepsilon_n \operatorname{ES}_G[X,L\,|\,\mathcal{G}]\big) \quad \text{a.s.},\end{aligned}$$

and hence

$$1_M \frac{\operatorname{ES}_G[L+\varepsilon_n X\,|\,\mathcal{G}] - \operatorname{ES}_G[L\,|\,\mathcal{G}]}{\varepsilon_n} \geq 1_M \operatorname{ES}_G[X,L\,|\,\mathcal{G}] \quad \text{a.s.} \tag{7.20}$$

By Definition 7.2 and consistency (a), we have

$$\begin{aligned}1_M \operatorname{ES}_G[L, L+\varepsilon_n X\,|\,\mathcal{G}] &= 1_M\big(\mathbb{E}[f_{\mathcal{G},G,L+\varepsilon_n X}\,L\,|\,\mathcal{G}] + (1-G(1-,\cdot))\,L^{\mathcal{G}}\big) \\ &= 1_M\big(\mathbb{E}[f_{\mathcal{G},G,L+\varepsilon_n X}\,(L+\varepsilon_n)\,|\,\mathcal{G}] - \varepsilon_n\,\mathbb{E}[f_{\mathcal{G},G,L+\varepsilon_n X}\,X\,|\,\mathcal{G}]\big) \\ &= 1_M\big(\operatorname{ES}_G[L+\varepsilon_n X\,|\,\mathcal{G}] - \varepsilon_n \operatorname{ES}_G[X, L+\varepsilon_n X\,|\,\mathcal{G}]\big) \quad \text{a.s.},\end{aligned}$$

where, by diversification (b),

$$1_M \operatorname{ES}_G[L, L+\varepsilon_n X\,|\,\mathcal{G}] \leq 1_M \operatorname{ES}_G[L\,|\,\mathcal{G}] \quad \text{a.s.}$$

These last two observations imply

$$1_M \operatorname{ES}_G[X, L+\varepsilon_n X\,|\,\mathcal{G}] \geq 1_M \frac{\operatorname{ES}_G[L+\varepsilon_n X\,|\,\mathcal{G}] - \operatorname{ES}_G[L\,|\,\mathcal{G}]}{\varepsilon_n} \quad \text{a.s.} \tag{7.21}$$

To get the corresponding results on $\{\varepsilon_n < 0\}$, simply apply (7.20) and (7.21) to $\varepsilon_n' = -\varepsilon_n$, as well as $X' = -X$ and use the fact that $\operatorname{ES}_G[X,L\,|\,\mathcal{G}] = -\operatorname{ES}_G[X',L\,|\,\mathcal{G}]$ a.s., by linearity as given in (g).

Note that $(L+\varepsilon_n X)$ converges to L almost surely and thus in conditional probability as $n \to \infty$. Hence, using (7.20) and (7.21), we can apply (k) to obtain

$$\operatorname{ES}_G[X,L\,|\,\mathcal{G}] = \lim_{n\to\infty} \frac{\operatorname{ES}_G[L+\varepsilon_n X\,|\,\mathcal{G}] - \operatorname{ES}_G[L\,|\,\mathcal{G}]}{\varepsilon_n} \quad \text{in } L^1\,.$$

□

Chapter 8

Illustrative Applications of Conditional Risk Measures

In this section we give a few illustrative examples where different conditional risk measures are calculated explicitly. We start with time series applications in Examples 8.1 and 8.2 which are based on a study of McNeil and Frey [84]. Moreover, Example 8.3 shows some basic calculations and proves that conditional expected shortfall in general is neither time-consistent nor, equivalently, recursive, cf. Delbaen [31, Section 6] and see Remark 5.25(e). Finally, Example 8.5 shows another application of conditional risk measures with a scope to scenario analysis where conditioning is subject to risk factors in the credit risk model extended CreditRisk$^+$, see Schmock [111, Section 6]. This example is closely related to Example 2.5 and to the annuity model discussed in the second part of this thesis.

Example 8.1. The first example is taken from McNeil and Frey [84] where negative log returns of financial asset prices are modelled in order to calculate conditional lower quantiles and conditional expected shortfall. Let $(X_t)_{t\in\mathbb{Z}}$ be a family of $\mathcal{F}$-measurable real-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Dynamics are given by

$$X_t = \mu_t + \sigma_t Z_t\,, \quad t \in \mathbb{Z}\,,$$

where $(Z_t)_{t\in\mathbb{Z}}$ is an independent and identically distributed sequence of random variables with mean zero, unit variance and marginal distribution function F. For $t \in \mathbb{Z}$ define the filtration $\mathcal{F}_t := \sigma(X_s, s \le t)$ and assume that μ_t and σ_t are measurable with respect to $\mathcal{F}_{t-1}$ where, in addition, $\sigma_t > 0$ a.s. for every $t \in \mathbb{Z}$.

Due to the assumption of independence of $(Z_t)_{t\in\mathbb{Z}}$, we can use Lemma 10.3(b) to conclude that, for every $t \in \mathbb{Z}$ and every $\mathcal{F}_{t-1}$-measurable $Z\colon \Omega \to \mathbb{R}$,

$$\mathbb{P}(X_t \le Z \,|\, \mathcal{F}_{t-1}) = \mathbb{P}\Big(Z_t \le \frac{Z-\mu_t}{\sigma_t}\,\Big|\,\mathcal{F}_{t-1}\Big) = F\Big(\frac{Z-\mu_t}{\sigma_t}\Big) \quad \text{a.s.}$$

Thus, conditional lower quantiles at $\mathcal{F}_{t-1}$-measurable level $\delta_t\colon \Omega \to [0,1]$, see Definition 3.7, are given by

$$q_{\delta_t,\mathcal{F}_{t-1}}(X_t) = \mu_t + \sigma_t F^{\leftarrow}(\delta_t) \quad \text{a.s., for all } t \in \mathbb{Z}\,,$$

where $F^{\leftarrow}$ denotes the lower quantile function of F, see Remarks 3.11(a).

Next, we want to calculate conditional expected shortfall at $\mathcal{F}_{t-1}$-measurable level $\delta_t\colon \Omega \to [0,1]$ using Definitions 5.1 and 5.3. Note that $f_{\mathcal{F}_{t-1},\delta_t,X_t} = f_{\mathcal{F}_{t-1},\delta_t,Z_t}$ a.s. for every $t \in \mathbb{Z}$. Thus, by Remarks 5.2(b) as well as Remarks 3.3(h) and (i),

$$\mathrm{ES}_{\delta_t}[X_t\,|\,\mathcal{F}_{t-1}] = \mu_t + \sigma_t\,\mathrm{ES}_{\delta}[Z_t\,|\,\mathcal{F}_{t-1}] \quad \text{a.s.},$$

which, by Lemma 5.12, can also be written as

$$\mathrm{ES}_{\delta_t}[X_t\,|\,\mathcal{F}_{t-1}] = \mu_t + \frac{\sigma_t}{1-\delta}\int_{[\delta,1)} F^{\leftarrow}(u)\,du \quad \text{a.s., on } \{0<\delta_t<1\}\,.$$

The conditional distortion risk measure $\rho_g[X_t\,|\,\mathcal{F}_{t-1}]$ with distortion process g adapted to the constant filtration $(\mathcal{F}_{t-1})_{s\in[0,1]}$, see Definitions 4.1 and 4.5, is then given by

$$\rho_g[X_t\,|\,\mathcal{F}_{t-1}] = \mu_t + \sigma_t\int_{[0,1]} F^{\leftarrow}(1-u)\,g(du,\cdot) \quad \text{a.s.}$$

For GARCH-type models McNeil and Frey [84] provide an estimation procedure for conditional lower quantiles and conditional expected shortfall where the tail of F is modelled via extreme value distributions.

Example 8.2. Considering a similar setting as before, we want to show an easy application of conditional expected shortfall contributions as defined in Chapter 7. Consider the same assumptions as in Example 8.1 where, in addition, the stochastic process $(X_t)_{t\in\mathbb{Z}}$ is a sum of two processes, i.e.,

$$X_t = X_{t,1} + X_{t,2}\,, \quad t\in\mathbb{Z}\,,$$

where $X_{t,i} := \mu_{t,i} + \sigma_{t,i}Z_t$ with $\mu_{t,i}$ and $\sigma_{t,i}$ being $\mathcal{F}_{t-1}$-measurable for $i=1,2$ and $t\in\mathbb{Z}$ such that $\sigma_{t,1} > \sigma_{t,2} > 0$ a.s. Thus, we are given a process where its components $X_{t,1}$ and $X_{t,2}$ are conditionally comonotonic with respect to $\mathcal{F}_{t-1}$ for all $t\in\mathbb{Z}$. Using the results of the previous example, we get, for all $t\in\mathbb{Z}$,

$$\mathrm{ES}_{\delta_t}[X_t\,|\,\mathcal{F}_{t-1}] = (\mu_{t,1}+\mu_{t,2}) + (\sigma_{t,1}+\sigma_{t,2})\,\mathrm{ES}_{\delta_t}[Z_t\,|\,\mathcal{F}_{t-1}] \quad \text{a.s.}$$

Since $f_{\mathcal{F}_{t-1},\delta_t,X_t} = f_{\mathcal{F}_{t-1},\delta_t,Z_t}$ a.s., conditional expected shortfall contributions of $X_{t,i}$ to X_t at $\mathcal{F}_{t-1}$-measurable level $\delta_t\colon\Omega\to[0,1]$ are, for all $t\in\mathbb{Z}$, given by

$$\mathrm{ES}_{\delta_t}[X_{t,i},X_t\,|\,\mathcal{F}_{t-1}] = \mu_{t,i} + \sigma_{t,i}\,\mathrm{ES}_{\delta_t}[Z_t\,|\,\mathcal{F}_{t-1}] \quad \text{a.s., for } i=1,2\,.$$

Alternatively, assume that $X_{t,2} := \mu_{t,2} + \sigma_{t,2}(-Z_t)$ for all $t\in\mathbb{Z}$, i.e., the components $X_{t,1}$ and $-X_{t,2}$ are conditionally comonotonic with respect to $\mathcal{F}_{t-1}$ for all $t\in\mathbb{Z}$. Just as before, for all $t\in\mathbb{Z}$, we get

$$\mathrm{ES}_{\delta_t}[X_t\,|\,\mathcal{F}_{t-1}] = (\mu_{t,1}+\mu_{t,2}) + (\sigma_{t,1}-\sigma_{t,2})\,\mathrm{ES}_{\delta_t}[Z_t\,|\,\mathcal{F}_{t-1}] \quad \text{a.s.}$$

and

$$\mathrm{ES}_{\delta_t}[X_{t,1},X_t\,|\,\mathcal{F}_{t-1}] = \mu_{t,1} + \sigma_{t,1}\,\mathrm{ES}_{\delta_t}[Z_t\,|\,\mathcal{F}_{t-1}] \quad \text{a.s.}$$

This example illustrates the sensitivity of conditional expected shortfall contributions to the underlying dependence structure. In particular, conditional expected shortfall contributions of a subportfolio can be greater, as well as smaller than conditional expected shortfall of the whole portfolio itself for different dependence structures, e.g., we may set $\mu_{t,1} = \mu_{t,2} = 0$ for all $t\in\mathbb{Z}$.

Example 8.3. For a continuous or discrete time interval $I \subset [0,\infty)$ let $X = (X_t)_{t\in I}$ be a stochastic process such that $\mathbb{E}[|X_t|] < \infty$ for all $t \in I$ with independent increments and let $(\mathcal{F}_t)_{t\in I}$ denote the natural filtration induced by X. Thus, for all $t, T \in I$ with $t < T$, $(X_T - X_t)$ is independent of $\mathcal{F}_t$. Moreover, for all $t \in I$, assume level $\delta_t\colon \Omega \to (0,1)$ to be $\mathcal{F}_t$-measurable.

We start with a derivation of conditional lower quantiles, see Definition 3.7, and of conditional expected shortfall, see Definition 5.3. By translation invariance and a property under independence of conditional lower quantiles, see Lemma 3.18(c) and (f), we immediately get

$$q_{\mathcal{F}_t,\delta_t}(X_T) = X_t + q_{\mathcal{F}_t,\delta_t}(X_T - X_t) = X_t + F^{\leftarrow}(\delta_t) \quad \text{a.s.},$$

for all $t, T \in I$ with $t < T$. Using the quantile representation of conditional expected shortfall of Lemma 5.12, we get

$$\mathrm{ES}_{\delta_t}[X_T\,|\,\mathcal{F}_t] = \frac{1}{1-\delta_t}\int_{[\delta_t,1)} X_t + F^{\leftarrow}(u)\,du = X_t + \mathrm{ES}_{\delta_t}[X_T - X_t] \quad \text{a.s.}, \tag{8.4}$$

where $\mathrm{ES}_{\delta_t}[X_T - X_t]$ denotes classical, unconditional expected shortfall of $(X_T - X_t)$, see Remarks 5.5(a). From now on, let $s, t, T \in I$ with $s < t \le T$, as well as $\mathcal{F}_s$-measurable level $\delta_s\colon \Omega \to (0,1)$ be fixed. Using Equation (8.4) twice, together with the supermartingale property of Corollary 5.27, we get

$$\mathrm{ES}_{\delta_s}[X_T\,|\,\mathcal{F}_s] \ge \mathbb{E}[\mathrm{ES}_{\delta_s}[X_T\,|\,\mathcal{F}_t]\,|\,\mathcal{F}_s] = \mathbb{E}[X_t\,|\,\mathcal{F}_s] + \mathrm{ES}_{\delta_s}[X_T - X_t] \quad \text{a.s.},$$

which implies

$$\mathrm{ES}_{\delta_s}[X_T - X_s] - \mathbb{E}[X_T - X_s] \ge \mathrm{ES}_{\delta_s}[X_T - X_t] - \mathbb{E}[X_T - X_t]\ .$$

Note that this inequality is related to the uncertainty decrease of projections as given in Lemma 5.23(o). For a standard Brownian motion, the inequality above is immediate since both expectations equal zero and since in that case expected shortfall is given by, cf. McNeil, Frey and Embrechts [85, Example 2.18],

$$\mathrm{ES}_{\delta'}[X_T - X_t] = \sqrt{T-t}\,\frac{\phi(\Phi^{-1}(\delta'))}{1-\delta'}\,, \quad \delta' \in (0,1)\,,$$

where ϕ denotes the continuous density, as well as where Φ^{-1} denotes the inverse distribution function of a standard normal distribution. Moreover, Equation (8.4) and subadditivity of expected shortfall, cf. Lemma 5.23(d), imply

$$\mathrm{ES}_{\delta_s}[\mathrm{ES}_{\delta_s}[X_T\,|\,\mathcal{F}_t]\,|\,\mathcal{F}_s] = X_s + \mathrm{ES}_{\delta_s}[X_T - X_t] + \mathrm{ES}_{\delta_s}[X_t - X_s] \ge \mathrm{ES}_{\delta_s}[X_T\,|\,\mathcal{F}_s] \quad \text{a.s.}$$

Note that in the case of a standard Brownian motion the inequality is strict if $s < t < T$. Thus, conditional expected shortfall is not recursive, in general, and hence not time-consistent. For the equivalence of recursiveness and time-consistency see Delbaen [31, Section 6].

Conditional distortion risk measures are straight-forward to derive. Moreover, using the result of (8.4), conditional G-weighted expected shortfall of X_T, with $t, T \in I$ and $t < T$, as well as $(\mathcal{F}_t)_{u\in I}$-adapted process G as given in Definition 6.2, is given by

$$\mathrm{ES}_G[X\,|\,\mathcal{F}_t] = \int_{[0,1]} X_t + \mathrm{ES}_r[X_T - X_t]\ G(dr,\cdot) = X_t + \mathrm{ES}_G[X_T - X_t] \quad \text{a.s.},$$

where $\mathrm{ES}_G[X_T - X_t]$ is given as in Lemma 6.5(h). Hence, we see that in an independent-increment-setting all conditional risk measures described in this thesis can be simplified to an unconditional representation.

Example 8.5. Let us consider a simplified version of the collective risk model extended CreditRisk$^+$ which is related to Example 2.5 and to the annuity model discussed in the second part of this thesis. Also see Schmock [111, Section 6] for further information and a detailed introduction to this model. Modifications of this model are given in Section 11.2 and Section 19.1.1. We assume $m \in \mathbb{N}$ obligors and a collection G of non-empty subsets of all obligors $\{1, \dots, m\}$, called risk groups, which are subject to joint defaults. Each single obligor at least belongs to one risk group. We are given $K \in \mathbb{N}$ non-idiosyncratic risk factors $\Lambda := (\Lambda_1, \dots, \Lambda_K)$ which are the joint drivers of the number of occurring defaults. Their support is assumed to be countable and strictly positive with $\mathrm{Var}(\Lambda_k) > 0$ and $\mathbb{E}[\Lambda_k] = 1$ for $k \in \{1, \dots, K\}$. Moreover, we are given one-year default probabilities $p_g \in [0, 1]$ for all risk groups $g \in G$. Weight, or susceptibility, of idiosyncratic risk is denoted by $w_{g,0} \in [0, 1]$ and weights of non-idiosyncratic risks are denoted by $w_{g,k} \in [0, 1]$ for $k \in \{1, \dots, K\}$ such that

$$\sum_{k=0}^{K} w_{g,k} = 1 \,.$$

For each group $g \in G$, the default number due to idiosyncratic risk, denoted by $N_{g,0}$, is independent of every other random variable and has a Poisson distribution

$$\mathcal{L}(N_{g,0}) = \mathrm{Poisson}(p_g w_{g,0}) \,.$$

For each group $g \in G$, the default numbers due to non-idiosyncratic risk factors Λ, denoted by $N_{g,1}, \dots, N_{g,K}$, are conditionally Poisson-distributed

$$\mathcal{L}(N_{g,k} \,|\, \Lambda) = \mathrm{Poisson}(p_g w_{g,k} \Lambda_k) \,.$$

Moreover, assume that conditionally on Λ non-idiosyncratic defaults

$$\big\{ N_{g,k} \,\big|\, g \in G, k \in \{1, \dots, K\} \big\}$$

are independent. The sequence of $\mathbb{N}_0^g$-valued random losses of obligor $i \in \{1, \dots, m\}$ in risk group $g \in G$ due to risk $k \in \{0, \dots, K\}$ for default number $n \in \mathbb{N}$, denoted by $(L_{g,i,k,n})_{i \in g}$, is independent and identically distributed and independent of all other random variables. In particular, for all $g \in G$, $i \in \{1, \dots, m\}$, $k \in \{0, \dots, K\}$ and $n \in \mathbb{N}$, we assume $L_{g,i,k,n}$ to be Bernoulli distributed with

$$p := \mathbb{P}(L_{g,i,k,n} = 1) = 1 - \mathbb{P}(L_{g,i,k,n} = 0) \,.$$

Moreover, define $G_i := \{g \in G \,|\, i \in g\}$ for each obligor $i \in \{1, \dots, m\}$. Then, the total loss in the extended CreditRisk$^+$ framework is given by

$$L := \sum_{i=1}^{m} \sum_{g \in G_i} \sum_{k=0}^{K} \sum_{n=1}^{N_{g,k}} L_{g,i,k,n} \,.$$

Due to our assumptions, see Schmock [111, Section 6.9.2], L has a conditional Poisson distribution

$$\mathcal{L}(L \,|\, \Lambda) = \mathrm{Poisson}(\lambda(\Lambda))$$

where the random parameter $\lambda(\Lambda)$ is given by

$$\lambda(q) := p \sum_{i=1}^{m} \sum_{g \in G_i} p_g \left(w_{g,0} + \sum_{k=1}^{K} w_{g,k} q_k \right), \quad q = (q_1, \dots, q_K) \in (0, \infty)^K \,.$$

In the following, we want to calculate conditional lower quantiles and conditional expected shortfall for different levels δ, conditioned on Λ. Note that $\sigma(\Lambda)$-measurable random variables can be written as measurable functions of Λ. Thus, since L has a conditional Poisson distribution,

$$\mathbb{P}(L \leq f(\Lambda) \,|\, \Lambda) = \sum_{q \in \operatorname{supp} \Lambda} 1_{\{\Lambda = q\}} \mathbb{P}(L \leq f(q) \,|\, \Lambda) = e^{-\lambda(\Lambda)} \sum_{j=0}^{\lfloor f(\Lambda) \rfloor} \frac{(\lambda(\Lambda))^j}{j!} \quad \text{a.s.},$$

which equals the distribution function of a Poisson distribution with parameter $\lambda(\Lambda)$ evaluated at $f(\Lambda)$. For simplicity, we assume a deterministic level $\delta \in (0,1)$. The task, when calculating conditional lower quantiles as given in Definition 3.7, is to find minimal—in the sense of an essential infimum—$f(\Lambda)$ such that the expression above is greater or equal than δ. Thus, obviously,

$$q_{\delta,\sigma(\Lambda)}(L) = F_\Lambda^{\leftarrow}(\delta)\,, \quad \delta \in (0,1)\,, \tag{8.6}$$

where $F_q^{\leftarrow}$ for $q > 0$ denotes the lower quantile function of a Poisson distribution with parameter $\lambda(q)$, see Remarks 3.11(a). Explicitly, we have

$$F_q^{\leftarrow}(\delta) = \min\left\{ k \in \mathbb{N}_0 \,\middle|\, e^{-\lambda(q)} \sum_{j=0}^{k} \frac{(\lambda(q))^j}{j!} \geq \delta \right\}, \quad \delta \in (0,1) \text{ and } q \in (0,\infty)^K\,.$$

Using the quantile representation of conditional expected shortfall of Lemma 5.12, we finally get a representation for conditional expect shortfall as follows:

$$\mathrm{ES}_\delta[L \,|\, \sigma(\Lambda)] = \int_{[\delta,1)} F_\Lambda^{\leftarrow}(t)\, dt \quad \text{a.s.} \tag{8.7}$$

With respect to scenario analysis, (8.6) and (8.7) show how a change in the underlying risk factors transforms risk in a credit or a annuity portfolio using CreditRisk$^+$. In Chapter 12 we provide estimation procedures for a simplified version of extended CreditRisk$^+$ in the context of life insurance and annuity portfolios. Using this approach, realisations of risk factors Λ under various scenarios can be estimated, see Chapter 14. Just to mention a few, such scenarios can correspond to macro economic events, to an overall shift in credit ratings after a financial crisis, or to the introduction of new effective treatments in the context of longevity risk in life insurance.

Chapter 9

Conclusion to Advanced Conditional Risk Measurement

Conclusively, in the first part of this thesis we give a mathematically sound and very general approach to several classes of well-known and also new conditional risk measures which provide the basis for an assessment of acceptable risk in a dynamic environment. We give and collect rigorous, explicit definitions for sometimes intuitively used conditional versions of notions like essential supremum and quantile, distortion risk measure, expected shortfall, weighted expected shortfall and risk contributions. Various notable properties for all these conditional risk measures and dynamic counterparts are transferred or proven so that we provide a useful toolbox for practitioners, as well as researchers. As an alternative for time-consistency we suggest two other dynamic properties which are satisfied by conditional expected shortfall, as well as weighted conditional expected shortfall. We also introduce risk contributions to weighted conditional expected shortfall quantifying the risk of a subportfolio to the whole portfolio and prove notable properties. Ultimately, we observe that most properties which hold in the classical, unconditional risk measures can be proven in the conditional case as well. Thus, in some cases, conditional risk measures may seem as a trivial extensions to classical, unconditional cases. But, in particular, if we use random levels of risk aversion and if we condition on general sub-σ-algebras, as in the case of stochastic processes, definitions of conditional risk measures have to be chosen very carefully so that all quantities are well-defined and so that classical properties are preserved.

Chapter 10

Appendix to Advanced Conditional Risk Measurement

In this chapter we give some essential definitions and basic results which are used throughout this thesis. We start with the definition of σ-integrable random variables and then introduce the notion of time-consistency as given, for example, in Delbaen [31, Section 6], as well as its weaker forms. Conditional entropic risk measures are mentioned as an important class of time-consistent risk measures. Moreover, we define conditional versions of comonotonicity and stochastic ordering. We also give some preliminary technical results including a Fubini-type theorem for conditional expectations. Moreover, in Section 10.2 we provide alternative proofs to some items in Lemma 5.23 where the usage of results taken from conditional distortion risk measures is avoided.

10.1 Basic concepts and definitions

Definition 10.1 (σ-integrability)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$. Then, X is called *σ-integrable with respect to $\mathcal{G}$* if there exists a sequence $(\Omega_n)_{n\in\mathbb{N}} \subset \mathcal{G}$, where $\Omega_n \subset \Omega_{n+1}$ for all $n \in \mathbb{N}$, and $\mathbb{P}(\Omega \setminus \bigcup_{n\in\mathbb{N}} \Omega_n) = 0$ such that $\mathbb{E}[|X 1_{\Omega_n}|] < \infty$ for all $n \in \mathbb{N}$.

Remark 10.2. The definition of σ-integrability with respect to $\mathcal{G}$ is taken from He, Wang and Yan [65, Definition 1.15]. Similar as in the classical case, we can define conditional expectations for quasi-σ-integrable random variables, i.e., for random variables where just the negative or just the positive part is σ-integrable. Various desired properties still hold.

Lemma 10.3 (Two properties of conditional expectation based on σ-integrability)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, consider an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$. Then, we get the following conditional properties:*

(a) Conditional Jensen: Assume that X is σ-integrable with respect to $\mathcal{G}$ and let $c\colon \mathbb{R} \to \mathbb{R}$ be a convex function. Then,

$$\mathbb{E}[c(X)\,|\,\mathcal{G}] \geq c(\mathbb{E}[X\,|\,\mathcal{G}]) \quad a.s.$$

(b) Assume that X^- is σ-integrable with respect to $\mathcal{G}$ and consider $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R}$. Let $h\colon \mathbb{R}^2 \to \mathbb{R}$ be a measurable function such that $h(X, Z)^-$ is σ-integrable with respect

to $\mathcal{G}$ and assume that X is independent of $\mathcal{G}$. Then,

$$\mathbb{E}[h(X,Z)\,|\,\mathcal{G}] = H(Z) \quad a.s.,$$

where $H(z) = \mathbb{E}[h(X,z)]$ for all $z \in \mathbb{R}$.

Proof. (a) The proof is similar to Williams [129, Chapter 9.8, Proof of (h)] by simply using the general version of conditional expectation and its properties from [65, Chapter I.4] instead.

(b) A similar result also appears in [86, Lemma A.0.1(v)], as well as [13, Example 34.3] and thus we will just sketch the proof. First note that the result is true for functions of the form $h = 1_{F\times G}$ where F is $\mathcal{F}$-measurable and where G is $\mathcal{G}$-measurable. Using a monotone class argument we get the result for bounded measurable h. The general case follows by taking the limit over bounded functions tending to h. □

Lemma 10.4 (A Fubini-type theorem for generalised conditional expectation)**.** *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and σ-finite measure space (S,Σ,μ), let $X\colon S\times\Omega \to \mathbb{R}$*[33] *be a $\Sigma\otimes\mathcal{F}$-measurable stochastic process such that $\int_S X_s\,\mu(ds)$ is σ-integrable with respect to $\mathcal{G}$. Moreover, assume that there exists a $\Sigma\otimes\mathcal{G}$-measurable version of $(\mathbb{E}[X_s\,|\,\mathcal{G}])_{s\in S}$. Then,*

$$\mathbb{E}\left[\int_S X_s\,\mu(ds)\,\middle|\,\mathcal{G}\right] = \int_S \mathbb{E}[X_s\,|\,\mathcal{G}]\;\mu(ds) \quad a.s.$$

Proof. Define $R_X = \big\{G\in\mathcal{G}\;\big|\;\mathbb{E}\big[\big|\int_S X_s\,\mu(ds)\big|1_G\big] < \infty\big\}$. Then, using Fubini's theorem, cf. [71, Theorem 1.27], as well as fundamental properties of generalised conditional expectation, see [65, Chapter I.4],

$$\mathbb{E}\left[1_G\int_S X_s\,\mu(ds)\right] = \int_S \mathbb{E}[1_G X_s]\;\mu(ds) = \int_S \mathbb{E}\big[1_G\,\mathbb{E}[X_s\,|\,\mathcal{G}]\big]\,\mu(ds) < \infty\,, \quad G\in R_X\,.$$

If we apply Fubini's theorem again, we get

$$\mathbb{E}\left[1_G\int_S X_s\,\mu(ds)\right] = \mathbb{E}\left[1_G\int_S \mathbb{E}[X_s\,|\,\mathcal{G}]\;\mu(ds)\right], \quad G\in R_X\,,$$

which yields the result due to the definition of generalised conditional expectation and since $\int_S \mathbb{E}[X_s\,|\,\mathcal{G}]\;\mu(ds)$ is $\mathcal{G}$-measurable. □

Remark 10.5 (Joint measurability). Let $S=[0,\infty)$ and $\Sigma = \mathcal{B}([0,\infty))$. Then the condition that X is $\Sigma\otimes\mathcal{F}$-measurable is, for example, satisfied if X_s is $\mathcal{F}$-measurable for every $s\in S$, and if $[0,\infty)\ni t\mapsto X(t,\omega)$ is right-continuous or left-continuous for every $\omega\in\Omega$, cf. Karatzas and Shreve [72, Proposition 1.13]. If in addition those paths are increasing, then there exists a version of $(\mathbb{E}[X_s\,|\,\mathcal{G}])_{s\in S}$ which is $\Sigma\otimes\mathcal{G}$-measurable. This follows by a similar argumentation as in Lemma 3.15.

Next, we define time-consistency based on the papers of Delbaen [31, Section 6] and Acciaio and Penner [3].

[33] We define $X(s,\omega) = X_s(\omega)$ for all $s\in S$ and $\omega\in\Omega$.

Definition 10.6 (Time-consistency)**.** Recalling Definitions 1.1 and 1.3, as well as given a filtration $(\mathcal{F}_t)_{t\in I}$ with index set $I \subset [0,\infty)$, consider a dynamic risk measure $(\rho_t)_{t\in I}$ defined on a suitable subset $\mathcal{L} \subset L^0(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, consider stopping times $\tau, \sigma\colon \Omega \to I$ such that $\sigma \leq \tau$ a.s. For $X, Y \in \mathcal{L}$ consider property

$$\rho_\tau[X] \leq \rho_\tau[Y] \quad (\geq \text{ resp.}) \quad \Rightarrow \quad \rho_\sigma[X] \leq \rho_\sigma[Y] \quad (\geq \text{ resp.})\,. \tag{10.7}$$

Then, define the following:

(a) $(\rho_t)_{t\in I}$ is called *(strongly) time-consistent* if, for every $X \in \mathcal{L}$, either direction of (10.7) holds for all $Y \in \mathcal{L}$.

(b) $(\rho_t)_{t\in I}$ is called *middle rejection* or *acceptance consistent*, respectively, if, for every $X \in \mathcal{L}$, Implication (10.7) holds for all $Y \in \mathcal{L} \cap L^0(\Omega, \mathcal{F}_\sigma, \mathbb{P})$ where $\mathcal{F}_\sigma = \{A \in \mathcal{F} \mid A \cap \{\sigma \leq t\} \in \mathcal{F}_t, \text{ for all } t \in T\}$.

(c) $(\rho_t)_{t\in I}$ is called *weakly rejection* or *acceptance consistent*, respectively, if, for every $X \in \mathcal{L}$, Implication (10.7) holds for all constants $Y \in \mathbb{R}$.

Remarks 10.8. (Time-consistency).

(a) Note that definitions of time-consistency slightly vary in the literature.

(b) Every form of time-consistency can be interpreted as follows. If a financial position X is preferable to some other financial position or benchmark Y at a random time τ in the future, then this position should also be preferable at an earlier time σ. The richer the class of benchmarks Y, the stronger the result.

(c) For a definition of time-consistency based on bounded random variables we refer to Delbaen [31, Section 6], as well as Acciaio and Penner [3, Section 1.4.2] and the references therein. In [31, Section 6] equivalent characterisations are shown, including recursiveness.

(d) The definitions of rejection and acceptance consistency, as well as corresponding middle and weak versions are adapted from Acciaio and Penner [3, Section 1.4.2]. Note that in our case losses are assumed to be positive.

(e) Time-consistency is the strongest of the above properties and implies (b) and (c). Middle rejection and acceptance consistency imply weak rejection and acceptance consistency, respectively.

(f) For further readings on dynamic risk measures and time-consistency we refer to Acciaio and Penner [3], Cheridito, Delbaen and Kupper [23], Delbaen [31], Detlefsen and Scandolo [35], as well as Kupper and Schachermayer [77] and the references therein.

(g) In Rosazza Gianin [106] time-consistent risk measures are constructed using non-linear g-expectations arising from backward stochastic differential equations.

(h) Time consistency can also be defined for processes instead of random variables, cf. Cheridito, Delbaen and Kupper [23].

The following risk measure, called *conditional entropic risk measure*, is not analysed in this thesis but essential in the context of time-consistency in a continuous-time setting, see Kupper and Schachermayer [77, Subsection 1.2]. Entropic risk measures and conditional entropic risk measures are also analysed in Acciaio and Goldammer [2], Cheridito, Delbaen and Kupper [23], Detlefsen and Scandolo [35], as well as Föllmer and Schied [49, 50].

Definition 10.9 (Conditional entropic risk measure)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $\gamma\colon \Omega \to (-\infty, \infty]$ be $\mathcal{G}$-measurable. Then, recalling Definition 3.1 of the upper envelope $X^{\mathcal{G}}$, for an $\mathcal{F}$-measurable $X\colon \Omega \to \mathbb{R}$ the *conditional entropic risk measure* with parameter γ is given by

$$\rho_\gamma^{\mathrm{ent}}[X\,|\,\mathcal{G}] = \begin{cases} X^{\mathcal{G}} & \text{on } \{\gamma = \infty\}\,, \\ \frac{1}{\gamma}\log \mathbb{E}[e^{\gamma X}\,|\,\mathcal{G}] & \text{on } \{\gamma \in \mathbb{R}\setminus\{0\}\}\,, \\ \operatorname{ess\,inf}_{\gamma_0>0} \frac{1}{\gamma_0}\log \mathbb{E}[e^{\gamma_0 X}\,|\,\mathcal{G}] & \text{on } \{\gamma = 0\}\,. \end{cases}$$

Remarks 10.10. Given the assumptions of Definition 10.9, we get the following:

(a) If $\exp(\gamma X)$ is σ-integrable with respect to $\mathcal{G}$, then, using Jensen's inequality as given in Lemma 10.3(a), we get $\rho_0^{\mathrm{ent}}[X\,|\,\mathcal{G}] = \mathbb{E}[X\,|\,\mathcal{G}]$ and $1_{\{\gamma>0\}}\,\rho_\gamma^{\mathrm{ent}}[X\,|\,\mathcal{G}] \geq 1_{\{\gamma>0\}}\,\mathbb{E}[X\,|\,\mathcal{G}]$, as well as $1_{\{\gamma<0\}}\,\rho_\gamma^{\mathrm{ent}}[X\,|\,\mathcal{G}] \leq 1_{\{\gamma<0\}}\,\mathbb{E}[X\,|\,\mathcal{G}]$, all a.s.

(b) Conditional entropic risk measures are conditional convex risk measures. But, in general, they are not positively homogeneous and, thus, not conditionally coherent. In the special case $\gamma = \infty$ a.s. conditional entropic risk measures are conditionally positively homogeneous, see Remarks 3.3(i).

(c) Given a filtration $(\mathcal{F}_t)_{t\in I}$ with $I \subset [0,\infty)$, it is easy to see that conditional entropic risk measures with $\mathcal{F}_0$-measurable parameter γ are time-consistent. On $\{\gamma \in \mathbb{R}\setminus\{0\}\}$ this can be shown by direct calculation. The result immediately extends to $\{\gamma = 0\}$. On $\{\gamma = \infty\}$ time-consistency follows by Remark 3.6.

(d) Given a filtration $(\mathcal{F}_t)_{t\in I}$ with some index set $I \subset [0,\infty)$, assume that $\exp(\gamma X)$ is σ-integrable with respect to $\mathcal{G}$. Then, for a given $\mathcal{F}_0$-measurable parameter γ with $\gamma > 0$ a.s., $(\rho_\gamma^{\mathrm{ent}}[X\,|\,\mathcal{F}_t])_{t\in I}$ is a supermartingale which follows by Jensen's inequality as, for example, given in Lemma 10.3(a).

The definition below is geared for our purposes and gives the definition of *conditional comonotonicity.* In particular, if σ-algebra $\mathcal{G}$ is trivial, see Footnote 17, then the definition of conditional comonotonicity corresponds to the classical definition of comonotonicity as given in McNeil, Frey and Embrechts [85, Definition 5.15], i.e., for all $x, y \in \mathbb{R}$ we require $\mathbb{P}(X \leq x, Y \leq y) = \min\big\{\mathbb{P}(X \leq x), \mathbb{P}(Y \leq y)\big\}$.

Definition 10.11 (Conditional comonotonicity)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, two random variables $X, Y\colon \Omega \to \mathbb{R}$ are said to be *conditionally comonotonic* with respect to $\mathcal{G}$ if $\mathbb{P}(X \leq x, Y \leq y\,|\,\mathcal{G}) = \min\big\{\mathbb{P}(X \leq x\,|\,\mathcal{G}), \mathbb{P}(Y \leq y\,|\,\mathcal{G})\big\}$ a.s. for all $x, y \in \mathbb{R}$.

Remark 10.12. (Comonotonicity and conditional comonotonicity). An alternative definition of conditional comonotonicity and several other characterisations, including the definition we use here, are given in Jouini and Napp [68]. There, existence of regular conditional distributions and corresponding transition kernels is assumed. There exist further equivalent definitions of comonotonicity, e.g., definitions via comonotonic sets, which can potentially be generalised to the conditional case.

Lemma 10.13. *Given Definition 10.11, consider a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$. If X and Y are comonotonic, then they are conditionally comonotonic with respect to $\mathcal{G}$ as well. The reverse statement is not true, in general.*

Proof. Given $x, y \in \mathbb{R}$, note that by definition of classical comonotonicity we either have $\mathbb{P}(X > x, Y \leq y) = 0$ or $\mathbb{P}(X \leq x, Y > y) = 0$. Thus, either $\mathbb{P}(X > x, Y \leq y \,|\, \mathcal{G}) = 0$ a.s. or $\mathbb{P}(X \leq x, Y > y \,|\, \mathcal{G}) = 0$ a.s. since we would end up with a contradiction otherwise. The first result then follows since

$$\begin{aligned}\mathbb{P}(X \leq x, Y \leq y \,|\, \mathcal{G}) &= \mathbb{P}(X \leq x \,|\, \mathcal{G}) - \mathbb{P}(X \leq x, Y > y \,|\, \mathcal{G}) \\ &= \mathbb{P}(Y \leq y \,|\, \mathcal{G}) - \mathbb{P}(X > x, Y \leq y \,|\, \mathcal{G}) \quad \text{a.s.}\end{aligned}$$

For a counterexample that conditional comonotonicity does not imply comonotonicity in general, set $\mathcal{G} = \mathcal{F}$. Then,

$$\begin{aligned}\mathbb{P}(X \leq x, Y \leq y \,|\, \mathcal{G}) &= 1_{\{X \leq x, Y \leq y\}} = \min\{1_{\{X \leq x\}}, 1_{\{Y \leq y\}}\} \\ &= \min\big\{\mathbb{P}(X \leq x \,|\, \mathcal{G}), \mathbb{P}(Y \leq y \,|\, \mathcal{G})\big\} \quad \text{a.s.}\end{aligned}$$

Hence X and Y are always conditionally comonotonic with respect to $\mathcal{G} = \mathcal{F}$. If the second implication were true, all random variables would be comonotonic which obviously is a wrong statement. □

In the following, we define *conditional stochastic ordering.* A similar approach can be found in Rüschendorf [109, Section 2] which is a straight-forward generalisation of classical stochastic ordering, cf. Shaked and Shanthikumar [113]. The notation used here is adapted to the notation in [113].

Definition 10.14 (Conditional ordering)**.** Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $X, Y \colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable. Then, define the following:

(a) *Conditional first stochastic order:* If $\mathbb{E}[h(X) \,|\, \mathcal{G}] \leq \mathbb{E}[h(Y) \,|\, \mathcal{G}]$ for all increasing functions $h \colon \mathbb{R} \to \mathbb{R}$ such that $h(X)^-$ and $h(Y)^-$ are σ-integrable with respect to $\mathcal{G}$, define $X \leq_{\mathrm{st}(\mathcal{G})} Y$.

(b) *Conditional convex order:* If $\mathbb{E}[h(X) \,|\, \mathcal{G}] \leq \mathbb{E}[h(Y) \,|\, \mathcal{G}]$ for all convex functions $h \colon \mathbb{R} \to \mathbb{R}$ such that $h(X)^-$ and $h(Y)^-$ are σ-integrable with respect to $\mathcal{G}$, define $X \leq_{\mathrm{cx}(\mathcal{G})} Y$.

(c) *Conditional increasing convex order:* If $\mathbb{E}[h(X) \,|\, \mathcal{G}] \leq \mathbb{E}[h(Y) \,|\, \mathcal{G}]$, for all increasing and convex functions $h \colon \mathbb{R} \to \mathbb{R}$ such that $h(X)^-$ and $h(Y)^-$ are σ-integrable with respect to $\mathcal{G}$, define $X \leq_{\mathrm{icx}(\mathcal{G})} Y$.

Lemma 10.15. *Given a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $X, Y \colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable. Then, we have the following conditional properties:*

(a) If $\mathbb{E}[f(X) \,|\, \mathcal{G}] = \mathbb{E}[f(Y) \,|\, \mathcal{G}]$ a.s. for every bounded and continuous function $f \colon \mathbb{R} \to \mathbb{R}$, then, for every $\mathcal{G}$-measurable $Z \colon \Omega \to \mathbb{R}$,

$$\mathbb{P}(X \leq Z \,|\, \mathcal{G}) = \mathbb{P}(Y \leq Z \,|\, \mathcal{G}) \quad \textit{a.s.}$$

(b) If $X \leq_{\mathrm{st}(\mathcal{G})} Y$, then, for every $\mathcal{G}$-measurable $Z \colon \Omega \to \mathbb{R}$,

$$\mathbb{P}(X \leq Z \,|\, \mathcal{G}) \geq \mathbb{P}(Y \leq Z \,|\, \mathcal{G}) \quad \textit{a.s.}$$

(c) Let $h \colon \mathbb{R}^2 \to \mathbb{R}$ be a measurable function such that $\mathbb{R} \ni x \mapsto h(x, z)$ is increasing, as well as convex for all $z \in \mathbb{R}$ and such that $\mathbb{R} \ni z \mapsto h(x, z)$ is continuous for each $x \in \mathbb{R}$. If $X \leq_{\mathrm{icx}(\mathcal{G})} Y$, then, for all $\mathcal{G}$-measurable $Z \colon \Omega \to \mathbb{R}$ such that $h(X, Z)^-$ and $h(Y, Z)^-$ are σ-integrable with respect to $\mathcal{G}$,

$$\mathbb{E}[h(X, Z) \,|\, \mathcal{G}] \leq \mathbb{E}[h(Y, Z) \,|\, \mathcal{G}] \quad \textit{a.s.}$$

(d) Given $\mathcal{F}$-measurable $f_X, f_Y, X, Y\colon \Omega \to [0,\infty)$, assume that X and Y share the same conditional distribution given $\mathcal{G}$, i.e., $\mathcal{L}(X\,|\,\mathcal{G}) = \mathcal{L}(Y\,|\,\mathcal{G})$ a.s., as well as that f_Y and Y are conditionally comonotonic given $\mathcal{G}$, see Definition 10.11. Then, if $f_X \leq_{\mathrm{icx}(\mathcal{G})} f_Y$,

$$\mathbb{E}[f_X X\,|\,\mathcal{G}] \leq \mathbb{E}[f_Y Y\,|\,\mathcal{G}] \quad \textit{a.s.} \tag{10.16}$$

Proof. (a) Define $f_k(x,z) = (1 - k(x-z)^+)^+$ for all $k \in \mathbb{N}$ and $x, z \in \mathbb{R}$. Then, note that f_k is bounded and continuous in both variables for every $k \in \mathbb{N}$ and that

$$1_{\{X \leq Z\}} = \lim_{k\to\infty} f_k(X,Z) \quad \text{a.s.}$$

Moreover, note that there exists a sequence $(Z_n)_{n\in\mathbb{N}}$ of simple $\mathcal{G}$-measurable real-valued random variables with $Z_n \to Z$ a.s., i.e., Z_n is of the form $Z_n = \sum_{i=1}^{j_n} \alpha_{i,n} 1_{A_{i,n}}$ with $j_n \in \mathbb{N}$ and $\alpha_{i,n} \in \mathbb{R}$ where $\{A_{i,n}\}_{1\leq i\leq j_n}$ is a $\mathcal{G}$-measurable partition of Ω for every $n \in \mathbb{N}$. Thus, by conditional bounded convergence and conditional monotone convergence, see He, Wang and Yan [65, Theorems 1.19(1) and 1.20],

$$\mathbb{P}(X \leq Z\,|\,\mathcal{G}) = \lim_{k\to\infty} \lim_{n\to\infty} \sum_{i=1}^{j_n} 1_{A_{i,n}}\, \mathbb{E}\big[f_k(X, \alpha_{i,n})\,\big|\,\mathcal{G}\big] \quad \text{a.s.}$$

Consequently, since $f_k(\cdot, z)$ is bounded and continuous for every $k \in \mathbb{N}$ and $z \in \mathbb{R}$,

$$\mathbb{P}(X \leq Z\,|\,\mathcal{G}) = \lim_{k\to\infty} \lim_{n\to\infty} \sum_{i=1}^{j_n} 1_{A_{i,n}}\, \mathbb{E}\big[f_k(Y, \alpha_{i,n})\,\big|\,\mathcal{G}\big] \quad \text{a.s.},$$

which yields the result using the same argumentation for Y as well.

(b) Similarly as in the proof of (a), define $f_k(x,z) = (1 - k(x-z)^+)^+$ for $k \in \mathbb{N}$ and $x, z \in \mathbb{R}$ and note that these functions are bounded and continuous in both variables. Using the same notation and argumentation as above, we get

$$\mathbb{P}(X \leq Z\,|\,\mathcal{G}) = \lim_{k\to\infty} \lim_{n\to\infty} \sum_{i=1}^{j_n} 1_{A_{i,n}}\, \mathbb{E}\big[f_k(X, \alpha_{i,n})\,\big|\,\mathcal{G}\big] \quad \text{a.s.}$$

Consequently, since $1 - f_k(\cdot, z)$ is increasing in the first variable for $z \in \mathbb{R}$ fixed and since $X \leq_{\mathrm{st}(\mathcal{G})} Y$, we have

$$\mathbb{P}(X \leq Z\,|\,\mathcal{G}) \geq \lim_{k\to\infty} \lim_{n\to\infty} \sum_{i=1}^{j_n} 1_{A_{i,n}}\, \mathbb{E}\big[f_k(Y, \alpha_{i,n})\,\big|\,\mathcal{G}\big] \quad \text{a.s.}$$

Again, the result follows by conditional bounded and conditional monotone convergence, see [65, Theorems 1.19(1) and 1.20].

(c) The proof is similar to the cases before. Note that there exists a decreasing sequence of simple $\mathcal{G}$-measurable real-valued random variables with $Z_n \searrow Z$ a.s. where $Z_n = \sum_{i=1}^{k_n} \alpha_{i,n} 1_{A_{i,n}}$ with the same notation as in (a). Thus, since h is continuous in the second argument and by conditional monotone convergence, see [65, Theorem 1.19(1)], we get

$$\mathbb{E}[h(X,Z)\,|\,\mathcal{G}] = \lim_{n\to\infty} \sum_{i=1}^{k_n} 1_{A_{i,n}}\, \mathbb{E}\big[h(X, \alpha_{i,n})\,\big|\,\mathcal{G}\big] \quad \text{a.s.},$$

where the result remains true if we replace X by Y. Thus, the result follows by the definition of conditional increasing convex order.

(d) On $\{\mathbb{E}[f_Y Y \,|\, \mathcal{G}] = \infty\}$ the result is clear. Thus, we may assume $\mathbb{E}[f_Y Y \,|\, \mathcal{G}] < \infty$ a.s. Note that we have

$$\mathbb{E}[f_Y Y \,|\, \mathcal{G}] = \mathbb{E}\left[\int_0^\infty \int_0^\infty 1_{\{Y>u, f_Y>v\}} \, du \, dv \,\middle|\, \mathcal{G}\right] \quad \text{a.s.}$$

Using the conditional Fubini theorem of Lemma 10.4 twice, we get

$$\mathbb{E}[f_Y Y \,|\, \mathcal{G}] = \int_0^\infty \int_0^\infty \mathbb{P}(Y > u, f_Y > v \,|\, \mathcal{G}) \, du \, dv \quad \text{a.s.}$$

Note that all assumptions of Lemma 10.4 are satisfied. This can be shown by using monotone convergence as in the proof of Corollary 5.21 and by using suitable versions with measurable paths of $(\mathbb{P}(Y > u, f_Y > v \,|\, \mathcal{G}))_{u \geq 0}$, for fixed $v \geq 0$, as well as of $(\mathbb{E}[1_{\{f_Y > v\}} Y \,|\, \mathcal{G}])_{v \geq 0}$. Then, since Y and $f_{\mathcal{G},G,Y}$ are conditionally comonotonic with respect to $\mathcal{G}$, we get

$$\mathbb{E}[f_Y Y \,|\, \mathcal{G}] = \int_0^\infty \int_0^\infty \min\left\{\mathbb{P}(Y > u \,|\, \mathcal{G}), \mathbb{P}(f_Y > v \,|\, \mathcal{G})\right\} du \, dv \quad \text{a.s.}$$

Define $\bar{F}(u) := \mathbb{P}(X > u \,|\, \mathcal{G}) = \mathbb{P}(Y > u \,|\, \mathcal{G})$, as well as $\bar{G}_Y(u) := \mathbb{P}(f_Y > u \,|\, \mathcal{G})$ for all $u \geq 0$. Then, the integrand in the equation above can, for fixed $u \geq 0$, be written as

$$\min\{\bar{F}(u), \bar{G}_Y(v)\} = \bar{F}(u) 1_{[0,\alpha_Y(u)]}(v) + \bar{G}_Y(v) 1_{(\alpha_Y(u),\infty)}(v) \,, \quad v \geq 0 \,,$$

where $\alpha_Y(u) = \operatorname{ess\,inf}\{Z \geq 0 \,|\, Z \in L^0(\Omega, \mathcal{G}, \mathbb{P}) \text{ and } \bar{G}_Y(Z) \leq \bar{F}(u)\}$, see Footnote 3. Thus,

$$\mathbb{E}[f_Y Y \,|\, \mathcal{G}] = \int_0^\infty \left(\bar{F}(u) \alpha_Y(u) + \int_{\alpha_Y(u)}^\infty \bar{G}_Y(v) \, dv\right) du \quad \text{a.s.} \tag{10.17}$$

Similarly as above, define $\alpha_X(u) := \operatorname{ess\,inf}\{Z \geq 0 \,|\, Z \in L^0(\Omega, \mathcal{G}, \mathbb{P}) \text{ and } \bar{G}_X(Z) \leq \bar{F}(u)\}$ for $u \geq 0$, where $\bar{G}_X(z) := \mathbb{P}(f_X > z \,|\, \mathcal{G})$ for all $z \geq 0$. Similarly as for Equation (10.17), we get

$$\mathbb{E}[f_X X \,|\, \mathcal{G}] \leq \int_0^\infty \left(\bar{F}(u) \alpha_X(u) + \int_{\alpha_X(u)}^\infty \bar{G}_X(v) \, dv\right) du \quad \text{a.s.,}$$

with an inequality instead. Thus, we have the desired result if we can show, for every $u \geq 0$,

$$\bar{F}(u) \alpha_X(u) + \int_{\alpha_X(u)}^\infty \bar{G}_X(v) \, dv \leq \bar{F}(u) \alpha_Y(u) + \int_{\alpha_Y(u)}^\infty \bar{G}_Y(v) \, dv \quad \text{a.s.} \tag{10.18}$$

Note that the conditional Fubini theorem of Lemma 10.4 implies, for every $u \geq 0$,

$$\int_{\alpha_Y(u)}^\infty \bar{G}_Y(v) \, dv = \int_0^\infty \mathbb{E}\big[1_{\{\alpha_Y(u) \leq v < f_Y\}} \,\big|\, \mathcal{G}\big] \, dv = \mathbb{E}\big[(f_Y - \alpha_Y(u))^+ \,\big|\, \mathcal{G}\big] \quad \text{a.s.}$$

Since $f_X \leq_{\mathrm{icx}(\mathcal{G})} f_Y$, we have $\mathbb{E}[(f_X - \alpha)^+ \,|\, \mathcal{G}] \leq \mathbb{E}[(f_Y - \alpha)^+ \,|\, \mathcal{G}]$ a.s. for all $\alpha \in \mathbb{R}$. Hence, by approximating $\alpha_Y(u)$ with $\mathcal{G}$-measurable simple functions and by using conditional monotone convergence, we may conclude

$$\int_{\alpha_Y(u)}^\infty \bar{G}_X(v) \, dv \leq \int_{\alpha_Y(u)}^\infty \bar{G}_Y(v) \, dv \quad \text{a.s., for all } u \geq 0 \,.$$

Hence, given $u \geq 0$, set $M := \{\alpha_Y(u) \geq \alpha_X(u)\}$, as well as $M^{\mathrm{c}} := \Omega \setminus M$ and conclude that the inequality in (10.18) follows if

$$\bar{F}(u)\,(\alpha_X(u) - \alpha_Y(u)) + 1_M \int_{\alpha_X(u)}^{\alpha_Y(u)} \bar{G}_X(v)\,dv - 1_{M^{\mathrm{c}}} \int_{\alpha_Y(u)}^{\alpha_X(u)} \bar{G}_X(v)\,dv \leq 0 \quad \text{a.s.},$$

or, equivalently,

$$0 \leq 1_M \int_{\alpha_X(u)}^{\alpha_Y(u)} (\bar{F}(u) - \bar{G}_X(v))\,dv + 1_{M^{\mathrm{c}}} \int_{\alpha_Y(u)}^{\alpha_X(u)} (\bar{G}_X(v) - \bar{F}(u))\,dv \quad \text{a.s.} \tag{10.19}$$

Since we have $\bar{F}(u) \geq \bar{G}_X(v)$ a.s., on $\{v > \alpha_X(u)\}$, as well as $\bar{F}(u) \leq \bar{G}_X(v)$ a.s., on $\{0 \leq v < \alpha_X(u)\}$, the integrals in (10.19) are both positive and finally (10.16) follows. □

10.2 Some alternative proofs for conditional expected shortfall

Some alternative proofs of Lemma 5.23. (a) On $\{0 < \delta < 1\}$, by definition of conditional expected shortfall, $\mathrm{ES}_\delta[0\,|\,\mathcal{G}] = \frac{1}{1-\delta}\,\mathbb{E}[f_{\mathcal{G},\delta,0}\,0\,|\,\mathcal{G}] = 0$ a.s. Thus, on $\{\delta = 0\}$,

$$\mathrm{ES}_0[0\,|\,\mathcal{G}] = \operatorname*{ess\,inf}_{\delta \in (0,1)} \frac{1}{1-\delta}\,\mathbb{E}[f_{\mathcal{G},\delta,0}\,0\,|\,\mathcal{G}] = 0 \quad \text{a.s.}$$

Obviously, the upper envelope satisfies $X^{\mathcal{G}} = 0$ a.s. and thus the statement follows on $\{\delta = 1\}$.

(b) On $\{0 < \delta < 1\}$, by Lemma 3.18(b), we have $q_{\mathcal{G},\delta}(X\,Z) = q_{\mathcal{G},\delta}(X)\,Z$ a.s. Therefore, we know that $\beta_{\mathcal{G},\delta,X\,Z} = \beta_{\mathcal{G},\delta,X}$ a.s. and $f_{\mathcal{G},\delta,X\,Z} = f_{\mathcal{G},\delta,X}$ a.s. As Z is $\mathcal{G}$-measurable and bounded, we get $\mathrm{ES}_\delta[X\,Z\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}]\,Z$ a.s. using [65, Theorem 1.21]. On $\{\delta = 0\}$ the result follows from the representation of conditional expected shortfall at level zero using the essential infimum. On $\{\delta = 1\}$ the result follows from Remark 3.3(i) as $\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = q_{\mathcal{G},\delta}(X) = X^{\mathcal{G}}$.

(c) On $\{0 < \delta < 1\}$, by Lemma 3.18(c), we have $q_{\mathcal{G},\delta}(X + Z) = q_{\mathcal{G},\delta}(X) + Z$ a.s. Again, this implies $\beta_{\mathcal{G},\delta,X+Z} = \beta_{\mathcal{G},\delta,X}$ a.s. and $f_{\mathcal{G},\delta,X+Z} = f_{\mathcal{G},\delta,X}$ a.s. Using linearity of conditional expectation again and $\mathcal{G}$-measurability of Z, we get $\mathrm{ES}_\delta[X + Z\,|\,\mathcal{G}] = \mathrm{ES}_\delta[X\,|\,\mathcal{G}] + Z$ a.s. by [65, Theorems 1.18 and 1.21]. On $\{\delta = 0\}$ the result follows from the representation of conditional expected shortfall using the essential infimum. On $\{\delta = 1\}$ the result follows from Remark 3.3(h) as $\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = q_{\mathcal{G},\delta}(X) = X^{\mathcal{G}}$ a.s.

(d) On $\{0 < \delta < 1\}$, as

$$\mathbb{E}[f_{\mathcal{G},\delta,X}\,|\,\mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,Y}\,|\,\mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X+Y}\,|\,\mathcal{G}] = 1 - \delta \quad \text{a.s.},$$

note that $\mathcal{F}^1_{\mathcal{G},\delta,X} = \mathcal{F}^1_{\mathcal{G},\delta,Y} = \mathcal{F}^1_{\mathcal{G},\delta,X+Y}$ a.s. By Lemma 5.23(t1) and linearity of conditional expectation, see [65, Theorem 1.18], we get

$$\begin{aligned}
\mathrm{ES}_\delta[X + Y\,|\,\mathcal{G}] &= \frac{1}{1-\delta} \operatorname*{ess\,sup}_{f \in \mathcal{F}^1_{\mathcal{G},\delta,X+Y}} \mathbb{E}[f\,(X+Y)\,|\,\mathcal{G}] \\
&\leq \frac{1}{1-\delta}\Big(\operatorname*{ess\,sup}_{f \in \mathcal{F}^1_{\mathcal{G},\delta,X+Y}} \mathbb{E}[f\,X\,|\,\mathcal{G}] + \operatorname*{ess\,sup}_{f \in \mathcal{F}^1_{\mathcal{G},\delta,X+Y}} \mathbb{E}[f\,Y\,|\,\mathcal{G}] \Big) \\
&= \mathrm{ES}_\delta[X\,|\,\mathcal{G}] + \mathrm{ES}_\delta[Y\,|\,\mathcal{G}] \quad \text{a.s.}
\end{aligned}$$

On $\{\delta = 0\}$ the result follows from the representation of conditional expected shortfall using the essential infimum. On $\{\delta = 1\}$ the result follows from Remarks 3.3(f).

(e) This result follows from (b) and (d).

(f) and (g) On $\{0 < \delta < 1\}$ both results follow from Lemma 3.18(d) and (e), the quantile representation in Lemma 5.12 and the linearity of Lebesgue integrals. On $\{\delta = 0\}$ the same result follows from the representation of conditional expected shortfall using the essential infimum.

(k) On $\{0 < \delta < 1\}$, the result immediately follows by the quantile representation of conditional expected shortfall, see Lemma 5.12, and by Lemma 3.18(h). The result extends to the infimum, i.e. on $\{\delta = 0\}$. On $\{\delta = 1\}$, the result also follows by Lemma 3.18(h).

(l) On $\{0 < \delta < 1\}$, the result follows by a fundamental property of generalised conditional expectation, see [65, Theorem 1.21]. The result extends to the infimum, i.e., on $\{\delta = 0\}$. On $\{\delta = 1\}$, the result also follows by Lemma 3.3(i).

(p) By translation invariance from (c), we may assume that without loss of generality X_n is non-negative for every $n \in \mathbb{N}$. Similar to the proof of (d), we can show that $\mathcal{F}^1_{\mathcal{G},\delta,X} = \mathcal{F}^1_{\mathcal{G},\delta,X_n}$ a.s. for every $n \in \mathbb{N}$. By Definition 5.3, we may write

$$(1-\delta)\,\mathrm{ES}_\delta[X\,|\,\mathcal{G}] = \mathbb{E}[f_{\mathcal{G},\delta,X}\,X\,|\,\mathcal{G}]\ .$$

Using Fatou's Lemma for conditional expectation, see [65, Theorem 1.19(2)], we get

$$\mathbb{E}[f_{\mathcal{G},\delta,X}\,X\,|\,\mathcal{G}] \le \liminf_{n\to\infty} \mathbb{E}[f_{\mathcal{G},\delta,X}\,X_n\,|\,\mathcal{G}] \quad \text{a.s.}$$

Furthermore, by Lemma 5.7, for every $n \in \mathbb{N}$ we have

$$\mathbb{E}[f_{\mathcal{G},\delta,X}\,X_n\,|\,\mathcal{G}] \le \operatorname*{ess\,sup}_{f\in\mathcal{F}^1_{\mathcal{G},\delta,X_n}} \mathbb{E}[f\,X_n\,|\,\mathcal{G}] = (1-\delta)\,\mathrm{ES}_\delta[X_n\,|\,\mathcal{G}] \quad \text{a.s.}$$

Dividing by $1-\delta$ proves the result. □

Part II

Risk Aggregation with Applications to Credit and Life Insurance

Chapter 11

Modelling Annuity Portfolios with Extended CreditRisk$^+$

In this chapter we develop an approach for modelling annuity, life insurance and credit portfolios using a special version of extended CreditRisk$^+$ as given in Schmock [111, Section 6]. Dependence is introduced via common stochastic risk factors which can be identified with different death causes in the context of life insurance and annuities. Within this model, there exists an efficient, numerically stable algorithm for deriving loss distributions exactly. Furthermore, we point out possible generalisations of our annuity model and give an introductory example. In particular, this chapter illustrates our way of thinking and prepares the reader for all further applications.

11.1 Annuity portfolios

In this section we introduce the key components of our annuity model. The setting can immediately be applied to other life insurance portfolios.

Definition 11.1 (Policyholders and death indicators)**.** Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\{1, \ldots, m\}$ with $m \in \mathbb{N}$ denote the set of *policyholders* in the annuity portfolio and let $\mathcal{F}$-measurable death indicators $N_1, \ldots, N_m \colon \Omega \to \mathbb{N}_0$ indicate the *number of deaths* of each policyholder in the following period. Event $\{N_i = 0\}$ indicates no death for $i \in \{1, \ldots, m\}$.

In reality, death indicators are Bernoulli random variables[34] as each person can just die once. Unfortunately in practice, such an approach is not tractable for calculating loss distributions of large portfolios as execution times of implementations explode. Alternatively, one can always rely on Monte Carlo techniques which are computationally expensive if numerical errors should be small. On the contrary, we will assume the number of deaths of a each policyholder to be compound Poisson distributed. As we are going to see in Lemma 11.19, assuming our model with Poisson distributed deaths gives an efficient way for calculating loss distributions using an algorithm based on Panjer's recursion, also for large portfolios. Ultimately, calibration of the model also gets easier since sums of independent Poisson distributions are Poisson distributed again, with a modified intensity.

There are mainly two possibilities how to calibrate death indicators $N_1, \ldots, N_m$.

[34] A random variable X is Bernoulli distributed with parameter $q \in [0, 1]$ if $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = q$.

Definition 11.2 (Scaling via survival probabilities)**.** Given Definition 11.1, assume that $\mathbb{P}(N_i \geq 1) = q_i^*$ for all $i \in \{1, \ldots, m\}$ where q_i^* denotes the probability of death of policyholder i in the following period.

Remark 11.3 (Alternative scaling via expectations)*.* Instead of matching survival probabilities as described above, one can also set $\mathbb{E}[N_i] = q_i^*$ for all $i \in \{1, \ldots, m\}$. Several numerical trials in our setting show that this alternative approach mostly gives worse results than the approach from the definition above, see Section 19.2. In particular, in the tail where just few deaths happen. This alternative scaling approach is more risk averse since, in that case, survival probabilities are higher than in the approach of Definition 11.2. In Section 16.1, we compare our annuity model to the Lee–Carter model and see better fits if we set $\mathbb{E}[N_i] = q_i^*$, especially for older age categories.

Remark 11.4 (Multiple deaths)*.* Obviously, the proposed model has a major shortcoming as it allows for multiple deaths of each policyholder. From a theoretical point of view, the approach with random sums, in particular random Poisson sums, is justified by the Poisson approximation and generalisations of it, see for example Vellaisamy and Chaudhuri [124]. Since annual death probabilities for ages up to 85 are less than 10 percent, multiple deaths are relatively unlikely for all major ages. See Remarks 11.23(d) for a short comparison of errors made by Monte Carlo to errors made by the Poisson mixture approach.

Definition 11.5 (Payments)**.** Given Definition 11.1, let $d \in \mathbb{N}$ denote the dimension of payments including a dimension for annuity payments to policyholders[35]. The independent $\mathcal{F}$-measurable random vectors $X_1, \ldots, X_m \colon \Omega \to \mathbb{N}_0^d$ denote portfolio *payments* within the following period given survival, i.e., on $\{N_i = 0\}$ for all $i \in \{1, \ldots, m\}$. Correspondingly, the independent[36] $\mathcal{F}$-measurable random vectors $Y_1, \ldots, Y_m \colon \Omega \to \mathbb{N}_0^d$ denote portfolio payments in the following period which need not be paid or which are not received due to death, i.e., on $\{N_i \geq 1\}$ for all $i \in \{1, \ldots, m\}$, and are assumed to be independent of $N_1, \ldots, N_m$.

Remarks 11.6. (Annuity payments and reserves).

(a) Given a policyholder $i \in \{1, \ldots, m\}$, each dimension of the d-valued random vector X_i represents positive payments to or from this policyholder in the case of survival over the next period such as annuities paid to i, premiums paid by i or actuarial reserves[37] being declared at the end of the next period. In the case many policyholders hold several insurance contracts, further dimensions for different lines of business can be added. In practice, not more than three dimensions are recommended as otherwise the recursive algorithm described in Lemma 11.19 can become very time-consuming. Correspondingly, Y_i represents payments of or to policyholder i in the following period which need not be paid in the case i dies within this period. Thus, note that X_i and Y_i do not necessarily share the same distribution due to possible sub periodical payments. Positivity in every component of X_i and Y_i is required as otherwise Panjer's recursion does not work. Nevertheless, we can model payments with opposite signs with our d-dimensional setting, see Remark 11.24.

[35] Further dimensions may represent for paid premiums, actuarial reserves to be declared and payments for various other lines of business, see Remarks 11.6.

[36] To prepare our model for Panjer's recursion, see Lemma 11.19, we assume independence amongst payments. Dependence is later on introduced via dependent number of deaths.

[37] The actuarial reserve of a contract at the time t is the conditional expected value of all discounted future cash flows and thus, in general, stochastic.

(b) For all $i \in \{1, \dots, m\}$, X_i and Y_i may be stochastic as in the case of unit-linked annuities, for monthly payments, when using stochastic discount factors or for annuities with optionality, as well as deterministic as in the case of fixed pension payments and premiums with deterministic discounting.

(c) A possible setting. Let policyholder $i \in \{1, \dots, m\}$ be fixed and let A_i denote annuity payments to this policyholder within the following period given that he or she survives. Moreover, in the case i survives, let P_i denote the premium that has to be paid and let R_i denote the actuarial reserve for the corresponding contract which has to be declared at the end of this period. Correspondingly, for a random variable A_i' having the same distribution as A_i, let $A_i' U_i$ denote annuity payments within the next period which need not be paid in the case of death of policyholder i where U_i is continuously uniformly distributed on $(0, 1]$ or discretely uniformly distributed on $\{1/m, 2/m, \dots, 1\}$ with $m \in \mathbb{N}$. This indicates continuous or periodic payments throughout a period. Also, for a random variable P_i' having the same distribution as P_i, let $P_i' U_i$ denote premiums which are not paid by i due to death. Then, set $X_i = (A_i, P_i, R_i)$ and $Y_i = (A_i' U_i, P_i' U_i, R_i)$. Note that Y_i thus becomes the sub-periodic fraction of payments which need need not be paid in the case of death appearing uniformly throughout the period.

(d) Using the technique of *stochastic rounding*, see Schmock [111], we may assume X_i and Y_i to be $[0, \infty)^d$-valued for all $i \in \{1, \dots, m\}$.

Remark 11.7 (Time issues). For notational convenience we omit time indices as we are mostly confronted with a one-period setting. If required, we add a time index t to all quantities appearing in our model as, for example, done in the context of parameter estimation or forecasting.

Definition 11.8 (Total loss)**.** Given Definitions 11.1 and 11.5, define cumulative payments which need not be paid due to deaths

$$S := \sum_{i=1}^{m} \sum_{j=1}^{N_i} Y_{i,j} \,,$$

where $(Y_{i,j})_{j \in \mathbb{N}}$ for every $i \in \{1, \dots, m\}$ is an i.i.d. sequence of random variables with $\mathcal{L}(Y_{i,j}) = \mathcal{L}(Y_i)$ for all $i \in \{1, \dots, m\}$ and $j \in \mathbb{N}$ where $\mathcal{L}$ denotes the distribution of the argument. Then, the total portfolio loss is defined as

$$L := \sum_{i=1}^{m} X_i - S \,.$$

Remarks 11.9. (Total loss).

(a) S is the sum of all annuity payments, premiums and actuarial reserves which need not be paid, are received and declared, respectively, in the following period due to deaths of policyholders. L on the other hand is the total portfolio loss over the next period. Note that we are interested in large losses of L, i.e., the right tail of its distribution. This translates into the case where just few policyholders die such that many annuity payments have to be made. Correspondingly, small values of S, i.e., the left tail of its distribution, is the part of major interest and major risk.

(b) Since Poisson approximation just works properly for small values of death probabilities q_i^* for all $i \in \{1, \ldots, m\}$, extended CreditRisk$^+$ is not suitable to calculate loss L directly via the sum

$$\sum_{i=1}^{m} \sum_{j=1}^{\overline{N}_i} Y_{i,j}$$

where $\overline{N}_i$ denotes the survival indicator with $\mathbb{P}(\overline{N}_i \geq 1) = 1 - q_i^*$ and where $(Y_{i,j})_{j \in \mathbb{N}}$ for all $i \in \{1, \ldots, m\}$ are i.i.d. copies of Y_i.

(c) Appropriate dependence structures between $\sum_{i=1}^{m} X_i$ and S have to be assumed. The cases of independence, as well as perfect positive and negative dependence, called comonotonicity and countermonotonicity, are easy to calculate. The illustrative example in Section 19.2 suggests that assuming independence will be sufficient in many applications. This is intuitive in the presence of monthly or fortnightly payments due to diversification effects over time.

(d) In Section 19.2 we give an illustrative example which compares the model with Poisson distributed deaths to the model with Bernoulli distributed deaths.

(e) (Bounds for value at risk and expected shortfall) Letting $d = 1$ and given marginal distributions of $\sum_{i=1}^{m} X_i$ and S, it is always possible to derive approximative bounds for value at risk of L using techniques given in the works of Embrechts, Rüschendorf and Puccetti [42, 97], for example. Note that upper and lower bounds for quantiles of L are in general not obtained by the extreme dependence scenarios of comonotonicity and countermonotonicity as shown in Embrechts and Puccetti [41]. Upper bounds for expected shortfall on arbitrary levels are easy to obtain as this risk measure is comonotonically additive, as well as sub additive, see Schmock [111], which implies that risk is maximised under countermonotonicity of $\sum_{i=1}^{m} X_i$ and S. Techniques for further bounds of expected shortfall can be found in Puccetti [96].

(f) The sum $\sum_{i=1}^{m} X_i$ can be calculated using usual convolution, fast Fourier transform (FFT) or normal approximation, given that all required additional assumptions are satisfied, respectively.

(g) If N_i is a Bernoulli random variable and if $X_i = Y_{i,1}$ a.s. for $i \in \{1, \ldots, m\}$, then the sum

$$L^* = \sum_{i=1}^{m} Y_{i,1} - S = \sum_{i=1}^{m} (1 - N_i) Y_{i,1}$$

calculates the exact loss and, therefore, we refer to L^* as the loss of the *exact model*.

11.2 Annuity model with independent risk factors

To make our model applicable in practical situations and to ensure a flexible handling in terms of multi-level dependence, we introduce stochastic risk factors. Risk factors are designed to model effects which simultaneously influence death probabilities of many policyholders due to a common exposure to the same type of risk. In the context of annuities and life insurance, risk factors can be identified with causes of death such as neoplasms, cardiovascular diseases or idiosyncratic components. In terms of credit risk, risk factors may correspond to economic variates such as gas prices or political stability.

Definition 11.10 (Common stochastic risk factors)**.** Given Definitions 11.1, 11.2 and 11.5, consider $\mathcal{F}$-measurable *risk factors* $\Lambda_1, \dots, \Lambda_K \colon \Omega \to [0, \infty)$ with $K \in \mathbb{N}_0$ and corresponding weights $w_{i,0}, \dots, w_{i,K} \in [0, 1]$ for every policyholder $i \in \{1, \dots, m\}$. Risk index zero represents idiosyncratic risk and we require $w_{i,0} + \cdots + w_{i,K} = 1$ for all $i \in \{1, \dots, m\}$.

To guarantee a flexible and yet numerically tractable model, we need to make probabilistic assumptions. The approach here is based on Schmock [111, Section 6]. This model is referenced as extended CreditRisk$^+$ and enables us to apply an algorithm based on iterated Panjer's recursion.

Definition 11.11 (The annuity model)**.** Given Definitions 11.5, 11.1, 11.2 and 11.10, we call our model an *annuity model* if in addition the following is satisfied:

(a) Death indicators $N_{1,0}, \dots, N_{m,0} \colon \Omega \to \mathbb{N}_0$ are independent from one another, as well as all other random variables and, for all $i \in \{1, \dots, m\}$, they are Poisson distributed with intensity $q_i w_{i,0}$ where $q_i := -\log(1 - q_i^*)$, i.e.,

$$\mathbb{P}\Big(\bigcap_{i=1}^{m} \{N_{i,0} = n_{i,0}\}\Big) = \prod_{i=1}^{m} e^{-q_i w_{i,0}} \frac{(q_i w_{i,0})^{n_{i,0}}}{n_{i,0}!}\,, \quad n_{1,0}, \dots, n_{m,0} \in \mathbb{N}_0\,.$$

(b) Risk factors $\Lambda_1, \dots, \Lambda_K \colon \Omega \to [0, \infty)$ are independent and, for all $k \in \{1, \dots, K\}$, they have a gamma distribution with mean $e_k = 1$ and variance $\sigma_k^2 > 0$, i.e., with shape and inverse scale parameter $1/\sigma_k^2$ such that their densities are given by

$$f_{\Lambda_k}(x) = \begin{cases} \frac{(e_k/\sigma_k^2)^{e_k^2/\sigma_k^2}}{\Gamma(e_k^2/\sigma_k^2)} e^{-x e_k/\sigma_k^2} x^{e_k^2/\sigma_k^2 - 1} & \text{for } x > 0\,, \\ 0 & \text{for } x \le 0\,, \end{cases}$$

where $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t}\, dt$ for $x > 0$ denotes the gamma function. Also the degenerate case with $\sigma_k^2 = 0$ for $k \in \{1, \dots, K\}$ is allowed.

(c) Given risk factors, death indicators $(N_{i,k})_{i \in \{1,\dots,m\}, k \in \{1,\dots,K\}} \colon \Omega \to \mathbb{N}_0^{m \times K}$ are independent and, for every policyholder $i \in \{1, \dots, m\}$ and $k \in \{1, \dots, K\}$, they are Poisson distributed with random intensity $q_i w_{i,k} \Lambda_k$, i.e.,

$$\mathbb{P}\Big(\bigcap_{i=1}^{m} \bigcap_{k=1}^{K} \{N_{i,k} = n_{i,k}\} \,\Big|\, \Lambda_1, \dots, \Lambda_K\Big) = \prod_{i=1}^{m} \prod_{k=1}^{K} e^{-q_i w_{i,k} \Lambda_k} \frac{(q_i w_{i,k} \Lambda_k)^{n_{i,k}}}{n_{i,k}!} \quad \text{a.s.},$$

for all $(n_{i,k})_{i \in \{1,\dots,m\}, k \in \{1,\dots,K\}} \in \mathbb{N}_0^{m \times K}$.

(d) For every policyholder $i \in \{1, \dots, m\}$, the total number of deaths N_i is split up additively according to risk factors as

$$N_i = N_{i,0} + \cdots + N_{i,K}\,.$$

Thus, by our model construction, $\mathbb{E}[N_i] = q_i (w_{i,0} + \cdots + w_{i,K}) = -\log(1 - q_i^*)$.

Remarks 11.12. (Death probabilities and age categories).

(a) For notational convenience, q_i is termed as death probability even though it is an intensity and just an approximation to the true death probability $q_i^* = 1 - \exp(-q_i)$ for all $i \in \{1, \dots, m\}$.

(b) Usually, death probabilities q_i and weights $w_{i,k}$ are for each gender categorised into age groups so that policyholders within a certain age band and same gender share the same parameters. In our estimation example based on Australian data, we consider homogeneous age categories of five years length.

(c) Notation is kept general as individual information or risk behaviour of certain policyholders may be incorporated into death probabilities and weights. To be able to use individual information correctly, portfolio data are necessary and estimation procedures have to be adapted, see Remark 13.1.

Remark 11.13 (Interpretation of risk factors). Item (c) in Definition (11.11) states that if risk factor Λ_k for death cause $k \in \{1, \dots, K\}$ takes large or small values, then the likelihood of death due to cause k increases or decreases, respectively, simultaneously for all policyholders depending on the weight $w_{i,k}$. Given policyholder $i \in \{1, \dots, m\}$, note that weights $w_{i,0}, \dots, w_{i,K}$ indicate the vulnerability of policyholder i to risk factors $\Lambda_1, \dots, \Lambda_K$. For a practical example, assume that a new, very effective cancer treatment is available such that fewer people die from lung cancer. This situation would have a longevity effect on all policyholders, but particularly on smokers. Such a scenario would then correspond to the case when the risk factor for neoplasms shows a small realisation. The other way round, assume that we face a very hot summer. Then the likelihood to pass away due to heart failure increases. This example would correspond to a large realisation of the risk factor for cardiovascular diseases. Since such scenarios are previously unknown, it makes sense to model risk factors stochastically which, in our setting, immediately leads to stochastic death probabilities.

Remark 11.14 (Moments of $N_{i,k}$). Given the annuity model from Definition 11.11 with K non-idiosyncratic risk factors, let $k \in \{1, \dots, K\}$ and consider policyholder $i \in \{1, \dots, m\}$. Then, for the number of deaths $N_{i,k}$ due to risk factor Λ_k we have

$$\mathbb{E}[N_{i,k}] = \mathbb{E}[\mathbb{E}[N_{i,k} \,|\, \Lambda_k]] = \mathbb{E}[q_i \, w_{i,k} \, \Lambda_k] = q_i \, w_{i,k} \,, \tag{11.15}$$

and, using the law of total variance as in Schmock [111, Lemma 3.48],

$$\begin{aligned} \mathrm{Var}(N_{i,k}) &= \mathbb{E}[\mathrm{Var}(N_{i,k} \,|\, \Lambda_k)] + \mathrm{Var}(\mathbb{E}[N_{i,k} \,|\, \Lambda_k]) \\ &= \mathbb{E}[q_i \, w_{i,k} \, \Lambda_k] + \mathrm{Var}(q_i \, w_{i,k} \, \Lambda_k) \\ &= q_i \, w_{i,k} (1 + q_i \, w_{i,k} \, \sigma_k^2) \,. \end{aligned} \tag{11.16}$$

Analogously, for all $i, j \in \{1, \dots, m\}$ with $i \neq j$,

$$\begin{aligned} \mathrm{Cov}(N_{i,k}, N_{j,k}) &= \mathbb{E}[\mathrm{Cov}(N_{i,k}, N_{j,k} \,|\, \Lambda_k)] + \mathrm{Cov}(\mathbb{E}[N_{i,k} \,|\, \Lambda_k] \,, \mathbb{E}[N_{j,k} \,|\, \Lambda_k]) \\ &= 0 + q_i \, q_j \, w_{i,k} \, w_{j,k} \, \mathrm{Cov}(\Lambda_k, \Lambda_k) \\ &= q_i \, q_j \, w_{i,k} \, w_{j,k} \, \sigma_k^2 \,. \end{aligned} \tag{11.17}$$

This result will be used in Section 17 for model validation. A similar result also holds for the more general model with multi-level dependent risk factors, see Schmock [111, Section 6.5].

As already mentioned in the introduction, there exists a numerically stable algorithm to derive the loss distribution of S. Based on the more general approach as given in the lecture notes of Schmock [111, Section 6.7], we briefly recall the algorithm so that the reader can immediately implement it. For the more general algorithm and a pseudo implementation of it see Section 19.1.

Definition 11.18. Given the annuity model of Definition 11.11, for notational convenience in the next lemma define the cumulative Poisson intensity

$$\lambda_{k,\nu} := \sum_{i=1}^{m} q_i \, w_{i,k} \, \mathbb{P}(Y_i = \nu) \,,$$

for loss size $\nu \in \mathbb{N}_0^d \setminus \{0\}$ due to risk factor $k \in \{0, \dots, K\}$, and, correspondingly, the cumulative Poisson intensity for non-zero losses

$$\bar{\lambda}_k := \sum_{\nu \in \mathcal{S}_k} \lambda_{k,\nu} = \sum_{i=1}^{m} q_i \, w_{i,k} \, (1 - \mathbb{P}(Y_i = 0))$$

where $\mathcal{S}_k := \{\nu \in \mathbb{N}_0^d \setminus \{0\} \,|\, \lambda_{k,\nu} > 0\}$. For $k \in \{0, \dots, K\}$, if $\bar{\lambda}_k > 0$, define

$$q_{k,\nu} := \begin{cases} \lambda_{k,\nu}/\bar{\lambda}_k & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\} \,, \\ 0 & \text{for } \nu = 0 \in \mathbb{N}_0^d \,, \end{cases}$$

as well as if $\bar{\lambda}_k = 0$,

$$q_{k,\nu} := \begin{cases} 0 & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\} \,, \\ 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d \,. \end{cases}$$

Finally, define $p_k := \bar{\lambda}_k \sigma_k^2 / (1 + \bar{\lambda}_k \sigma_k^2) \in [0, 1)$ for all $k \in \{1, \dots, K\}$, as well as

$$\lambda := \bar{\lambda}_0 + \sum_{k=1}^{K} \frac{\bar{\lambda}_k}{1 + \bar{\lambda}_k \sigma_k^2} \, c(p_k)$$

where

$$c(p) := \sum_{n \in \mathbb{N}} \frac{p^{n-1}}{n} = \begin{cases} -\frac{\log(1-p)}{p} & \text{for } p \in (0, 1) \,, \\ 1 & \text{for } p = 1 \,. \end{cases}$$

Note that all definitions also work in the degenerate case $\sigma_k^2 = 0$ for $k \in \{1, \dots, K\}$.

Lemma 11.19 (Algorithm for exact derivation of the loss distribution)**.** *Given the annuity model of Definition 11.11 and considering Definition 11.18, there exists a numerically stable algorithm based on Panjer's recursion which allows an exact computation of the probability distribution of S up to every desired cumulative probability. More precisely,* $\mathbb{P}(S = 0) = \exp\big(\lambda(c_0 - 1)\big)$ *and, recursively,*[38]

$$\mathbb{P}(S = \nu) = \frac{\lambda}{\nu_i} \sum_{\substack{n = (n_1, \dots, n_d) \in \mathbb{N}_0^d \\ 0 < n \leq \nu}} n_i \, c_n \, \mathbb{P}(S = \nu - n) \,, \quad \nu = (\nu_1, \dots, \nu_d) \in \mathbb{N}_0^d \setminus \{0\} \,, \tag{11.20}$$

where $i \in \{1, \dots, d\}$ can be chosen arbitrarily such that $\nu_i \neq 0$ and where

$$c_\nu = \frac{1}{\lambda} \left(\bar{\lambda}_0 \, q_{0,\nu} - \sum_{k=1}^{K} b_{k,\nu} \frac{\bar{\lambda}_k}{1 + \bar{\lambda}_k \sigma_k^2} \, c(p_k) \right) , \quad \nu \in \mathbb{N}_0^d \,. \tag{11.21}$$

[38] The inequality $0 < n \leq \nu$ is to be understood in a component-wise sense where for the strict inequality it suffices to have a strict inequality in at least one component.

If $\lambda > 0$, then, for all $k \in \{1, \dots, K\}$, $b_{k,0} = q_{k,0}\, c(p_k\, q_{k,0})/c(p_k)$, as well as

$$b_{k,\nu} = \frac{1}{1 - p_k\, q_{k,0}} \left(\frac{q_{k,\nu}}{c(p_k)} + \frac{p_k}{\nu_i} \sum_{\substack{n \in \mathcal{S}_k, \\ n \leq \nu}} (\nu_i - n_i)\, q_{k,n}\, b_{k,\nu-n} \right), \quad \nu \in \mathbb{N}_0^d \setminus \{0\}, \tag{11.22}$$

Conversely, if $\lambda = 0$, then

$$c_\nu = \begin{cases} 0 & \text{for } \nu \in \mathbb{N}_0^d \setminus \{0\}, \\ 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d. \end{cases}$$

Proof. A detailed derivation of the more general formula in extended CreditRisk$^+$ is given in Schmock [111, Sections 6.6 and 6.7]. The main idea is to represent the random sum S as a Poisson sum. This can be achieved via deriving the probability-generating function of S which is, for at least all $z = (z_1, \dots, z_d) \in \mathbb{C}^d$ with $\|z\|_\infty \leq 1$, given by

$$\mathbb{E}\left[\prod_{i=1}^d z_i^{S_i}\right] = \sum_{\nu=(\nu_1,\dots,\nu_d)\in\mathbb{N}_0^d} \mathbb{P}(S = \nu) \prod_{i=1}^d z_i^{\nu_i} = \exp\left(\lambda(\tilde{\varphi}(z) - 1)\right),$$

where $\tilde{\varphi}(z) = \sum_{\nu\in\mathbb{N}_0^d} c_\nu \prod z_i^{\nu_i}$ with c_ν given by (11.21). The form of the probability-generating function implies that S is a Poisson sum which, by applying multi-variate Panjer's recursion, gives the result. □

Remarks 11.23. (Comments on the extended CreditRisk$^+$ algorithm).

(a) If $Y_1, \dots, Y_m$ are one-dimensional and deterministic, then the algorithm above is basically due to Giese [54] for which Haaf, Reiß and Schoenmakers [63] proved numerical stability. The relation to Panjer's recursion was first pointed out by Gerhold, Schmock and Warnung [53, Section 5.5]. Schmock [111, Section 5.1] generalised the algorithm to the multivariate case with dependent risk factors and risk groups, based on the multivariate extension of Panjer's algorithm given by Sundt [118].

(b) The recursive sums in (11.20) and (11.22) are due to the multivariate extension of Panjer's algorithm. Since just positive terms are added, the algorithm is numerically stable, in general. Nevertheless, numerical underflow may occur, see Remark 16.9 as well as Rudolph [108].

(c) The extended CreditRisk$^+$ model has no stochastic errors but approximates death indicators via compound Poisson distributions. Implementations of the algorithm described above are significantly faster than Monte Carlo approximations for comparable error levels. To avoid long execution times for implementations of extended CreditRisk$^+$ with large annuity portfolios, greater loss units can be used, i.e., random variables $Y_1, \dots, Y_m$ are rounded a priori to multiples of some $\mathbb{N}$-valued loss unit. Negative effects of this deviation from exact calculations can be reduced by using stochastic rounding, see Schmock [111, Section 6.2.2]. There, random variables are rounded to loss units such that expectations remain the same before and after rounding. Also for the calculation of value at risk and expected shortfall, smoothing algorithms can be used to get more accurate results.

(d) To compare execution times of extended CreditRisk$^+$ to Monte Carlo we may look at a portfolio with just idiosyncratic risk consisting of $m = 10\,000$ policyholders, each

having a death probability of $q = 0.015$ where we choose the alternative scaling as outlined in Remark 11.3. Losses $Y_1, \ldots, Y_m$ are deterministic and equal to one for all policyholders. Thus in the case of Bernoulli distributed death indicators N_i the sum S has binomial distribution with parameters $(10\,000, 0.015)$. Using Poisson approximation, see Schmock [111], we can conclude that the total variation[39] between the distributions for the model with Bernoulli distributed and Poisson distributed deaths is bounded above by 0.015. On the other hand, using Monte Carlo with 50 000 simulations as an approximation for the true model with Bernoulli distributed deaths, the total variation between those distributions is 0.0159 in our simulation and, thus, dominates the Poisson approximation in terms of total variation. Our implementation in 'R' has a system time of 21.6 seconds for the Monte Carlo approach and 0.01 seconds for extended CreditRisk$^+$ up to a cumulative probability of 0.999. Execution times for extended CreditRisk$^+$ depend on how clever you choose recursions in (11.20) as many quantities equal zero. This simple example illustrates that with similar accuracy Monte Carlo is significantly slower than extended CreditRisk$^+$.

Remark 11.24 (Approximation for multi-dimensional settings). Given our annuity model with $d \geq 2$, note that the algorithm described in Lemma 11.19 returns the exact distribution of S up to some cumulative level $\delta \in (0, 1)$—usually close to one—called a sub-distribution. If we are interested in the distribution of $f(S)$ for some measurable function $f\colon \mathbb{N}_0^d \to \mathbb{R}$, then the previously derived sub-distribution can be used to derive an approximation. More explicitly, let μ denote the probability measure induced by $f(S)$ and let ν denote the corresponding measure induced by the sub-distribution with $\nu(\mathbb{R}) = \delta < 1$. Then, the total variation distance between μ and ν, see Footnote 39, is given by

$$d_{\mathrm{TV}}(\mu, \nu) = 1 - \delta\,.$$

This, in particular, applies to settings where $Y_1, \ldots, Y_m$ are also allowed to take negative values. In that case, we can simply define

$$Y_i := (\max\{Y_i, 0\}, -\min\{Y_i, 0\})$$

and get an approximation for the total loss $S_1 - S_2$ via the extended CreditRisk$^+$ algorithm with $S = (S_1, S_2)$. Note that Panjer's recursion does not allow for a direct derivation of total loss distributions with positive and negative losses.

11.3 Generalised and alternative models

Up to now, we applied a simplified version of extended CreditRisk$^+$ to derive cumulative payments in annuity portfolios. A major shortcoming in this approach is the limited possibility of modelling dependencies amongst policyholders and death causes. In the most general form of extended CreditRisk$^+$ as described in Schmock [111, Section 6], it is possible to introduce risk groups which enable us to model joint deaths of several policyholders and

[39] The total variation distance d_{TV} between two probability measures μ and ν, e.g., push-forward measures induced by random variables, on a measurable space $(S, \mathcal{S})$ is defined by

$$d_{\mathrm{TV}}(\mu, \nu) := \sup_{A \in \mathcal{S}} (\mu(A) - \nu(A))\,.$$

See, e.g., Schmock [111, Definition 3.7] and the references therein.

it is possible to model dependencies amongst death causes, see Section 19.1.1. Dependencies can take a linear dependence structure combined with dependence scenarios to model negative correlations as well. Risk factors may then be identified with statistical variates such as average blood pressure, average physical activity or the average of smoked cigarettes, etc., and not directly with death causes. Moreover, for each policyholder individually, the general model allows for losses which depend on the underlying cause of death. This gives scope to the possibility of modelling—possibly new—life insurance products with payoffs depending on the cause of death as, for example, in the case of accidental death benefits. Including all extensions mentioned above, a similar algorithm as given in Lemma 11.19 may still be applied to derive loss distributions, again see Schmock [111, Section 6.7], as well as Section 19.1.1. Estimation of model parameters on the other hand gets more involved and is subject to current research.

Instead of using extended CreditRisk$^+$ to model annuity portfolios, i.e., an approach based on Poisson mixtures, we can assume a similar *Bernoulli mixture model*. In such a Bernoulli mixture model, conditionally Poisson distributed deaths are simply replaced by conditionally Bernoulli distributed deaths. A variation of a Bernoulli mixture model may in our case be given via replacing Definition 11.11(a), (c) and (d) by $\mathcal{L}(N_{i,0}) = \text{Bernoulli}(q_i\, w_{i,0})$ and

$$\mathcal{L}(N_{i,k}\,|\,\Lambda_1,\ldots,\Lambda_K) = \mathcal{L}(N_{i,k}\,|\,\Lambda_k) = \text{Bernoulli}(\min\{1, q_i\, w_{i,k}\,\Lambda_k\}) \quad \text{a.s.}\,,$$

as well as

$$N_i = \min\{1, N_{i,0} + N_{i,1} + \cdots + N_{i,K}\}\,,$$

respectively, for all $i \in \{1,\ldots,m\}$ and $k \in \{1,\ldots,K\}$. The textbook of McNeil, Frey and Embrechts [85, Section 8] gives a comprehensive introduction to credit risk models including Poisson and Bernoulli mixture models. In general, explicit and efficient derivation of loss distributions in the case of Bernoulli mixture models is not possible anymore. Thus, in this case, one has to rely on other methods such as Monte Carlo. Estimation of model parameters works similarly as discussed in Section 12 modulo some obvious changes in the posterior density and likelihood as illustrated in (16.12). For Bernoulli mixture models it is possible to give asymptotic distributions for large portfolios, see [85, Section 8.4.3] again and the references therein. As illustrated in Section 19.2, Poisson approximation, see for example Vellaisamy and Chaudhuri [124], suggests that loss distributions derived from Bernoulli and Poisson mixture models are similar in terms of total variation distance if death probabilities are small.

Another modelling approach is the usage of *threshold models* where default occurs if some critical random variable falls below a deterministic critical value, see McNeil, Frey and Embrechts [85, Section 8.3]. Threshold models use copulas to model dependence. Furthermore, in their work [85, Section 8.4.4] it is shown that threshold models may be written as Bernoulli mixture models. Thus, arguing with Poisson approximation and assuming independent risk factors, Bernoulli mixture models, as well as threshold models can be approximated by our proposed model. Versions of threshold models include CreditMetrics and KMV models, again see [85, Example 8.6], which both provide the feature of considering credit rating migrations.

11.4 An introductory example with a common risk factor

In this example we consider an annuity portfolio with the main objective of illustrating the effect of a common stochastic risk factor. Therefore, consider the annuity model of

Definition 11.11 with an artificial portfolio of five groups and deterministic annual payments 10, 20, 30, 40, as well as 50, each having 1 000 policyholders, i.e., $m = 5\,000$ in total. For simplicity, there is no other form of surrender or any other form of contract and there are no actuarial reserves. In each of those five groups, half of the people have an annual death probability of $q_i = 0.05$ whereas the other half has an annual death probability of $q_i = 0.1$. Thus, if no policyholder dies, the insurer has to face cumulative payments of 150 000.

To create dependence between policyholders, we introduce one non-idiosyncratic risk factor Λ_1 with $\sigma_1^2 := \mathrm{Var}(\Lambda_1) = 0.25$ and provide three different settings for corresponding weightings. For the first case define weights $w_{i,0} = w_{i,1} = 0.5$ for each policyholder $i \in \{1, \ldots, m\}$ which means that each policyholder is equally influenced by idiosyncratic risk and by Λ_1. For the second case set $w_{i,1} = 1$ for all policyholders $i \in \{1, \ldots, m\}$ which means that there is no idiosyncratic risk. This corresponds to the situation when a change in risk factor Λ_1 hits all policyholders simultaneously with 100 percent. Of course, this setting produces heavier tails, i.e., a higher likelihood that just very few people die. For the third case we switch to $w_{i,1} = 0$ for all policyholders $i \in \{1, \ldots, m\}$ which means that only idiosyncratic risk is present and deaths occur independently among all policyholders.

Table 11.1: Value at risk of L at different levels δ, i.e., $q_\delta(L)$, in our annuity model using the extended CreditRisk$^+$ algorithm with a loss unit of one.

level δ	$w_{i,0} = 0.5$ $w_{i,0} = 0.5$	$w_{i,0} = 1$ $w_{i,1} = 0$	$w_{i,0} = 0$ $w_{i,1} = 1$
0.950	142 600	139 800	146 220
0.990	143 470	140 220	147 750
0.999	144 210	140 690	148 870

Table 11.1 lists value at risk of L, see Definition 11.8, in our artificial annuity portfolio using the classical CreditRisk$^+$ algorithm[40] with a loss unit of one. Not surprisingly, the third case with no idiosyncratic risk creates the highest risk since there is a high probability that a low realisation of risk factor Λ_1 leads to just very few deaths. This is due to the fact that the Poisson intensity $q_i \Lambda_1$ of $N_{i,1}$ gets very small for all policyholders $i \in \{1, \ldots, m\}$ simultaneously and, therefore, increases the likelihood of surviving. Note that the model with $w_{i,1} = 0$ involves just idiosyncratic risk and is not influenced by the risk factor Λ_1, i.e., deaths occur independently with probability q_i for all policyholders $i \in \{1, \ldots, m\}$.

To demonstrate that the quantities derived in our model are close to those of a Bernoulli mixture model we compare them via Monte Carlo. The number of deaths in that case equals $\min\{N_{i,0} + N_{i,1}, 1\}$ for all policyholders $i \in \{1, \ldots, m\}$ where $N_{i,0}$ is Bernoulli distributed with $q_i(1 - w_{i,1})$ and where $N_{i,1}$ is conditionally Bernoulli distributed with given realisations of risk factor Λ_1, i.e., the probability of death of i due to risk factor Λ_1 given $\Lambda_1 = \lambda$ is $\min\{q_i w_{i,1} \lambda, 1\}$. Using 50 000 simulations of $\Lambda_1 = \lambda$, each followed by a simulation of $\min\{N_{i,0} + N_{i,1}, 1\}$ for all $i \in \{1, \ldots, m\}$, Table 11.2 gives corresponding value at risk for the total portfolio loss L at various levels. In brackets, conservative 95 percent confidence intervals for value at risk estimates in our simulation are given, i.e., intervals such that with a probability of at least 95 percent the true values of value at risk lie in them. The method

[40] In this simple case with deterministic losses we can use the traditional CreditRisk$^+$ algorithm as described in [16] and need not use extended CreditRisk$^+$. In 'R' the package 'crp.CSFP' [67] provides an implementation of CreditRisk$^+$ using the algorithm described in Giese [54] for which Haaf, Reißand Schoenmakers [63] proved numerical stability.

Table 11.2: Value at risk of L at different levels δ, i.e., $q_\delta(L)$, in a Bernoulli mixture model at different levels using 50 000 simulations with 95 percent binomial confidence intervals in brackets.

level δ	$w_{i,0} = 0.5$ $w_{i,1} = 0.5$	$w_{i,0} = 1$ $w_{i,1} = 0$	$w_{i,0} = 0$ $w_{i,1} = 1$
0.950	142 670 (−30;+30)	139 750 (−10;+10)	146 170 (−50;+40)
0.990	143 510 (−40;+50)	140 150 (−20;+20)	147 720 (−50;+80)
0.999	144 240 (−70;+40)	140 570 (−30;+50)	148 850 (−60;+90)

to calculate these intervals can be found in Shevchenko [114, Section 3.2.1]. Comparing the results of Table 11.1 and Table 11.2, we immediately see the close relationship amongst those two approaches.

Increasing the number of simulations of Λ_1 leads to decreased sample variances of derived value at risk and tighter error bounds. As one would expect, empirical variances of derived value at risk increase with higher levels of value at risk.

Conclusively, we observe that the calculations in extended CreditRisk$^+$ are very fast, yet accurate compared to a Bernoulli mixture model, besides all approximations. Changing weightings from purely idiosyncratic risk towards risk which is concentrated in a common stochastic risk factor creates heavier tails in loss distribution L. Thus, in that case, value at risk increases significantly.

Chapter 12

Parameter Estimation of our Annuity Model

In this chapter we provide several approaches for parameter estimation in our annuity model given publicly available data based on the whole population of a country. We develop the following four estimation approaches: Matching of moments, a version of maximum a posteriori, maximum likelihood and Markov chain Monte Carlo (MCMC). Whilst matching of moments estimates are easy to derive in real world applications, maximum a posterior and maximum likelihood estimates cannot be calculated by deterministic numerical optimisation. Thus, we use MCMC as a slow but powerful alternative. We later apply these different approaches in an illustrative example, Section 12.5, as well as in a real world example, see Chapter 15.

McNeil, Frey and Embrechts [85, Section 8.6] consider statistical inference for Poisson mixture models and Bernoulli mixture models. They briefly introduce moment estimators and maximum likelihood estimators for homogeneous groups in Bernoulli mixture models. Alternatively, they derive statistical inference via a generalised linear mixed model representation for mixture models which is distantly related to our setting. In their 'Notes and Comments' section the reader can find a comprehensive list of interesting references. Nevertheless, most of their results and arguments are not directly applicable to our case since we use a different parametrisation and since we usually have rich data of death counts compared to the sparse information of company defaults.

Our primary goal is to identify risk factors and estimate their variances, as well as corresponding weights and death probabilities. We consider trends in mortality, as well as trends in risk factor weights and model them as non-random events where overfitting should be avoided. Therefore, we suggest the usage of suitably easy trend curves which are parametrised by a few parameters. All remaining random fluctuations should be explained by risk factors and their variations. Note that all proposed parameter families of death probabilities and weights can be changed freely in order to meet specific needs. Such changes just result in minor, obvious adaptions in certain formulas. This issue is particularly easy to address within the Markov chain Monte Carlo approach as introduced in Section 12.4.

In order to be able to derive statistically sound estimates, we make the following simplifying assumptions:

Assumption 12.1 (Simplifying assumptions for estimation of risk factors)**.** *Given the annuity model from Definition 11.11, consider discrete-time periods* $1, \ldots, T$[41] *and additionally assume the following:*

(a) For all $t \in \{1, \ldots, T\}$*, quantities* $q_i(t)$ *and corresponding weights* $w_{i,k}(t)$*, respectively, are the same for all representative policyholders* $i \in \{1, \ldots, m\}$ *within the same age category* $a \in \{1, \ldots, A\}$*, same gender* $g \in \{\mathrm{f}, \mathrm{m}\}$ *and with respect to the same risk factor* $\Lambda_k(t)$ *with death cause* $k \in \{0, \ldots, K\}$*. For notational purposes we may therefore define* $q_{a,g}(t) := q_i(t)$ *and* $w_{a,g,k}(t) := w_{i,k}(t)$ *for a representative policyholder* i *of age category* a *and gender* g *with respect to risk factor* $\Lambda_k(t)$*.*

(b) All random variables at time $t \in \{1, \ldots, T\}$ *are assumed to be independent of random variables at some different point in time* $s \neq t$ *with* $s \in \{1, \ldots, T\}$*.*

(c) For each $k \in \{1, \ldots, K\}$*, risk factors* $\Lambda_k(1), \ldots, \Lambda_k(T)$ *are identically distributed.*

Remarks 12.2. (Simplifying assumptions).

(a) Assumption 12.1(a) is just needed for consistent estimation and is reasonable in the sense that we do not have individual information of dead people and how exposed they were to certain risk factors. For prediction purposes, within a portfolio of policyholders, individual death probabilities and weights can be considered since additional information as, for example, smoker or non-smoker may be available.

(b) Assumption 12.1(b) is also needed for estimation purposes but may easily be violated in practice. If, for example, fewer people die from neoplasms in a certain year due to a new treatment, then more people will die from other causes in subsequent years since everyone has to die at some point. This phenomenon can be seen as a serial correlation effect. But as we will remove a lot of dependence via trends in death probabilities and weights, see Assumption 12.12, such dependence effects seem to be negligible for Australian data which is shown in Section 17 via several validation techniques.

Data for the number of living people and deaths, as well as data for causes of deaths are usually freely available on governmental websites. In the case of Australia data can be found at the Australian Bureau of Statistics, AIHW, or related institutions. If suitable rich information of deaths and their causes is available for a certain portfolio of policyholders, estimation can of course be based on this specific data. Nevertheless, we suggest to base parameter estimation on data from the whole population of a country since this guarantees suitable rich information for all death causes and minimal selection effects.

Assumption 12.3 (Available data)**.** *For every age category* $a \in \{1, \ldots, A\}$*, gender* $g \in \{\mathrm{f}, \mathrm{m}\}$ *and year* $t \in \{1, \ldots, T\}$ *with* $T \geq 2$ *the database is assumed to contain historical population counts* $m_{a,g}(t)$[42] *and historical number of deaths* $n_{a,g,k}(t)$[43] *due to underlying death cause* $k \in \{0, 1, \ldots, K\}$*. An underlying death cause is to be understood as the disease or injury that initiated the train of morbid events leading directly to death.*

[41] In this section we add t as time index.

[42] For Australia, estimates for resident population data are available at the website of the Australian Bureau of Statistics where a detailed documentation of the used statistical methods is given. Based on census counts, several adjustment components such as census undercount and immigration are taken into account.

[43] For Australia, we may take ICD-9 and ICD-10 classified death data from AIHW.

Remark 12.4 (Death probabilities). To be consistent in our approach, we will estimate death probabilities $q_{a,g}(t)$ from death data $n_{a,g,k}(t)$ and $m_{a,g}(t)$. Usually, death probabilities are also publicly available in the form of second order life tables[44] where effects such as migration are taken into account as well. Note that estimation of death probabilities always requires a careful handling of mortality trends.

To make our model applicable to real world data, we have to specify common stochastic risk factors $(\Lambda_0(t), \ldots, \Lambda_K(t))_{t\in\{1,\ldots,T\}}$. We recommend the approach to directly identify risk factors with death causes $0, 1, \ldots, K$. This leads to the following assumption.

Assumption 12.5 (Data and model linkage)**.** *Given Assumption 12.3, as well as our annuity model of Definition 11.11, the observations of historical annual deaths* $n_{a,g,k}(t)$[45] *with age* $a \in \{1, \ldots, A\}$*, gender* $g \in \{\mathrm{f}, \mathrm{m}\}$*, due to death cause* $k \in \{0, \ldots, K\}$ *and at time* $t \in \{1, \ldots, T\}$ *correspond to realisations of the random variable*

$$N_{a,g,k}(t) := \sum_{i\in M_{a,g}(t)} N_{i,k}(t)\,,$$

where $M_{a,g}(t) \subset \{1, \ldots, m(t)\}$ *denotes the set of representative policyholders of specified age group and gender with* $|M_{a,g}(t)| = m_{a,g}(t)$*. Note that* $N_{i,k}(t)$ *is the number of deaths of policyholder* i *due to death cause* k *in year* t*. Death cause zero corresponds to ill-defined and not reported deaths, i.e., idiosyncratic components.*

Remark 12.6 (Weights). Given Assumption 12.5 and using Remark 11.14, we have

$$\mathbb{E}[N_{a,g,k}(t)] = m_{a,g}(t)\, q_{a,g}(t)\, w_{a,g,k}(t)\,,$$

for all $a \in \{1, \ldots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$, $k \in \{0, \ldots, K\}$ and $t \in \{1, \ldots, T\}$. In particular, this implies that in average the weight $w_{a,g,k}(t)$ gives the fraction of people dying from death cause k compared to all deaths, i.e., $\mathbb{E}[N_{a,g,k}(t)]\,/(m_{a,g}(t)\, q_{a,g}(t))$. Moreover, for all $a \in \{1, \ldots, A\}$ and $g \in \{\mathrm{f}, \mathrm{m}\}$ we have $\mathbb{E}[\sum_{k=0}^{K} N_{a,g,k}(t)] = m_{a,g}(t)\, q_{a,g}(t)$.

Since expectations of risk factors $(\Lambda_1(t), \ldots, \Lambda_K(t))_{t\in\{1,\ldots,T\}}$ are by assumption fixed to one, it remains to estimate variances of risk factors, corresponding weights and death probabilities. Note that in the parametrisation we use, see Assumption 11.11, we have $\mathrm{Var}(\Lambda_k(t)) = \sigma_k^2$ for all $k \in \{1, \ldots, K\}$ and $t \in \{1, \ldots, T\}$.

Whilst it is common knowledge that people tend to live longer, real world data also show that weights for certain death causes change heavily over time. This also happens on a short-term scale and mostly with a clear monotone trend. If we did not account for trends in weights, then estimated risk factor variances would be far too high and residuals would not be gamma distributed. To avoid overfitting on the other hand, we do not want to make trends too complicated. Thus, to account for mortality trends, we use the following family of death probabilities and weights. Note that once we have estimated parameters within this family, we can make projections of death probabilities and weights into the future, see Section 16 for further discussions in this topic. In order for these parameter families to be well-defined, we use the following functions.

[44] For Australia this information is available at the Australian Bureau of Statistics for 2002-2012.

[45] As a convention throughout this thesis, estimators are always denoted by capital letters whereas realisations of these estimators, as well as estimates are always written with corresponding lower case letters.

Definition 12.7 (Laplace distribution and trend reduction)**.** The Laplace distribution function with mean one and variance two is given by

$$F^{\mathrm{Lap}}(x) = \frac{1}{2} + \frac{1}{2}\mathrm{sign}(x)\big(1 - \exp(-|x|)\big)\,, \quad x \in \mathbb{R}\,, \tag{12.8}$$

with corresponding (lower) quantile function

$$(F^{\mathrm{Lap}})^{-1}(y) = -\mathrm{sign}(2y-1)\log(1-|2y-1|)\,, \quad y \in [0,1]\,. \tag{12.9}$$

Trend reduction with parameters $(\zeta, \eta) \in \mathbb{R} \times (0, \infty)$ is given by

$$\mathcal{T}_{\zeta,\eta}(t) = \frac{1}{\eta}\arctan(\zeta + \eta t)\,, \quad t \in \mathbb{R}\,. \tag{12.10}$$

Remarks 12.11. Given the definition above, we can draw some immediate conclusions.

(a) For $x < 0$, (12.8) becomes $\exp(x)/2$.

(b) Expression (12.10) will we used for a trend reduction technique which is motivated by Kainhofer, Predota and Schmock [69, Section 4.6.2]. There they replace linear time $t \in \mathbb{N}_0$ by time shift $\mathcal{T}_{0,\eta}(t)$ with $\eta = \frac{1}{t_0}$. Then, parameter η gives the inverse of the time t_0 when an initial trend is halved. Parameter ζ on the other hand gives the shift on the arctangent curve.

(c) Note that $\lim_{x\to\pm\infty}\arctan(x) = \pm\frac{\pi}{2}$.

(d) Expression (12.10) is related to the Cauchy distribution function $F^{\mathrm{C}}_{\zeta,\eta}$ with parameters $(\zeta, \eta) \in \mathbb{R} \times [0, \infty)$ via $F^{\mathrm{C}}_{\zeta,\eta}(x) = \frac{1}{2} + \frac{1}{\pi}\mathcal{T}_{\zeta,\eta}(x)$ for all $x \in \mathbb{R}$.

Assumption 12.12 (Parameter family for trends in death probabilities and weights)**.** *Given the annuity model of Definition 11.11 and Assumption 12.1, as well as* $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ *and* $t \in \{1, \dots, T\}$, *death probability* $q_{a,g}(t) \in [0, 1]$ *satisfies*

$$q_{a,g}(t) = F^{\mathrm{Lap}}\big(\alpha_{a,g} + \beta_{a,g}\mathcal{T}_{\zeta_{a,g},\eta_{a,g}}(t)\big)\,, \tag{12.13}$$

where $\alpha_{a,g}, \beta_{a,g}, \zeta_{a,g} \in \mathbb{R}$ *and* $\eta_{a,g} \in (0, \infty)$. *Additionally given* $k \in \{0, \dots, K\}$, *weight* $w_{a,g,k}(t) \in [0, 1]$ *satisfies*

$$w_{a,g,k}(t) = \frac{\exp\big(u_{a,g,k} + v_{a,g,k}\mathcal{T}_{\phi_k,\psi_k}(t)\big)}{\sum_{j=0}^{K}\exp\big(u_{a,g,j} + v_{a,g,j}\mathcal{T}_{\phi_j,\psi_j}(t)\big)}\,, \tag{12.14}$$

with $u_{a,g,0}, v_{a,g,0}, \phi_0, \dots, u_{a,g,K}, v_{a,g,K}, \phi_K \in \mathbb{R}$, *as well as* $\psi_0, \dots, \psi_K \in (0, \infty)$. *Define the support of parameters for death probabilities* $E := \mathbb{R}^{3\times A\times 2)} \times (0, \infty)^{A+2}$ *and for weights* $F := \mathbb{R}^{2\times A\times 2\times K+K} \times (0, \infty)^K$.

Remarks 12.15. Given Assumption 12.12, we may draw some immediate conclusions.

(a) Death probabilities (12.13) and weights (12.14) are between zero and one where, in particular, the constraint $w_{a,g,0} + \cdots + w_{a,g,K} = 1$ for all $a \in \{1, \dots, A\}$ and $g \in \{\mathrm{f}, \mathrm{m}\}$ is satisfied.

(b) Vectors α[46] and u can be interpreted as intercept parameters for death probabilities and weights, respectively. Henceforth, β and v are trend parameters, see Remarks 12.11(b). Parameters ζ, as well as ϕ give the shift of trend reduction and η, as well as ψ give the corresponding speed of trend reduction. The smaller the values of η and ψ, the slower the trend reduction. Kainhofer, Predota and Schmock [69, Section 4.6.2] suggest a value of 0.01 for the speed of trend reduction, i.e., trends have a half time of 100 years. Meaningful values for η and ψ usually lie in the interval $(0.001, 0.1)$. A clear trend reduction in mortality improvements can be observed in Japan since 1970, see Pasdika and Wolff [91, Section 4.2], and also for females in Australia, see Remark 16.14. Since particularly Japan has a very old population, it seems reasonable to assume trend reduction techniques for long-term forecasts in other countries.

(c) Estimation issues. To avoid a far too complicated modelling setup, we assume trend parameters ϕ_k and ψ_k in (12.14) to be depend solely on death cause $k \in \{0, \ldots, K\}$. The intuition behind this approach is that the evolution in the trend (initial point of trend reduction ϕ_k and speed of trend reduction ψ_k) is equal over all age categories and genders for a given death cause as better treatments influence all groups simultaneously. Still, trend parameters in (12.14) and also in (12.13) are usually hard to estimate when just few years of observations are available as various parameter values roughly yield the same trend curve—in particular for values of η and ψ close to zero and absolute values of ζ and ϕ above one. It is therefore suggested that all or some of these parameters are chosen to be fixed if long-term projections are not the primarily goal. Moreover, to avoid messy behaviour of estimation procedures, we suggest to assume one as an upper bound for parameters η and ψ, as well as an upper bound for absolute values of ζ and ϕ. This is not a major restriction as otherwise trends would be too extreme.

(d) As (12.10) gives roughly a linear function of t if parameter η is small, we can replace $\mathcal{T}_{\zeta_{a,g},\eta_{a,g}}(t)$ by t to provide a simpler setting guaranteeing easier estimation. Note that trend reduction (12.10) guarantees that limiting values for death probabilities and weights are non-degenerate.

(e) Note that as death probabilities are lower than 0.5 for most ages, (12.13) gives roughly an exponential decay in time, see Remarks 12.11(a). Thus, it gets obvious that (12.13) is motivated by the Lee–Carter model, see Lee and Carter [78], Brouhns, Denuit and Vermunt [17], as well as Kainhofer, Predota and Schmock [69, Section 4.5.1], where the time-dependent term $\mathcal{T}_{\zeta_{a,g},\eta_{a,g}}(t)$ is replaced by time-dependent trend components κ_t and then estimated via a combination of method of moments and a singular value decomposition. See Section 17 for a link between our approach and the Lee–Carter method. Furthermore, our approach is linked to the Swiss Nolfi-Ansatz, see, for example, Kainhofer, Predota and Schmock [69, Section 4.5].

(f) As also mentioned in Remark 12.22, we could base all parameter families for death probabilities and weights on logistic regression. Then, we unfortunately loose the link to the Lee–Carter approach.

(g) For old ages, the mortality trend given in (12.13) might not be sufficient and, therefore, models should be selected carefully. For a discussion on this topic see Kainhofer, Predota and Schmock [69, Section 4.7.2].

[46] For notational purposes in the context of estimation, we write α for $(\alpha_{a,g}(t))_{a\in\{1,\ldots,A\},g\in\{\mathrm{f,m}\},t\in\{1,\ldots,T\}}$ and analogously for all other high-dimensional parameters appearing in this thesis.

(h) For the maximum a posteriori approach in Section 12.2, the maximum likelihood approach in Section 12.51 and corresponding MCMC approaches in Section 12.4, families for death probabilities and weights can be modified arbitrarily without changing the principle of each method. In particular, phenomena such as cohort effects can be incorporated, see Cairns et al. [19], as well as Remark 16.14.

(i) Note that for fixed $a \in \{1, ..., A\}$ and $g \in \{\mathrm{f}, \mathrm{m}\}$ Equation (12.14) is invariant under a constant shift of parameters $(u_{\mathrm{a,g},k})_{k\in\{0,...,K\}}$ as well as of parameters $(v_{\mathrm{a,g},k})_{k\in\{0,...,K\}}$ if $\phi_0 = \cdots = \phi_K$ and $\psi_0 = \cdots = \psi_K$ for the latter. Thus, for each $\mathrm{a} \in 1, ..., A$ and $\mathrm{g} \in \{\mathrm{f}, \mathrm{m}\}$, we can always chose fixed and arbitrary values for $u_{\mathrm{a,g},0}$ and $v_{\mathrm{a,g},0}$, for example, if $\phi_0 = \cdots = \phi_K$ and $\psi_0 = \cdots = \psi_K$ for the latter case.

Remark 12.16 (Long-term projections). Given $a \in \{1, ..., A\}$ and $g \in \{\mathrm{f}, \mathrm{m}\}$, long-term projections of death probabilities using (12.13) give, for all $\mathrm{a} \in 1, ..., A$ and $\mathrm{g} \in \{\mathrm{f}, \mathrm{m}\}$,

$$\lim_{t\to\infty} q_{a,g}(t) = F^{\mathrm{Lap}}\Big(\alpha_{a,g} + \beta_{a,g}\,\frac{\pi}{2\,\eta_{a,g}}\Big)\,.$$

Likewise, long-term projections for weights using (12.14) are given by

$$\lim_{t\to\infty} w_{a,g,k}(t) = \frac{\exp\big(u_{a,g,k} + v_{a,g,k}\,\frac{\pi}{2\psi_k}\big)}{\sum_{j=0}^{K}\exp\big(u_{a,g,j} + v_{a,g,j}\,\frac{\pi}{2\psi_k}\big)}\,.$$

i.e., weights are peaked in death causes with highest trends. Also, alternative families for weights can be considered as outlined in Remark 12.17.

Remark 12.17 (Alternative families for weights). Given Assumption 12.12, let $a \in \{1, \ldots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$, $k \in \{0, \ldots, K\}$ and $t \in \{1, \ldots, T\}$. Instead of using (12.14), weights can be defined via

$$w_{a,g,k}(t) = \frac{F^{\mathrm{Lap}}\big(u_{a,g,k} + v_{a,g,k}\,\mathcal{T}_{\phi_k,\psi_k}(t)\big)}{\sum_{j=0}^{K} F^{\mathrm{Lap}}\big(u_{a,g,j} + v_{a,g,j}\,\mathcal{T}_{\phi_j,\psi_j}(t)\big)}\,.$$

The great advantage of this family is that long-term forecasts are approximatively (modulo trend reduction) equally weighted amongst death causes with positive trend. Thus, we obtain some long-term equilibrium. But, when it comes to estimation, this family can produce messy results as weights are often not uniquely determined since $F^{\mathrm{Lap}}(x + c) = \exp(c)\,F^{\mathrm{Lap}}(x)$ for all $x, c \leq 0$. Thus, alternative families for weights can be considered as briefly outlined in Remark 12.17. Another possibility is to use a (quasi) linear family of weights

$$w_{a,g,k}(t) = \frac{u_{a,g,k} + v_{a,g,k}\,\mathcal{T}_{\phi_k,\psi_k}(t)}{\sum_{j=0}^{K} u_{a,g,j} + v_{a,g,j}\,\mathcal{T}_{\phi_j,\psi_j}(t)}$$

where we have to assume $2\,u_{a,g,j} \geq \pi\,|v_{a,g,j}|$ for all $j \in \{0, \ldots, K\}$ to make weights positive. This constraint usually leads to an underestimation of trends which is why we do not recommend this approach.

Remark 12.18 (High dimensionality). It should be mentioned that in our proposed setup we are confronted with a model based on more than 300 parameters. Therefore, deterministic numerical optimisation of a posteriori functions and likelihood functions is difficult and even for Markov chain Monte Carlo (MCMC) methods, see Section 12.4, it is hard to judge whether mixing of MCMC chains is sufficient. The latter problem can be tackled via running several MCMC chains for each parameter with different starting values and check whether all chains converge to the same stationary distribution. Depending on the purpose of the model, the number of parameters can be reduced. For further discussions on this topic, see Chapter 17.

12.1 Estimation via matching of moments

This approach is straight forward but needs a simplifying assumption to guarantee independent and identical random variables over time. We refer to it as *matching of moments approach.*

Assumption 12.19 (I.i.d. setup)**.** *Given the annuity model of Definition 11.11, as well as Assumption 12.3 and Definition 12.5, assume that death counts* $(N_{a,g,k}(t))_{t\in\{1,\dots,T\}}$ *are i.i.d., i.e., assume that* $m_{a,g} := m_{a,g}(1) = \cdots = m_{a,g}(T)$ *and* $q_{a,g} := q_{a,g}(1) = \cdots = q_{a,g}(T)$*, as well as* $w_{a,g,k} := w_{a,g,k}(1) = \cdots = w_{a,g,k}(T)$ *for every* $a \in \{1,\dots,A\}$*,* $g \in \{\mathrm{f},\mathrm{m}\}$ *and* $k \in \{0,\dots,K\}$*.*

To approximately achieve such an i.i.d. setting, we suggest to transform death counts $N_{a,g,k}(t)$ such that $\mathbb{E}[N_{a,g,k}(1)] = \cdots = \mathbb{E}[N_{a,g,k}(T)]$ for all $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$ and $k \in \{0,\dots,K\}$ as outlined in the following remark.

Remark 12.20 (Modification of given data). Given the annuity model from Definition 11.11 and Assumption 12.3, modify the number of deaths $n_{a,g,k}(t)$, the total number of people $m_{a,g}(t)$, death probabilities $q_{a,g}(t)$ and weights $w_{a,g,k}(t)$ such that Assumption 12.19 is approximatively met for each age category $a \in \{1,\dots,A\}$, gender $g \in \{\mathrm{f},\mathrm{m}\}$, death cause $k \in \{0,\dots,K\}$ and year $t \in \{1,\dots,T\}$ as follows:

$$n'_{a,g,k}(t) := \left\lfloor \frac{m_{a,g}(T)\,q_{a,g}(T)\,w_{a,g,k}(T)}{m_{a,g}(t)\,q_{a,g}(t)\,w_{a,g,k}(t)}\,n_{a,g,k}(t)\right\rfloor, \quad t \in \{1,\dots,T\}\,,$$

and, correspondingly,

$$m_{a,g} := \frac{m_{a,g}(T)}{m_{a,g}(t)}\,m_{a,g}(t) = m_{a,g}(T)\,, \quad t \in \{1,\dots,T\}\,,$$

as well as

$$q_{a,g} := \frac{q_{a,g}(T)}{q_{a,g}(t)}\,q_{a,g}(t) = q_{a,g}(T)\,, \quad t \in \{1,\dots,T\}\,,$$

and

$$w_{a,g,k} := \frac{w_{a,g,k}(T)}{w_{a,g,k}(t)}\,w_{a,g,k}(t) = w_{a,g,k}(T)\,, \quad t \in \{1,\dots,T\}\,.$$

Remark 12.21. Using the modification of Remark 12.20, we manage to remove long term trends in mortality and therefore erase variability in the data which is not driven by stochastic events. Furthermore, we manage to keep $m_{a,g}(t)$, $q_{a,g}(t)$ and $w_{a,g,k}(t)$ constant over time such that it is legitimate to assume an i.i.d. setting for transformed data in the sense of Assumption 12.19. Time indices may then be dropped. In particular, this data modification will be used for model validation in Section 17.

To be able to modify data as described above, we have to estimate death probabilities and weights a priori. This can be done as follows:

Remark 12.22 (Estimation of death probabilities). Given Assumption 12.12, as well as recalling Remark 12.6, for $a \in \{1,\dots,A\}$ and $g \in \{\mathrm{f},\mathrm{m}\}$ we may derive estimates

$$\big(\hat{q}^{MM}_{a,g}(t)\big)_{t\in\{1,\dots,T\}} = \Big(F^{\mathrm{Lap}}\big(\hat{\alpha}^{\mathrm{MM}}_{a,g} + \hat{\beta}^{\mathrm{MM}}_{a,g}\,\mathcal{T}_{\hat{\zeta}^{\mathrm{MM}}_{a,g},\hat{\eta}^{\mathrm{MM}}_{a,g}}(t)\big)\Big)_{t\in\{1,\dots,T\}}$$

for death probabilities $(q_{a,g}(t))_{t\in\{1,\dots,T\}}$ via minimising the mean squared error, i.e.,

$$\arg\inf_{\alpha_{a,g},\beta_{a,g},\zeta_{a,g},\eta_{a,g}} \sum_{t=1}^{T} \left(\frac{\sum_{k=0}^{K} n_{a,g,k}(t)}{m_{a,g}(t)} - F^{\mathrm{Lap}}\big(\alpha_{a,g} + \beta_{a,g}\,\mathcal{T}_{\zeta_{a,g},\eta_{a,g}}(t)\big) \right)^2 .$$

If parameters ζ and η are previously fixed, this result can be obtained by simply regressing

$$\left((F^{\mathrm{Lap}})^{-1} \left(\frac{\sum_{k=0}^{K} n_{a,g,k}(t)}{m_{a,g}(t)} \right) \right)_{t\in\{1,\dots,T\}}$$

on $(\mathcal{T}_{\zeta_{a,g},\eta_{a,g}}(t))_{t\in\{1,\dots,T\}}$. Rougher estimates for $\alpha_{a,g}$ and $\beta_{a,g}$ can always be derived by using linear regression on logarithmic death rates, see Remarks 12.15(d). Alternatively, we can use logistic regression which implies that death probabilities take the form

$$\log \frac{q_{a,g}(t)}{1-q_{a,g}(t)} = \alpha_{a,g} + t\beta_{a,g}\,, \quad t \in \{1,\dots,T\}\,.$$

In that case, we loose the link to Lee–Carter models.

Remark 12.23 (Estimation of weights). Given Assumption 12.12 and Remark 12.6, as well as Remark 12.22, we may derive estimates $(\hat{u}^{\mathrm{MM}}_{a,g,k}, \hat{v}^{\mathrm{MM}}_{a,g,k}, \hat{\phi}^{\mathrm{MM}}_{k}, \hat{\psi}^{\mathrm{MM}}_{k})_{t\in\{1,\dots,T\}}$ for parameters $(u_{a,g,k}, v_{a,g,k}, \phi_k, \psi_k)_{t\in\{1,\dots,T\}}$ for all $a \in \{1,\dots,A\}$ and $g \in \{\mathrm{f},\mathrm{m}\}$ via minimising the mean squared error to death rates, i.e.,

$$\arg\inf_{u_{a,g},v_{a,g},\phi_k,\psi_k} \sum_{t=1}^{T} \left(\frac{n_{a,g,k}(t)}{m_{a,g}(t)\,\hat{q}^{\mathrm{MM}}_{a,g}(t)} - \exp\big(u_{a,g} + v_{a,g}\,\mathcal{T}_{\phi_k,\psi_k}(t)\big) \right)^2 ,$$

for all age categories $a \in \{1,\dots,A\}$, genders $g \in \{\mathrm{f},\mathrm{m}\}$ and $k \in \{0,\dots,K\}$. Again, if parameters ϕ and ψ are previously fixed, this can be obtained by simply regressing

$$\left(\log \frac{n_{a,g,k}(t)}{m_{a,g}(t)\,\hat{q}^{\mathrm{MM}}_{a,g}(t)} \right)_{t\in\{1,\dots,T\}}$$

on $(\mathcal{T}_{\phi_k,\psi_k}(t))_{t\in\{1,\dots,T\}}$. Estimates $(\hat{w}^{\mathrm{MM}}_{a,g,k}(t))_{t\in\{1,\dots,T\}}$ are then given by (12.14).[47] Note that, while using regression techniques, we always have to check carefully if necessary assumptions such as constant variances of residuals are satisfied. Otherwise, we can switch to other generalised linear regression models or weighted least squares, depending on the data.

Once death probabilities and weights, as well as trends have been estimated such that Assumption 12.19 is satisfied (approximatively) via modifications suggested in Remark 12.20, risk factor variances may be estimated.

Lemma 12.24. *Given Assumptions 12.1, 12.5 and 12.19, for each age $a \in \{1,\dots,A\}$, gender $g \in \{\mathrm{f},\mathrm{m}\}$, death cause $k \in \{0,\dots,K\}$ and time $t \in \{1,\dots,T\}$, define*

$$W^*_{a,g,k}(t) := \frac{N_{a,g,k}(t)}{m_{a,g}\,q_{a,g}}\,,$$

[47] These are rough estimate but, as

$$\sum_{j=0}^{K} \exp\big(\hat{u}^{\mathrm{MM}}_{a,g,k} + \hat{v}^{\mathrm{MM}}_{a,g,k}\,\mathcal{T}_{\phi^{\mathrm{MM}}_k,\eta^{\mathrm{MM}}_k}(t)\big)$$

is usually close to one, they provide suitable starting values for the more sophisticated approaches below.

as well as

$$\overline{W}^*_{a,g,k} := \frac{1}{T}\sum_{t=1}^{T} W^*_{a,g,k}(t)\,.$$

Then, $\mathbb{E}\big[\overline{W}^*_{a,g,k}\big] = \mathbb{E}\big[W^*_{a,g,k}(t)\big] = w_{a,g,k}$*, i.e.,* $\overline{W}^*_{a,g,k}$ *and* $W^*_{a,g,k}(t)$ *are unbiased estimators for* $w_{a,g,k}$*.*

Proof. Since $n_{a,g,k}(t)$ is a realisation of $\sum_{i\in M_{a,g}(t)} N_{i,k}(t)$, we have

$$\mathbb{E}\big[\overline{W}^*_{a,g,k}\big] = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i\in M_{a,g}}\mathbb{E}[N_{i,k}(t)]}{m_{a,g}\,q_{a,g}}\,,$$

for each $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$, $c_k \in \{c_0,\dots,c_K\}$ and $t \in \{1,\dots,T\}$. Thus, since $\mathbb{E}[N_{i,k}(t)] = q_{a,g}\,w_{a,g,k}$ for every representative policyholder i of the specified category, the result follows. □

Lemma 12.25. *Given Assumptions 12.1, 12.5 and 12.19, define the estimator for the variance of* $W^*_{a,g,k}(t)$ *as*

$$\widehat{\Sigma}^2_{a,g,k} = \frac{1}{T-1}\sum_{t=1}^{T}\big(W^*_{a,g,k}(t) - \overline{W}^*_{a,g,k}\big)^2\,, \tag{12.26}$$

for all age categories $a \in \{1,\dots,A\}$*, genders* $g \in \{\mathrm{f},\mathrm{m}\}$ *and death causes* $k \in \{0,\dots,K\}$*. Then, recalling Assumption 11.11, we have*

$$\mathbb{E}\big[\widehat{\Sigma}^2_{a,g,k}\big] = \mathrm{Var}\big(W^*_{a,g,k}(t)\big) = \frac{w_{a,g,k}}{m_{a,g}\,q_{a,g}} + \sigma_k^2 w^2_{a,g,k}\,. \tag{12.27}$$

Proof. For notational convenience and without loss of generality we omit time parameters in all random variables in this proof as we have an i.i.d. setting. Also, fix $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$ and $k \in \{1,\dots,K\}$. Note that $(W^*_{a,g,k}(t))_{t\in\{1,\dots,T\}}$ is an i.i.d. sequence. Thus, since $\widehat{\Sigma}_{a,g,k}$ is an unbiased estimator for the standard deviation of $W^*_{a,g,k}(t)$ and $\overline{W}^*_{a,g,k}$, see Lehmann and Romano [79, Example 11.2.6], we immediately get

$$\mathbb{E}\big[\widehat{\Sigma}^2_{a,g,k}\big] = \mathrm{Var}\big(\overline{W}^*_{a,g,k}\big) = \mathrm{Var}\bigg(\frac{1}{m_{a,g}\,q_{a,g}}\sum_{i\in M_{a,g}} N_{i,k}\bigg)\,.$$

Using the law of total variance as in [111, Lemma 3.48] together with Definition 11.11(c) gives

$$\begin{aligned} m^2_{a,g}q^2_{a,g}\,\mathbb{E}\big[\widehat{\Sigma}^2_{a,g,k}\big] &= \mathbb{E}\bigg[\mathrm{Var}\bigg(\sum_{i\in M_{a,g}} N_{i,k}\,\bigg|\,\Lambda_k\bigg)\bigg] + \mathrm{Var}\bigg(\mathbb{E}\bigg[\sum_{i\in M_{a,g}} N_{i,k}\,\bigg|\,\Lambda_k\bigg]\bigg) \\ &= \sum_{i\in M_{a,g}}\mathbb{E}[\mathrm{Var}(N_{i,k}\,|\,\Lambda_k)] + \mathrm{Var}\bigg(\sum_{i\in M_{a,g}}\mathbb{E}[N_{i,k}\,|\,\Lambda_k]\bigg)\,. \end{aligned}$$

Since $\mathrm{Var}(N_{i,k}|\Lambda_k) = \mathbb{E}[N_{i,k}\,|\,\Lambda_k] = q_{a,g}\,w_{a,g,k}\,\Lambda_k$ a.s. for all representative policyholders $i \in M_{a,g}$ with $|M_{a,g}| = m_{a,g}$, the equation above simplifies to

$$\mathbb{E}\big[\widehat{\Sigma}^2_{a,g,k}\big] = \frac{w_{a,g,k}}{m_{a,g}\,q_{a,g}} + w^2_{a,g,k}\,\mathrm{Var}(\Lambda_k)\,,$$

which gives the result. □

Remark 12.28. Having obtained (12.27) and recalling Assumption 11.11, we get, for all $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $k \in \{0, \dots, K\}$,

$$\mathbb{E}\Big[\widehat{\Sigma}^2_{a,g,k} - \frac{w_{a,g,k}}{m_{a,g}\, q_{a,g}}\Big] = \sigma_k^2 w^2_{a,g,k}$$

and, thus, summing up over all age categories and genders gives

$$\mathbb{E}\left[\frac{\sum_{a=1}^{A}\sum_{g\in\{\mathrm{f},\mathrm{m}\}}\big(\widehat{\Sigma}^2_{a,g,k} - \frac{w_{a,g,k}}{m_{a,g}\, q_{a,g}}\big)}{\sum_{a=1}^{A}\sum_{g\in\{\mathrm{f},\mathrm{m}\}} w^2_{a,g,k}}\right] = \sigma_k^2\,, \quad k \in \{1, \dots, K\}\,. \tag{12.29}$$

Replacing $q_{a,g}$ and $w_{a,g,k}$ by their estimates $\hat{q}^{\mathrm{MM}}_{a,g}(T)$ and $\hat{w}^{\mathrm{MM}}_{a,g,k}(T)$ in Equation (12.29), see Remarks 12.22 and 12.23, we may define the following matching of moments estimates for risk factor variances.

Definition 12.30 (Estimates for risk factor variances)**.** Given Assumptions 12.1, 12.5 and 12.19 as well as Remarks 12.23 and 12.22, the *matching of moments estimate* for σ_k for all $k \in \{1, \dots, K\}$ is defined as

$$\hat{\sigma}_k^{\mathrm{MM}} := \sqrt{\max\left\{0, \frac{\sum_{a=1}^{A}\sum_{g\in\{\mathrm{f},\mathrm{m}\}}\left(\hat{\sigma}^2_{a,g,k} - \frac{w^{\mathrm{MM}}_{a,g,k}(T)}{m_{a,g}\, q^{\mathrm{MM}}_{a,g}(T)}\right)}{\sum_{a=1}^{A}\sum_{g\in\{\mathrm{f},\mathrm{m}\}}(w^{\mathrm{MM}}_{a,g,k}(T))^2}\right\}}\,, \tag{12.31}$$

where $\hat{\sigma}^2_{a,g,k}$ is the estimate corresponding to estimator $\widehat{\Sigma}^2_{a,g,k}$.

Remark 12.32. $\hat{\sigma}_k^{\mathrm{MM}}$ can equal zero and therefore may not detect variation in data properly. With a similar argumentation as for (12.29), we could define an alternative matching of moments estimator using

$$\mathbb{E}\left[\frac{1}{2A}\sum_{a=1}^{A}\sum_{g\in\{\mathrm{f},\mathrm{m}\}}\frac{\widehat{\Sigma}^2_{a,g,k} - \frac{w_{a,g,k}}{m_{a,g}\, q_{a,g}}}{w^2_{a,g,k}}\right] = \sigma_k^2\,, \quad k \in \{1, \dots, K\}\,.$$

The problem of this definition is that for categories with few observations of deaths summands can become very large. In particular, if weights are zero, then fractions may not even be defined.

12.2 Estimation via a maximum a posteriori approach

While the matching of moments approach requires several modifications of the data to gain constant weights and death probabilities, the approach in this section does not require any of these. It is a variation of *maximum a posteriori estimation* based on Bayesian inference. For an introduction to Bayesian inference see, for example, Shevchenko [114, Section 2.9]. In particular, Definition 11.11(c) will be of great importance. One main advantage of this approach is the fact that we obtain estimates for risk factor realisations which is very useful for scenario analysis, see Chapter 14. Also, handy approximations for estimates of risk factor realisations and variances are obtained in this section. The basic idea is to express the joint posterior distribution of all parameters via conditional distributions.

Lemma 12.33 (Posterior density)**.** *Given Assumptions 12.1 and 12.5, as well as 12.12, consider parameters* $\theta_q := (\alpha, \beta, \zeta, \eta) \in E$*,* $\theta_w := (u, v, \phi, \psi) \in F$*, risk factor realisations* $\lambda := (\lambda_k(t)) \in (0,\infty)^{K\times T}$ *of* $\Lambda := (\Lambda_k(t)) \in (0,\infty)^{K\times T}$ *and* $\sigma := (\sigma_k) \in [0,\infty)^K$*, as well as data* $n := (n_{a,g,k}(t)) \in \mathbb{N}_0^{A\times 2\times (K+1)\times T}$*. Assume that parameters are independent so that their prior distribution may be written as*[48]

$$\pi(\theta_q, \theta_w, \sigma) := 1_E(\theta_q)\, 1_F(\theta_w)\, 1_{(0,\infty)^K}(\sigma)\,. \tag{12.34}$$

Then, the posterior density $\pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n)$ *of parameters given data* $N = n$ *is up to constant given by*[49]

$$\begin{aligned}
\pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n) &\propto \pi(\theta_q, \theta_w, \sigma)\pi(\lambda \,|\, \theta_q, \theta_w, \sigma)\ell(n \,|\, \theta_q, \theta_w, \lambda, \sigma) \\
&= \prod_{t=1}^{T}\Bigg(\Bigg(\prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}} \frac{e^{-\rho_{a,g,0}(t)}\,\rho_{a,g,0}(t)^{n_{a,g,0}(t)}}{n_{a,g,0}(t)!}\Bigg)\prod_{k=1}^{K}\Bigg(\frac{e^{-\lambda_k(t)/\sigma_k^2}\,\lambda_k(t)^{1/\sigma_k^2-1}}{\Gamma(1/\sigma_k^2)\,(\sigma_k^2)^{1/\sigma_k^2}} \\
&\qquad \times \prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}} \frac{e^{-\rho_{a,g,k}(t)\lambda_k(t)}\,(\rho_{a,g,k}(t)\,\lambda_k(t))^{n_{a,g,k}(t)}}{n_{a,g,k}(t)!}\Bigg)\Bigg)\pi(\theta_q, \theta_w, \sigma)\,,
\end{aligned} \tag{12.35}$$

where $\pi(\lambda \,|\, \theta_q, \theta_w, \sigma)$ *denotes the prior density of risk factors at* $\Lambda = \lambda$ *given all other parameters, where* $\ell(n \,|\, \theta_q, \theta_w, \lambda, \sigma)$ *denotes the likelihood of* $N = n$ *given all parameters and where we have* $\rho_{a,g,k}(t) = m_{a,g}(t)\,q_{a,g}(t)\,w_{a,g,k}(t)$ *for all* $a \in \{1,\dots,A\}$*,* $g \in \{\mathrm{f},\mathrm{m}\}$*,* $k \in \{1,\dots,K\}$ *and* $t \in \{1,\dots,T\}$*.*

Proof. The first proportional equality follows by Bayes' theorem which is also widely used in Bayesian inference, see, for example, Shevchenko [114, Section 2.9]. Due to independence amongst risk factors and since they are gamma distributed with mean one and variances σ^2, we have

$$\pi(\lambda \,|\, \theta_q, \theta_w, \sigma) = \prod_{k=1}^{K}\prod_{t=1}^{T}\left(\frac{e^{-\lambda_k(t)/\sigma_k^2}\,\lambda_k(t)^{1/\sigma_k^2-1}}{\Gamma(1/\sigma_k^2)\,(\sigma_k^2)^{1/\sigma_k^2}}\right).$$

If $\theta_q \in E$, $\theta_w \in F$, $\lambda \in (0,\infty)^{K\times T}$ and $\sigma \in [0,\infty)^K$, then note that due to Definition 11.11, as well as Assumption 12.1 we have

$$\begin{aligned}
\ell(n \,|\, \theta_q, \theta_w, \lambda, \sigma) &= \mathbb{P}\Bigg(\bigcap_{a=1}^{A}\bigcap_{g\in\{\mathrm{f},\mathrm{m}\}}\bigcap_{k=0}^{K}\bigcap_{t=1}^{T}\big\{N_{a,g,k}(t) = n_{a,g,k}(t)\big\}\,\Bigg|\,\Lambda = \lambda\Bigg) \\
&= \prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}}\prod_{t=1}^{T}\Bigg(e^{-\rho_{a,g,0}(t)}\,\frac{\rho_{a,g,0}(t)^{n_{a,g,0}(t)}}{n_{a,g,0}(t)!} \\
&\qquad \times \prod_{k=1}^{K}\mathbb{P}\big(N_{a,g,k}(t) = n_{a,g,k}(t)\,\big|\,\Lambda_k(t) = \lambda_k(t)\big)\Bigg)\,,
\end{aligned}$$

[48] Here we are confronted with a so-called *improper prior*, see Shevchenko [114, Section 2.9.5], since it is not a density with respect to the Lebesgue–Borel measure in the usual sense due to the infinite support of σ^2. This prior distribution does not carry any information about the parameters to be estimated and it corresponds to independent uniform distributions of all components with respective supports.

[49] The symbol '$\propto$' denotes proportionality almost everywhere, i.e., equality almost everywhere up to a multiplicative constant which is independent of the parameters, see Shevchenko [114, Theorem 2.3]. If we restrict to continuous densities, then we can drop almost everywhere.

which then gives (12.35) since, for all $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $k \in \{1, \dots, K\}$, as well as $t \in \{1, \dots, T\}$,

$$\begin{aligned}
&\mathbb{P}\big(N_{a,g,k}(t) = n_{a,g,k}(t) \,\big|\, \Lambda_k(t) = \lambda_k(t)\big) \\
&\quad = \mathbb{P}\Bigg(\sum_{i \in M_{a,g}(t)} N_{i,k}(t) = n_{a,g,k}(t) \,\Bigg|\, \Lambda_k(t) = \lambda_k(t) \Bigg) \\
&\quad = \exp\big(- m_{a,g}(t)\, q_{a,g}\, w_{a,g,k}\, \lambda_k(t) \big) \frac{\big(m_{a,g}(t)\, q_{a,g}\, w_{a,g,k}\, \lambda_k(t)\big)^{n_{a,g,k}(t)}}{n_{a,g,k}(t)!} \,,
\end{aligned}$$

where $i \in M_{a,g}(t)$ with $|M_{a,g}(t)| = m_{a,g}(t)$ are representatives of the specified category. □

Remark 12.36. (Maximum a posteriori approach)

(a) Notation for posterior and conditional densities is adapted to the notation used in the textbook of Shevchenko [114, Section 2.9].

(b) The approach described above may look like a pure Bayesian inference approach but note that risk factors $\Lambda_k(t)$ are truly stochastic and, therefore, we refer to it as a maximum a posteriori estimation approach.

(c) Consider the assumptions of Lemma 12.33. Since the products in (12.35) can become very small, it is recommended to use the logarithm of posterior densities which are denoted by $\log \pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n)$. For $n \in \mathbb{N}_0^{A \times 2 \times (K+1) \times T}$ they are given by

$$\begin{aligned}
\log \pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n) = \sum_{t=1}^{T} \Bigg(& \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f},\mathrm{m}\}} \big(n_{a,g,0}(t) \log \rho_{a,g,0}(t) - \rho_{a,g,0}(t) - \log(n_{a,g,0}(t)!) \big) \\
& + \sum_{k=1}^{K} \Bigg(- \log \Gamma\Big(\frac{1}{\sigma_k^2}\Big) - \frac{\log \sigma_k^2}{\sigma_k^2} - \frac{\lambda_k(t)}{\sigma_k^2} + \Big(\frac{1}{\sigma_k^2} - 1\Big) \log \lambda_k(t) \\
& + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f},\mathrm{m}\}} \big(n_{a,g,k}(t) \log\big(\rho_{a,g,k}(t)\, \lambda_k(t)\big) - \rho_{a,g,k}(t)\, \lambda_k(t) - \log(n_{a,g,k}(t)!) \big) \Bigg) \Bigg) \,,
\end{aligned} \tag{12.37}$$

if $\theta_q \in E$, $\theta_w \in F$, $\lambda \in (0, \infty)^{K \times T}$ and $\sigma \in [0, \infty)^K$. Otherwise, the logarithmic posterior density takes the value $-\infty$.

Having derived the posterior density, we can now define corresponding maximum a posteriori estimates.

Definition 12.38 (Maximum a posteriori estimates)**.** Recalling (12.35) and (12.37) as well as given the assumptions of Lemma 12.33, *maximum a posteriori estimates* for parameters $\theta_q, \theta_w, \lambda$ and σ, given uniqueness, are defined by

$$\begin{aligned}
\big(\hat{\theta}_q^{MAP}, \hat{\theta}_w^{\mathrm{MAP}}, \hat{\lambda}^{\mathrm{MAP}}, \hat{\sigma}^{\mathrm{MAP}}\big) &:= \operatorname*{arg\,sup}_{\theta_q, \theta_w, \lambda, \sigma} \pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n) \\
&= \operatorname*{arg\,sup}_{\theta_q, \theta_w, \lambda, \sigma} \log \pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n) \,.
\end{aligned} \tag{12.39}$$

Remarks 12.40. (Maximum a posteriori estimates).

(a) Using real world data, risk factor variances usually take small values less than 0.1. Thus, assuming an upper bound for these parameters is legitimate so that (12.34) becomes a proper density modulo a normalisation constant. Estimates for risk factor realisations λ should then take values close to one, except outliers. Death probabilities should be close to values derived by the Lee–Carter method and close to values in life tables. Weights $\hat{w}^{\mathrm{MAP}}$ should be close to $\hat{w}^{\mathrm{MM}}$.

(b) In general, there exists no closed form solution for maximum a posteriori estimates. Deterministic or stochastic numerical optimisation schemes have to be applied but, for suitable data, approximations exist, see Remark 12.47.

(c) Numerical issues. Deterministic optimisation in (12.39) may quickly lead to numerical issues due to high dimensionality and due to the flat surface of the function to be optimised which may yield to failure of some gradient methods. Adaption of convergence tolerance can lead to better results. In 'R' the optimisation routine nlminb, see [99], gives stable results in simple examples. But, also this procedure quickly breaks down in more involved settings. One alternative is to use Markov chain Monte Carlo as described in Section 12.4 or as in Shevchenko [114, Section 2.11]. Otherwise, we suggest to estimate weights and death probabilities as outlined in Section 12.1 a priori and then proceed with optimisation in (12.39) over σ and λ. Alternatively, Lemma 12.41 or Remark 12.47 can be used.

As maximum a posteriori estimates are hard to obtain, we can give some necessary characterisations of the solutions which can then be used as easy-to-calculate approximations.

Lemma 12.41 (Conditions for maximum a posteriori estimates)**.** *Given Definition 12.38, estimates $\hat{\lambda}^{\mathrm{MAP}}$ and $\hat{\sigma}^{\mathrm{MAP}}$ satisfy, for every $k \in \{1, \dots, K\}$ and $t \in \{1, \dots, T\}$,*

$$\hat{\lambda}_k^{\mathrm{MAP}}(t) = \frac{1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 - 1 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} n_{a,g,k}(t)}{1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} \rho_{a,g,k}(t)} \tag{12.42}$$

if $1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 - 1 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} n_{a,g,k}(t) > 0$, as well as

$$2 \log \hat{\sigma}_k^{\mathrm{MAP}} + \frac{\Gamma'\big(1/(\hat{\sigma}_k^{\mathrm{MAP}})^2\big)}{\Gamma\big(1/(\hat{\sigma}_k^{\mathrm{MAP}})^2\big)} = \frac{1}{T} \sum_{t=1}^{T} \big(1 + \log \hat{\lambda}_k^{\mathrm{MAP}}(t) - \hat{\lambda}_k^{\mathrm{MAP}}(t)\big)\,, \tag{12.43}$$

where, for given $\hat{\lambda}_k^{\mathrm{MAP}}(1), \dots, \hat{\lambda}_k^{\mathrm{MAP}}(T) > 0$, (12.43) has a unique solution which is strictly positive. In particular, for every $k \in \{1, \dots, K\}$,

$$\begin{aligned} 2 \log \hat{\sigma}_k^{\mathrm{MAP}} = \frac{1}{T} \sum_{t=1}^{T} \Bigg(& 1 + \log \frac{1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 - 1 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} n_{a,g,k}(t)}{1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} \rho_{a,g,k}(t)} \\ & - \frac{1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 - 1 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} n_{a,g,k}(t)}{1/(\hat{\sigma}_k^{\mathrm{MAP}})^2 + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} \rho_{a,g,k}(t)} \Bigg) - \frac{\Gamma'\big(1/(\hat{\sigma}_k^{\mathrm{MAP}})^2\big)}{\Gamma\big(1/(\hat{\sigma}_k^{\mathrm{MAP}})^2\big)}\,. \end{aligned} \tag{12.44}$$

Remark 12.45. The term $\Gamma'(x)/\Gamma(x)$, known as digamma function or ψ-function, is extensively discussed in the literature, see for example Chaudhry and Zubair [22] and Qi et al. [98], as well as the references therein.

Proof of Lemma 12.41. First, set

$$\pi^*(n) := \log \pi(\theta_q, \theta_w, \lambda, \sigma \,|\, n)\,.$$

Then, for every $k \in \{1, \dots, K\}$ and $t \in \{1, \dots, T\}$, differentiating (12.37) gives

$$\frac{\partial \pi^*(n)}{\partial \lambda_k(t)} = \frac{1/\sigma_k^2 - 1}{\lambda_k(t)} - \frac{1}{\sigma_k^2} + \sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} \Big(\frac{n_{a,g,k}(t)}{\lambda_k(t)} - \rho_{a,g,k}(t) \Big)\,.$$

Setting this term equal to zero and solving for $\Lambda_k(t)$ gives (12.42). Similarly, for every $k \in \{1, \dots, K\}$, we obtain

$$\frac{\partial \pi^*(n)}{\partial \sigma_k^2} = \frac{1}{\sigma_k^4} \sum_{t=1}^{T} \Big(\log \sigma_k^2 - 1 + \frac{\Gamma'(1/\sigma_k^2)}{\Gamma(1/\sigma_k^2)} - \log \lambda_k(t) + \lambda_k(t) \Big)\,.$$

Again, setting this term equal to zero and rearranging the terms gives (12.43).

For existence and uniqueness of a solution in (12.43), given $\hat{\lambda}_k^{\mathrm{MAP}}(1), \dots, \hat{\lambda}_k^{\mathrm{MAP}}(T) > 0$, let $k \in \{1, \dots, K\}$ and note that the right side in the equation is strictly negative unless $\hat{\lambda}_k^{\mathrm{MAP}}(1) = \dots = \hat{\lambda}_k^{\mathrm{MAP}}(T) = 1$ in which case there is no variability in the risk factor, i.e., $\sigma_k^2 = 0$. Then, note that

$$f(x) := \log x - \frac{\Gamma'(x)}{\Gamma(x)}\,, \quad x > 0\,,$$

is continuous and

$$\frac{1}{2x} < f(x) < \frac{1}{2x} + \frac{1}{12x^2}\,, \quad x > 0\,, \tag{12.46}$$

which follows by Qi et al. [98, Corollary 1] together with $f(x+1) = 1/x + f(x)$ for all $x > 0$. As we want to solve $-f(1/x) = -c$ for some given $c > 0$, note that $f(0+) = \infty$, as well as $\lim_{x\to\infty} f(x) = 0$. Thus a solution of Equation (12.43) has to exist for given $\hat{\lambda}_k^{\mathrm{MAP}}(1), \dots, \hat{\lambda}_k^{\mathrm{MAP}}(T) > 0$. Furthermore,

$$f'(x) = \frac{1}{x} - \sum_{i=0}^{\infty} \frac{1}{(x+i)^2} < \frac{1}{x} - \int_x^{\infty} \frac{1}{z^2}\, dz = 0\,, \quad x > 0\,,$$

where the first equality follows by Chaudhry and Zubair [22]. This implies that $-f(1/x)$ and $f(x)$ are strictly decreasing. Thus, the solution of (12.43) is unique. Equation (12.44) then follows by substituting (12.42) into (12.43). □

Remark 12.47 (Approximations). Given Definition 12.38, let weights and death probabilities, as well as risk factor variances be estimated a priori using, for example, matching of moments as given in Section 12.1, as well as Remark 12.22 and Remark 12.23. Then, (12.42) provides an approximation for risk factor realisations. Alternatively, we can use a rougher approach to derive approximative maximum a posteriori estimates for λ and σ. Based on (12.42), for all $k \in \{1, \dots, K\}$ and $t \in \{1, \dots, T\}$, if

$$\sum_{a=1}^{A} \sum_{g \in \{\mathrm{f,m}\}} n_{a,g,k}(t)\,, \quad k \in \{1, \dots, K\} \text{ and } t \in \{1, \dots, T\}\,,$$

is large, it is reasonable to define

$$\hat{\lambda}_k^{\mathrm{MAPappr}}(t) := \frac{-1 + \sum_{a=1}^{A} \sum_{g\in\{\mathrm{f,m}\}} n_{a,g,k}(t)}{\sum_{a=1}^{A} \sum_{g\in\{\mathrm{f,m}\}} \rho_{a,g,k}(t)} \tag{12.48}$$

as an approximative estimate for $\lambda_k(t)$ where $\rho_{a,g,k}(t) := m_{a,g}(t)\, q_{a,g}(t)\, w_{a,g,k}(t)$. In particular, this approximation is independent of estimates for σ. Having derived approximative estimates for λ, we can use (12.43) to get estimates for σ which exist and are unique. Alternatively, note that due to (12.46), we get

$$-2\log\hat{\sigma}_k^{\mathrm{MAP}} - \frac{\Gamma'\big(1/(\hat{\sigma}_k^{\mathrm{MAP}})^2\big)}{\Gamma\big(1/(\hat{\sigma}_k^{\mathrm{MAP}})^2\big)} = \frac{(\hat{\sigma}_k^{\mathrm{MAP}})^2}{2} + \mathcal{O}\big((\hat{\sigma}_k^{\mathrm{MAP}})^4\big)\,, \quad k\in\{1,\dots,K\}\,.$$

Furthermore, if we use second order Taylor expansion for the logarithm, then the right hand side of (12.43) gets, for all $k \in \{1,\dots,K\}$,

$$\frac{1}{T}\sum_{t=1}^{T}\big(\hat{\lambda}_k^{\mathrm{MAP}}(t) - 1 - \log\hat{\lambda}_k^{\mathrm{MAP}}(t)\big) = \frac{1}{2T}\sum_{t=1}^{T}\Big(\big(\hat{\lambda}_k^{\mathrm{MAP}}(t)-1\big)^2 + \mathcal{O}\big(\big(\hat{\lambda}_k^{\mathrm{MAP}}(t)-1\big)^3\big)\Big)\,.$$

This approximation is better the closer the values of λ are to one. Thus, using these observations, an approximation for risk factor variances σ^2 is given by

$$\hat{\sigma}_k^{\mathrm{MAPappr}} := \sqrt{\frac{1}{T}\sum_{t=1}^{T}\big(\hat{\lambda}_k^{\mathrm{MAPappr}}(t)-1\big)^2}\,, \quad k\in\{1,\dots,K\}\,, \tag{12.49}$$

which is simply the sample variance of $\hat{\lambda}^{\mathrm{MAP}}$. Note that this estimate would be an intuitive guess for estimating the variance of risk factors given realisations.

Remark 12.50 (An easy but accurate approach). We have two possibilities to avoid optimisation of the maximum a posteriori function in (12.39). In both cases, we estimate death probabilities and weights, i.e., parameters θ_q and θ_w, a priori via matching of moments. Then, we can use Equation (12.44) to find estimates for risk factor variances σ^2 which then yield estimates for risk factor realisations λ via Equation (12.42). Note that, in general, Equation (12.44) does not have a unique solution as the function is oscillating around zero as $\sigma_k \searrow 0$. The second possibility to estimate λ and σ^2 is to use approximations in (12.48) and (12.49). Note that $|\hat{\lambda}_k^{\mathrm{MAP}}(t)-1| < |\hat{\lambda}_k^{\mathrm{MAPappr}}(t)-1|$ for all $k\in\{1,\dots,K\}$ and $t\in\{1,\dots,T\}$, implying that (12.49) will dominate solutions obtained by (12.43) in most cases.

12.3 Estimation via maximum likelihood

Thirdly, we propose a classical estimation approach following *maximum likelihood*. Maximum likelihood estimation immediately guarantees nice asymptotic properties of estimators under mild regularity conditions. Unfortunately, similarly as for the maximum a posteriori approach, estimates are not given explicitly and deterministic numerical optimisation easily breaks down due to high dimensionality.

Lemma 12.51 (Likelihood function)**.** *Given Assumptions 12.1, 12.5 and 12.12, define*

$$n_k(t) := \sum_{a=1}^{A} \sum_{g\in\{\mathrm{f},\mathrm{m}\}} n_{a,g,k}(t)\,, \quad k \in \{0,\dots,K\} \text{ and } t \in \{1,\dots,T\}\,,$$

as well as $\rho_{a,g,k}(t) := m_{a,g}(t)\,q_{a,g}(t)\,w_{a,g,k}(t)$ *for all* $a \in \{1,\dots,A\}$ *and* $g \in \{\mathrm{f},\mathrm{m}\}$ *and*

$$\rho_k(t) := \sum_{a=1}^{A} \sum_{g\in\{\mathrm{f},\mathrm{m}\}} \rho_{a,g,k}(t)\,.$$

Then, the likelihood function $\ell(n\,|\,\theta_q,\theta_w,\sigma)$ *of parameters* $\theta_q := (\alpha,\beta,\zeta,\eta) \in E$*, as well as* $\theta_w := (u,v,\phi,\psi) \in F$ *and* $\sigma := (\sigma_k) \in [0,\infty)^K$ *given data* $n := (n_{a,g,k}(t)) \in \mathbb{N}_0^{A\times 2\times(K+1)\times T}$ *is given by*

$$\begin{aligned}\ell(n\,|\,\theta_q,\theta_w,\sigma) = \prod_{t=1}^{T}\Bigg(\Bigg(\prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}} \frac{e^{-\rho_{a,g,0}(t)}\,\rho_{a,g,0}(t)^{n_{a,g,0}(t)}}{n_{a,g,k}(t)!}\Bigg)\\ \times \prod_{k=1}^{K}\Bigg(\frac{\Gamma(1/\sigma_k^2+n_k(t))}{\Gamma(1/\sigma_k^2)\,(\sigma_k^2)^{1/\sigma_k^2}\,(1/\sigma_k^2+\rho_k(t))^{1/\sigma_k^2+n_k(t)}}\prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}}\frac{\rho_{a,g,k}(t)^{n_{a,g,k}(t)}}{n_{a,g,k}(t)!}\Bigg)\Bigg)\,.\end{aligned} \tag{12.52}$$

Proof. Analogously to the derivation of (12.35), we get

$$\begin{aligned}\ell(n\,|\,\theta_q,\theta_w,\sigma) = \mathbb{P}(N=n\,|\,\theta_q,\theta_w,\sigma) = \prod_{t=1}^{T}\Bigg(\Bigg(\prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}}\frac{e^{-\rho_{a,g,0}(t)}\,\rho_{a,g,0}(t)^{n_{a,g,0}(t)}}{n_{a,g,0}(t)!}\Bigg)\\ \times\prod_{k=1}^{K}\mathbb{E}\Bigg[\mathbb{P}\Bigg(\bigcap_{a=1}^{A}\bigcap_{g\in\{\mathrm{f},\mathrm{m}\}}\big\{N_{a,g,k}(t)=n_{a,g,k}(t)\big\}\,\Bigg|\,\Lambda_k(t)\Bigg)\Bigg]\Bigg)\,,\end{aligned}$$

where $\mathbb{P}(N=n\,|\,\theta_q,\theta_w,\sigma)$ denotes the probability of the event $\{N=n\}$ given parameters. Note that this expression is not a conditional probability per se. Then, for all $k \in \{1,\dots,K\}$ and $t \in \{1,\dots,T\}$, $\Lambda_k(t)$ is gamma distributed with mean one and variance σ_k^2. Therefore, taking expectations in the equation above gives

$$\begin{aligned}&\mathbb{E}\Bigg[\mathbb{P}\Bigg(\bigcap_{a=1}^{A}\bigcap_{g\in\{\mathrm{f},\mathrm{m}\}}\big\{N_{a,g,k}(t)=n_{a,g,k}(t)\big\}\,\Bigg|\,\Lambda_k(t)\Bigg)\Bigg]\\ &= \mathbb{E}\Bigg[e^{-\rho_k(t)\,\Lambda_k(t)}\prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}}\frac{(\rho_{a,g,k}(t)\,\Lambda_k(t))^{n_{a,g,k}(t)}}{n_{a,g,k}(t)!}\Bigg]\\ &= \Bigg(\prod_{a=1}^{A}\prod_{g\in\{\mathrm{f},\mathrm{m}\}}\frac{\rho_{a,g,k}(t)^{n_{a,g,k}(t)}}{n_{a,g,k}(t)!}\Bigg)\int_0^\infty e^{-\rho_k(t)\,x_t}\,x_t^{n_k(t)}\,\frac{x_t^{1/\sigma_k^2-1}\,e^{-x_t/\sigma_k^2}}{\Gamma(1/\sigma_k^2)\,(\sigma_k^2)^{1/\sigma_k^2}}\,dx_t\,.\end{aligned}$$

The integrand above is a density of a gamma distribution—modulo the normalisation constant—with parameters $1/\sigma_k^2+n_k(t)$ and $1/\sigma_k^2+\rho_k(t)$. Therefore, the corresponding integral equals the multiplicative inverse of the normalisation constant, i.e.,

$$\left(\frac{(1/\sigma_k^2+\rho_k(t))^{1/\sigma_k^2+n_k(t)}}{\Gamma(1/\sigma_k^2+n_k(t))}\right)^{-1}, \quad k \in \{1,\dots,K\} \text{ and } t \in \{1,\dots,T\}\,.$$

Putting all results together gives (12.52). □

Since the products in (12.52) can become very small, we recommend to use the log-likelihood function instead which is given in the following remark.

Remark 12.53 (Log-likelihood function). The log-likelihood function $\log \ell(n\,|\,\theta_q, \theta_w, \sigma)$ is, for $n \in \mathbb{N}_0^{A\times 2\times (K+1)\times T}$, given by

$$\begin{aligned}\log \ell(n\,|\,\theta_q, \theta_w, \sigma) = \sum_{t=1}^{T} \Bigg(& \sum_{a=1}^{A} \sum_{g\in\{\mathrm{f,m}\}} \big(n_{a,g,0}(t) \log \rho_{a,g,0}(t) - \rho_{a,g,0}(t) - \log(n_{a,g,0}(t)!)\big) \\ & + \sum_{k=1}^{K} \Bigg(\log \frac{\Gamma(1/\sigma_k^2 + n_k(t))}{\Gamma(1/\sigma_k^2)} - \frac{\log \sigma_k^2}{\sigma_k^2} - \Big(\frac{1}{\sigma_k^2} + n_k(t)\Big) \log \Big(\frac{1}{\sigma_k^2} + \rho_k(t)\Big) \\ & + \sum_{a=1}^{A} \sum_{g\in\{\mathrm{f,m}\}} \big(n_{a,g,k}(t) \log \rho_{a,g,k}(t) - \log(n_{a,g,k}(t)!)\big)\Bigg)\Bigg),\end{aligned} \tag{12.54}$$

if $\theta_q \in E$ and $\theta_w \in F$, as well as $\sigma \in [0,\infty)^K$. Otherwise, the log-likelihood function takes the value $-\infty$. For implementations we recommend to write the first term in the third row as

$$\log \frac{\Gamma(1/\sigma_k^2 + n_k(t))}{\Gamma(1/\sigma_k^2)} = \log \Gamma\Big(\frac{1}{\sigma_k^2} + n_k(t)\Big) - \log \Gamma\Big(\frac{1}{\sigma_k^2}\Big)$$

and to use the log-gamma function, e.g., the lgamma function in 'R' see [99], as $\Gamma(1/\sigma_k^2 + n_k(t))$ may lead to overflow errors. Alternatively, the identity $\Gamma(x+n)/\Gamma(x) = \prod_{j=1}^{n}(x+j-1)$ for all $n \in \mathbb{N}_0$, as well as $x > 0$ can be used to obtain

$$\log \frac{\Gamma(1/\sigma_k^2 + n_k(t))}{\Gamma(1/\sigma_k^2)} = \sum_{j=1}^{n_k(t)} \log \Big(\frac{1}{\sigma_k^2} + j - 1\Big).$$

Definition 12.55 (Maximum likelihood estimates)**.** Recalling (12.52) and (12.54), as well as given the assumptions of Lemma 12.51, *maximum likelihood estimates* for parameters θ_q, θ_w and σ, given uniqueness, are defined by

$$\begin{aligned}\big(\hat\theta_q^{MLE}, \hat\theta_w^{\mathrm{MLE}}, \hat\sigma^{\mathrm{MLE}}\big) &:= \sup_{\theta_q,\theta_w,\sigma} \ell(n\,|\,\theta_q, \theta_w, \sigma) \\ &= \sup_{\theta_q,\theta_w,\sigma} \log \ell(n\,|\,\theta_q, \theta_w, \sigma)\,.\end{aligned} \tag{12.56}$$

Remark 12.57 (Numerical issues). In many examples, maximum likelihood estimates are unique but numerical optimisation is needed to finally derive them. However, numerical issues can occur as outlined in Remarks 12.40(c). Switching to a Bayesian setting, Markov chain Monte Carlo can be used to derive estimates with stochastic numerical optimisation, see Section 12.4.

Remark 12.58 (Setup embedding and asymptotic variance). Given Definition 12.55, assume that we have a priori estimated death probabilities and weights such that a transformation as suggested in Remark 12.23 leads to an i.i.d. setting. Moreover, let $k \in \{1,\dots,K\}$ be fixed. Using a suitable embedding, we can identify the random vectors $(N_{a,g,k}(t))_{a\in\{1,\dots,A\},g\in\{\mathrm{f,m}\}}$ for all $t \in \{1,\dots,T\}$ with a one-dimensional random variable and can therefore assume that we are confronted with a classical i.i.d. maximum likelihood setting. Then, the estimator

$\widehat{\Sigma}_k^{\mathrm{MLE}}$ of estimate $\hat{\sigma}_k^{\mathrm{MLE}}$ given by (12.56)—and correspondingly for all other parameters—is asymptotically unbiased and asymptotically efficient as $T \to \infty$. In particular, this estimator is asymptotically normally distributed with asymptotic variance

$$\begin{aligned}\lim_{T\to\infty} \operatorname{Var}\big(\widehat{\Sigma}_k^{\mathrm{MLE}}\big)^{-1} &= \mathbb{E}\bigg[\frac{\partial^2 \log \ell(n\,|\,\theta_q,\theta_w,\sigma)}{\partial \sigma_k^2}\bigg] \\ &= \sum_{n\in\mathbb{N}_0^{A\times 2\times(K+1)\times T}} \bigg(\frac{\partial \log \ell(n\,|\,\theta_q,\theta_w,\sigma)}{\partial \sigma_k}\bigg)^2 \ell(n\,|\,\theta_q,\theta_w,\sigma)\,.\end{aligned}$$

This term is known as Fisher information and is widely discussed in the statistical literature, see, for example, Harville [64], as well as Lehmann and Romano [79, Section 12.4.1].

12.4 Estimation via Markov chain Monte Carlo

As briefly outlined in Remarks 12.40(c) and 12.57, deriving maximum a posteriori estimates and maximum likelihood estimates via deterministic numerical optimisation can be challenging or sometimes impossible due to high dimensionality. To give a rough estimate of the number of variables to be optimised, assume that we have eight age categories starting from age 50 for each gender, data for 15 years and ten non-idiosyncratic risk factors. In this case we end up with 394 parameters (362 to be optimised as weight parameters u and v for one risk factor can be chosen arbitrarily, see Remarks 12.12(i)) for the maximum likelihood approach.

Alternatively, we can use a stochastic optimisation method called *Markov chain Monte Carlo*—referred to as MCMC from now on in this section. Introductions to this topic can be found, for example, in Gilks, Richardson and Spiegelhalter [55], Gamerman and Lopes [51], as well as Shevchenko [114, Section 2.11]. Its original purpose is to approximate integrals of the form

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x_1,\dots,x_d)\,\pi(x_1,\dots,x_d)\,dx_1\dots dx_d\,,$$

for a measurable function $f\colon \mathbb{R}^d \to \mathbb{R}$, with $d \in \mathbb{N}$, and for some $\mathbb{R}^d$-valued random variable X with density π. Many different MCMC algorithms exist amongst which we find the *random walk Metropolis–Hastings within Gibbs algorithm*. This is the algorithm we are going to work with and which we are going to briefly introduce in this section. The basic idea is to sample from an $\mathbb{R}^d$-valued time-homogeneous Markov chain $(X_i)_{i\in\mathbb{N}}$ with stationary density[50] π, in the case a direct generation of π is complicated or very expensive, typically if d is large. In particular, if there exists a transition probability density[51] $p(x,y)$ for all

[50] Given a transition kernel $P(x,A) = \mathbb{P}(X_2 \in A\,|\,X_1 = x)$ for all $x \in \mathbb{R}^d$ and all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$ of an $\mathbb{R}^d$-valued time-homogeneous Markov chain $(X_i)_{i\in\mathbb{N}}$, density $\pi\colon \mathbb{R}^d \to [0,\infty)$ is called stationary if $\int_{\mathbb{R}^d} \pi(x)\,dx = 1$ and

$$\int_A \pi(x)\,dx = \int_{\mathbb{R}^d} \pi(x)\,P(x,A)\,dx\,, \quad A \in \mathcal{B}(\mathbb{R}^d)\,.$$

Note that these conditional probabilities always exist in our case as $\mathcal{B}(\mathbb{R}^d)$ is a Borel space, see, for example, Kallenberg [71, Theorem 6.3].

[51] Given a transition kernel $P\colon \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$, corresponding densities are given by

$$p(x,y) = \frac{\partial P(x,(-\infty,y_1]\times\cdots\times(-\infty,y_d])}{\partial y_1\dots\partial y_d}\,, \quad x,y = (y_1,\dots,y_d) \in \mathbb{R}^d\,,$$

if they exist. In our particular case, $p(x,y)$ is explicitly given for all $x,y \in \mathbb{R}^d$ satisfying $x \neq y$ and with an atom in $x = y$.

$x, y \in \mathbb{R}^d$ with $x \neq y$ which satisfies the detailed balance equation

$$\pi(x)p(x,y) = \pi(y)p(y,x)\,, \quad x, y \in \mathbb{R}^d\,, \tag{12.59}$$

then π is a stable density. To see this, note that, for every $A \in \mathcal{B}(\mathbb{R}^d)$, integrating the left side of (12.59) gives

$$\int_A \int_{\mathbb{R}^d} \pi(x)p(x,y)\,dy\,dx = \int_A \pi(x)\,dx - \int_A \pi(x)\,P(x,\{x\})\,dx$$

and integrating the right side of (12.59) gives

$$\begin{aligned}\int_A \int_{\mathbb{R}^d} \pi(y)p(y,x)\,dy\,dx &= \int_{\mathbb{R}^d} \pi(y)\big(P(y,A) - P(y,\{y\})1_{y\in A}\big)\,dx\,dy \\ &= \int_{\mathbb{R}^d} \pi(y)\,P(y,A)\,dy - \int_A \pi(y)\,P(y,\{y\})\,dy\,.\end{aligned}$$

Then, in the case of the random walk Metropolis–Hastings within Gibbs algorithm, we split up in the form $p(x,y) = q(x,y)\alpha(x,y)$ with an arbitrary[52] transition kernel density $q(x,y)$ for new proposals which are accepted with acceptance probabilities

$$\alpha(x,y) = \begin{cases} \min\big\{1, \frac{\pi(y)q(y|x)}{\pi(x)q(x|y)}\big\} & \text{for } x \neq y\,, \\ 1 & \text{for } x = y\,, \end{cases}$$

and rejected with probability $1 - \alpha(x,y)$. It is immediate that $p(x,y)$ satisfies (12.59) and, thus, has stationary density π. Given mild regularity conditions[53], sample chains generated by this method converge to the stationary distribution, see, for example, Tierney [121] and also Robert and Casella [103, Sections 6–10] for general properties of this algorithm.

In our context, an MCMC approach requires a Bayesian setting which we automatically have in the maximum a posteriori approach, see Section 12.2. Similarly, we can switch to a Bayesian setting in the maximum likelihood approach, see Section 12.3, by simply multiplying the likelihood function with some prior density of parameters, e.g., an improper constant prior. Thus, in the following, we base our argumentation solely on the maximum a posteriori approach and leave the straight-forward application to the maximum likelihood approach to the reader.

If we set $\pi = \pi(\theta_q, \theta_w, \lambda, \sigma)$ in the maximum a posteriori approach, the application of the random walk Metropolis–Hastings within Gibbs algorithm is straight-forward. Our goal is to get many samples $(\theta_q^i, \theta_w^i, \lambda^i, \sigma^i)$, with $i \in \mathbb{N}$, from the posterior distribution of (12.35) where the mode of these samples then corresponds to an approximation for (12.39). More stable estimates in terms of mean squared error, see Shevchenko [114, Section 2.10], are obtained by taking the mean over all samples once MCMC chains sample from the stationary distribution.

Remark 12.60 (Attention please). Taking the mean over all samples as an estimate, of course, can lead to troubles if posterior distributions of parameters are, e.g., bimodal, such that we end up in a region which is highly unlikely. It is therefore suggested to always have

[52] At least having the same support as π.

[53] If q is aperiodic, as well as irreducible and if $\alpha(x,y) > 0$ for every possible value $x, y \in \mathbb{R}^d$, then the Markov chain is irreducible and aperiodic with stationary density π. In particular, this holds if q is normally or truncated normally distributed with the right support.

a closer look at estimated posterior distributions and, if possible, to use every generated sample for further derivations. In Section 15.2, samples from MCMC are used to run our annuity model multiple times in order to extract parameter uncertainty. In that case it is possible to derive distributions of quantiles.

In the next step, we are giving a short sketch of the random walk Metropolis–Hastings within Gibbs algorithm as described in Shevchenko [114, Section 2.11.1] based on the maximum a posteriori approach. For notational convenience we use abbreviations $x^i = (x^i_j)_{j\in\{1,\dots,h\}} = (\theta^i_q, \theta^i_w, \lambda^i, \sigma^i)$, for $i \geq 0$. Note that the dimension of each sample equals $h = 8A + 2(2A(K+1) + K + 1) + KT + K$ for the maximum a posteriori approach, see Section 12.2.

Input: Based on the assumptions of Lemma 12.33, we require the posterior function $\pi(\cdot\,|\,n)\colon \mathbb{R}^h \to [0,\infty)$ and, for all $j \in \{1,\dots,h\}$, transition kernel densities $f_j(\cdot\,|\,x_j, \tau_j)$ given the previous state x_j and tuning parameter τ_j, e.g. from the normal or truncated normal density with mean $x_j \in \mathbb{R}$ and some standard deviation $\tau_j > 0$. This variance can be chosen arbitrarily at the beginning or be adapted throughout the procedure.

Output: Samples from a Markov chain with stationary density $\pi(\cdot\,|\,n)$.

initialise x^0 with a value in the support of $\pi(\cdot\,|\,n)$;
for $i = 1$ **to** *M such that sampled long enough from stationary distribution* **do**
  set $x^i = x^{i-1}$;
  for $j = 1$ **to** h **do**
    generate sample proposal $\hat{x}^i_j$ from transition kernel density $f_j(\cdot\,|\,x^i_j, \tau_j)$;
    derive acceptance probability

$$\alpha(i,j) = \min\left\{1, \frac{\pi(\hat{x}\,|\,n)\, f_j(x^i_j\,|\,\hat{x}^i_j, \tau_j)}{\pi(x^i\,|\,n)\, f_j(\hat{x}^i_j\,|\,x^i_j, \tau_j)}\right\}, \tag{12.61}$$

    where $\hat{x} := (x^i_1, \dots, x^i_{j-1}, \hat{x}^i_j, x^{i-1}_{j+1}, \dots)$;
    simulate u from a uniform distribution on $[0,1]$;
    if $u < \alpha_{i,j}$, **then**
      change position to proposal $x^i_j = \hat{x}^i_j$;
    else
      remain in previous position, i.e., $x^i_j = x^{i-1}_j$;
    end
  end
end

Algorithm 12.1: Single step random walk Metropolis–Hastings within Gibbs algorithm

Algorithm 12.1 is easy to implement and very powerful when many other methods break down due to high dimensionality. Acceptance probabilities do not depend on the normalisation constant of the posterior distribution. Thus, posterior densities just have to be specified up to a multiplicative constant which means that we can drop normalising factors. As already mentioned, depending on the chosen initial values and the chosen tuning parameter τ_j the method requires a certain burn-in period until the system becomes

stationary. This can be checked best through plotting chains. A typical class of transition kernel densities are normal or truncated normal distributions. The latter is bounded and, therefore, ensures the existence of a proper density of the posterior distribution. Again, note that MCMC in general returns an approximation for the joint posterior distribution of all parameters. It thus allows for error estimates, as well as probabilistic statements about estimators. But note that ultimately we are troubled with the curse of dimensionality as we will never be able to get an accurate approximation of the joint posterior distribution in a setting with several hundred parameters.

Remarks 12.62. (Useful hints for Algorithm 12.1).

(a) Estimates derived by matching of moments as described in Section 12.1 can be used as initial values x^0 to ensure a shorter burn-in period.

(b) Number of iterations $M \in \mathbb{N}$ has to be chosen such that we sample long enough from the stationary distribution in order to make the numerical error, due to finite number of samples, acceptably small. A measure for numerical error due to finite number of samples is the concept of standard errors as, for example, defined in the textbook of Shevchenko [114, Section 2.12.2].

(c) Tuning parameters τ_j with $j \in \{1, \dots, h\}$ can be chosen fixed or be adapted throughout the procedure. Badly chosen tuning parameters can lead to poor behaviour of MCMC chains, i.e., slow convergence towards the stable distribution. Typically, one tries to get average acceptance probabilities close to 0.234 which is asymptotically optimal for multivariate Gaussian proposals as shown in Roberts, Gelman and Gilks [104]. If the average acceptance probability of a parameter is too low, then proposals are too extreme and, therefore, not accepted very often. Then, a reduction of standard deviation τ_j in the proposal distribution may help. The reverse statement holds for high acceptance probabilities.

(d) Choosing an appropriate prior distribution, the stated algorithm works analogously using the likelihood function as given in (12.52).

(e) In many cases it is preferable to use the logarithm of posterior densities to avoid extreme values in high dimensions. Therefore, taking the logarithm of (12.61) gives

$$\begin{aligned}\log \alpha(i,j) = \min \big\{ 0, &\log \pi(\hat{x} \,|\, n) + \log f(x_j^i \,|\, \hat{x}_j^i, \tau_j) \\ &- \log \pi(x^i \,|\, n) - \log f(\hat{x}_j^i \,|\, x_j^i, \tau_j) \big\} \,.\end{aligned}$$

We then accept if $\log u < \log \alpha(i,j)$.

(f) Instead of generating a proposal for each parameter separately, it is legitimate and often better to sample proposals for several parameters, called blocks, in one step. Blocks can help to tackle issues with high correlation amongst parameters. For example, proposals for parameters $(u^i_{a,g,k})_{k \in \{0,\dots,K\}}$ may be sampled from a $(K+1)$-dimensional normal distribution. Such an approach leads to a faster algorithm but tuning gets more involved.

(g) As our implementations for high-dimensional MCMC settings face long execution times, it should be noted that there exist several possibilities of parallelisation and enhancements of the algorithm, ranging from easy to very sophisticated, see, e.g., Wilkinson [128] and Rosenthal [107, Section 4]. The easiest way is to run several independent MCMC chains with different starting points on different CPUs in a parallel way, each with a

reduced number of steps, e.g., 20 times 1 000 steps instead of 20 000 consecutive steps. It is recommended to use over-dispersed distributions for the different starting values. Special care has to be taken for random number generation in parallel codes as identical seeds can produce inconsistent results.

12.5 Illustrative example of estimation procedures

Consider the annuity model of Definition 11.11 with one age category a_0 having $m = 100\,000$ policyholders, one gender g_0 and one non-idiosyncratic risk factor Λ_1, i.e., $K = 1$, over a period of $T = 25$ years. Furthermore, recall Assumption 12.12 and set $\zeta_{a_0,g_0} = \phi_0 = \phi_1 = 0$, as well as $\psi_0 = \psi_1 =: \psi$, for sake of simplicity. Further parameter values are provided in Table 12.1. Note that with such a setting, values for $u_{a_0,g_0,0}$ and $v_{a_0,g_0,0}$ can be assumed to be fixed and need not be estimated, see Remarks 12.15(i).

Table 12.1: True parameter values for modelling setup.

	α_{a_0,g_0}	β_{a_0,g_0}	η_{a_0,g_0}	$u_{a_0,g_0,0}$	$u_{a_0,g_0,1}$	$v_{a_0,g_0,0}$	$v_{a_0,g_0,1}$	ψ	σ_1^2
true	−4.00	−0.01	0.01	0.00	1.00	0.02	−0.02	0.02	0.10

We proceed as follows: First, we start with a simulation of death counts. Therefore, we simulate realisations $(\lambda_1(t))_{t\in\{1,\dots,50\}}$ of risk factors $(\Lambda_1(t))_{t\in\{1,\dots,25\}}$. These realisations are then used to simulate the Poisson distributed number of deaths n, see Assumption 12.5, with parameters $q_{a_0,g_0}\, w_{a_0,g_0,0}$ for idiosyncratic deaths and $q_{a_0,g_0}\, w_{a_0,g_0,1}\, \lambda_k(t)$ for non-idiosyncratic deaths for all $t \in \{1, \dots, 50\}$.

As an illustration, we compare different estimation procedures given a simulation of death counts. Estimates are derived via matching of moments following the steps suggested in Remark 12.23 as well as Markov chain Monte Carlo (MCMC) methods based on (12.39) and (12.56) as described in Section 12.4. Starting values for maximum a posteriori and maximum likelihood estimates are derived in 'R' using the nlminb optimisation routine, see [99], but are not reliable as deterministic methods can get stuck in local maxima. For MCMC approaches we use 20 000 simulations with a burn-in period of 5 000 and proposals derived from truncated normal distributions. Standard deviations of the MCMC chains are abbreviated by 'standard dev.'. Adaptive tuning of MCMC is used such that mean acceptance probabilities are all close to the optimal[54] value of 0.234.

Table 12.2 summarises estimation results for some model parameters derived by all the different methods. Theses results illustrate that all estimation procedures give reasonable results for this simulation where, in particular, the matching of moments approach shows surprisingly accurate estimates whilst being easy and fast to calculate. True values of parameters are always between five and 95 percent quantiles of the chains generated by MCMC. Mode estimates of MCMC, i.e., parameter samples which give the highest value of the posteriori or likelihood function, are the analogue to corresponding point estimates of maximum a posteriori (MAP) and maximum likelihood (MLE) whereas mean estimates are more stable and, therefore, preferred. In particular, whilst all other procedures underestimate risk factor variance σ_1^2, mean estimates give better results as they account for skewness of posterior distributions. Trend reduction parameters η_{a_0,g_0} and ψ are particularly hard

[54] Asymptotically under Gaussian proposals, see Roberts, Gelman and Gilks [104].

Table 12.2: True values, estimates for matching of moment (MM), as well as for MCMC approaches with maximum a posteriori (MAP MCMC) and maximum likelihood (MLE MCMC). MCMC methods use 20 000 simulations and a burn-in period of 5 000. Standard errors are given in percent and defined as in Shevchenko [114, Section 2.12.2] with block size 40.

	α_{a_0,g_0}	β_{a_0,g_0}	η_{a_0,g_0}	$u_{a_0,g_0,1}$	$v_{a_0,g_0,1}$	ψ	σ_1^2
	deterministic						
true	−4.000	−0.010	0.010	1.000	−0.020	0.020	0.100
matching moments	−3.988	−0.012	0.000	0.956	−0.016	0.010	0.070
	MAP MCMC						
mode	−3.988	−0.012	0.000	0.975	−0.016	0.010	0.070
mean	−3.981	−0.013	0.032	0.987	−0.022	0.027	0.099
5% quantile	−4.048	−0.019	0.003	0.775	−0.044	0.002	0.057
95% quantile	−3.920	−0.007	0.078	1.204	−0.006	0.061	0.164
standard dev.	0.038	0.004	0.023	0.129	0.012	0.018	0.035
standard err. (in %)	0.188	0.018	0.106	0.655	0.060	0.086	0.084
	MLE MCMC						
mode	−3.994	−0.011	0.001	1.034	−0.025	0.034	0.071
mean	−3.990	−0.013	0.028	1.005	−0.022	0.027	0.096
5% quantile	−4.050	−0.018	0.003	0.812	−0.037	0.002	0.056
95% quantile	−3.930	−0.008	0.069	1.182	−0.007	0.064	0.153
standard dev.	0.037	0.003	0.021	0.113	0.009	0.019	0.031
standard err. (in %)	0.176	0.015	0.080	0.550	0.043	0.081	0.065

to estimate and confidence intervals are wide as surfaces of the posterior function and the likelihood function are flat along these parameters. It may therefore be useful to define trend reduction parameters a priori in order to avoid unstable behaviour of estimation procedures.

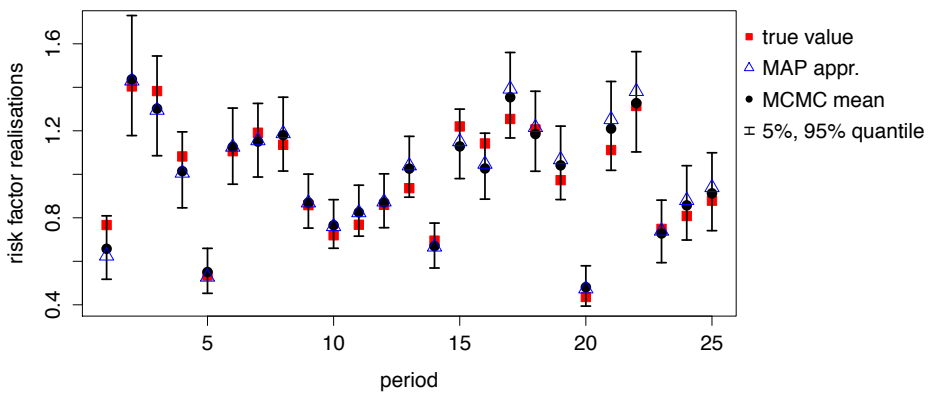


Figure 12.1: True and estimated risk factor realisations $\lambda_1(1), \ldots, \lambda_1(25)$ using the maximum a posteriori approach with deterministic optimisation (MAP), with its approximation in (12.48) (MAP appr.) and with the MCMC algorithm (MCMC mode, MCMC mean) , as well as corresponding error bars at five and 95 percent quantiles.

Figure 12.1 shows risk factor realisations for risk factors $\Lambda_1(1), \ldots, \Lambda_1(25)$, as well as their estimated values using the approximation given by (12.48) and using the MCMC method within the maximum a posteriori setting. Estimate for risk factor variance σ_1^2 obtained by (12.49) is given by 0.070. Again, note that estimates are reasonably accurate and, in particular, that Approximations (12.48) and (12.49) provide good results.

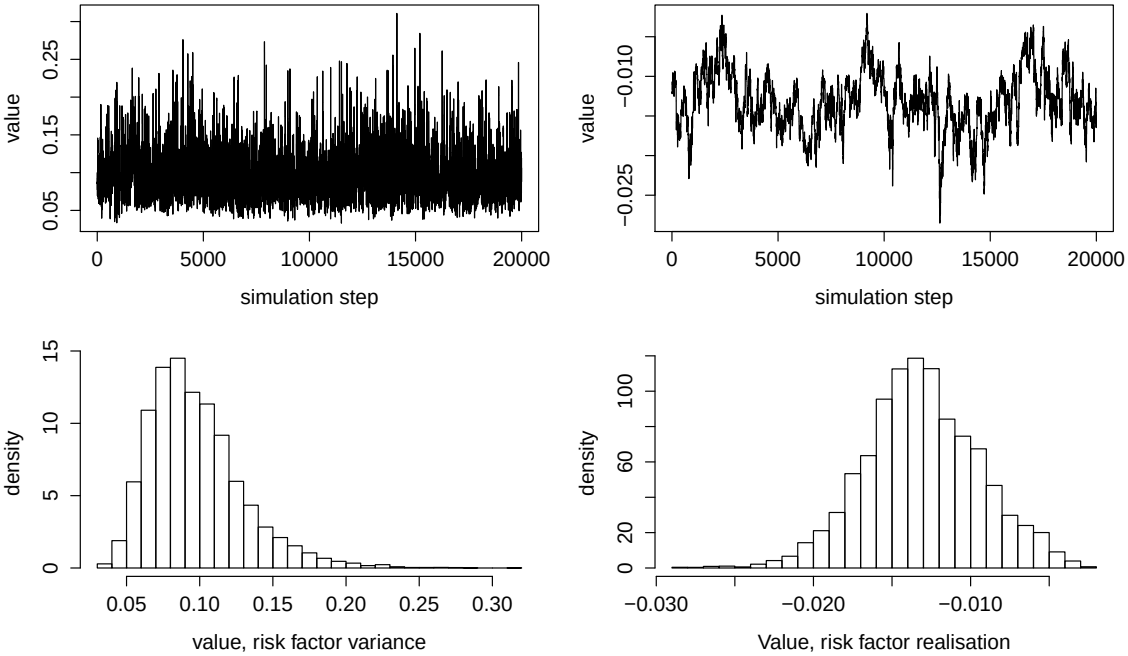


Figure 12.2: MCMC chains and density histograms for σ_1^2 (left) and $\lambda_1(1)$ (right).

Figure 12.2 then shows MCMC chains and corresponding density histograms for parameters σ_1^2 and β_{a_0,g_0} within the maximum likelihood and the maximum a posteriori setting, respectively. First of all, we can observe stationary behaviour of both MCMC chains. Remarkably, as illustrated in the left density histogram for parameter σ_1^2, posterior distributions of some parameters, i.e., the stationary distributions of the corresponding MCMC chains, are significantly right skewed. This observation outlines the fact that MCMC mode estimates may differ from MCMC mean estimates. Right skewed posterior distributions of risk factor variance σ_1^2 is reasonable as MCMC captures the risk of underestimating variances due to limited observations with possibly just few tail events.

Finally, Figure 12.3 shows estimates for death probabilities $q_{a_0,g_0}(1), \ldots, q_{a_0,g_0}(25)$ and weights $w_{a_0,g_0,1}(1), \ldots, w_{a_0,g_0,1}(25)$ of risk factor Λ_1 using matching of moments, as well as MCMC based on the maximum likelihood approach. The blue dash-dotted lines, denoted by MCMC mean, give estimates which are obtained by inserting means of estimated parameters into (12.13) and (12.14). The red dash dotted line gives five and 95 percent quantiles for death probabilities and weights from joint posterior distributions of parameters obtained by MCMC. True death probabilities and true weights always lie within these confidence intervals. Death rates at time $t \in \{1, \ldots, 25\}$ are simply given by $(n_{a_0,g_0,0}(t) + n_{a_0,g_0,1}(t))/(m \cdot q_{a_1,g}(t))$ for death probabilities and by $n_{a_0,g_0,i}(t)/(n_{a_0,g_0,0}(t) + n_{a_0,g_0,1}(t))$ for weights with $i \in \{0, 1\}$.

Remark 12.63 (Conclusion). This example suggests that matching of moments estimates, as well as estimates for risk factor realisations and variance given by (12.48) and (12.49) show

accurate and stable results while being straight-forward and fast to calculate. In general, maximum a posteriori estimates and maximum likelihood estimates usually show better results but are computationally much more expensive. Numerical optimisation routines such as gradient methods easily break down due to high number of parameters. MCMC methods provide very good results and give posterior distributions of estimates but, if not parallelised, execution times are higher.

Remark 12.64 (Blocks). Using proposal blocks for parameters $(u_{a_0,g_0,0}, u_{a_0,g_0,1})$, as well as $(v_{a_0,g_0,0}, v_{a_0,g_0,1})$ is also possible in this example. It makes tuning more involved whilst the reduction in proposals leads to faster execution times and reduced correlation amongst MCMC chains. This observation is a general pattern in our annuity model, i.e., sampling proposals from multidimensional distributions reduces correlations amongst MCMC chains but makes tuning more difficult.

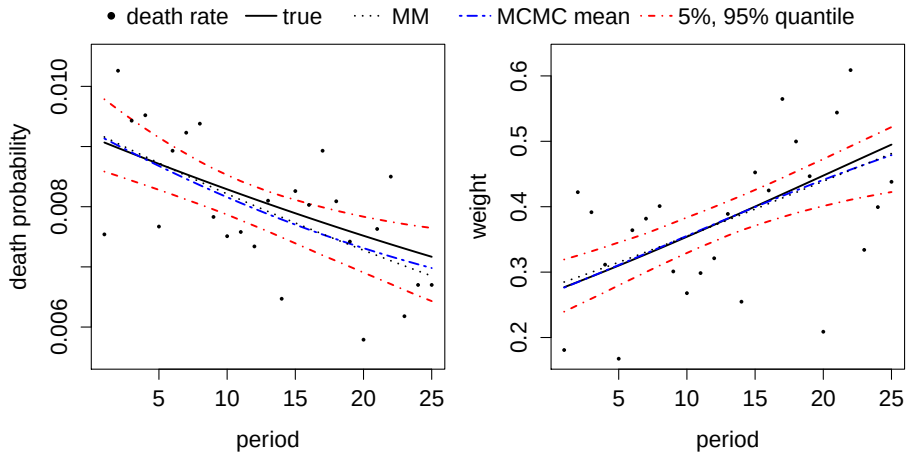


Figure 12.3: Death probability estimates $q_{a_0,g_0}(1), \ldots, q_{a_0,g_0}(25)$ *and weight estimates* $w_{a_0,g_0,1}(1), \ldots, w_{a_0,g_0,1}(25)$ *using matching of moments (MM) and mean MCMC estimates (MCMC mean) with five and 95 percent quantiles. Points show death rates.*

Chapter 13

Types of Risk

Regulators often require security margins in life tables when modelling annuity or certain life insurance products and portfolios to account for different sources of risk, including trends, volatility risk, model risk and parameter risk. Based on the requirements for Austria and Germany, see for example Kainhofer, Predota and Schmock [69], as well as Pasdika and Wolff [91], respectively, this chapter provides a short discussion on the main risks associated with annuity and life insurance portfolios as well as how they are incorporated in our annuity model of Definition 11.11. Note that the main risk associated with annuity portfolios is longevity which can be split into several sources.

The following sections do not cover the whole entity of different sources of risk but should encourage the reader to think critically about our modelling assumptions and how they may account for an advanced risk management. Some of the risks mentioned below are directly captured within our annuity model and others require additional portfolio information. If additional portfolio data are not available, certain security loadings can be added either to previously estimated death probabilities or as additional factor in the trend component $\beta_{a,g}(t)$ of Assumption 12.12, see Remark 13.1.

13.1 Trends

In our model, *mortality trends* are incorporated via Assumption 12.12 which is motivated by the Lee–Carter model. It is straight forward to arbitrarily change parameter families such that it fits the data as in the case when trends change fundamentally or when trends tend to increase. Such a phenomenon was observed in Austria around 1970, see Kainhofer, Predota and Schmock [69, Sections 4.5.3 and 4.6.2]. If other families for weights are used, one always has to check that they sum up to one over all death causes. Note that for certain alternative parameter families, mean estimates obtained from Markov chain Monte Carlo do not necessarily sum up to one anymore. Changing model parameter families may also be necessary when using long-term projections since long-term trends are fundamentally different from short-term trends. In this thesis, trend reduction techniques are incorporated via a time shift $\mathcal{T}_{\zeta,\eta}(t)$ to avoid vanishing death probabilities and weights in the far future based on the approach taken in Kainhofer, Predota and Schmock [69, Sections 4.6.2], see Remark 12.16. Since over the past few years mortality trends dramatically changed for higher ages, for example, in Austria, again see [69, Sections 4.7.2], it may be useful to

assume different trend families for different age categories in our model. Further estimation and testing procedures for trends in composite Poisson models in the context of convertible bonds can be found in Schmock [110].

Trends for weights are particularly interesting insofar as the model becomes sensitive to the change in the vulnerability of policyholders to different death causes over time. Cross dependencies over different death causes and different ages can occur. Such an effect can arise as a reduction in death rates of a particular cause can lead to increased death rates in another cause, several periods later, as people have to die at some point. Using Australian data, see Chapter 15, we see a general reduction in deaths due to circulatory diseases whereas, simultaneously, deaths due to mental and behavioural disorders get more frequent. Such observations may be crucial to forecast requirements for geriatric care, as well as medical supplies and resources. Note that our exponential family of weights, see (12.14), gives long-term forecasts which tend to peak in one risk factor. Nevertheless, estimation results are very accurate and mid-term forecasts show nice results, see Section 16.2.

Another major risk, which is usually not addressed in other annuity models, is the risk of unexpected deviations from a trend. In our model, this issue is captured with the variability introduced by common stochastic risk factors which effect all policyholders due to their weight simultaneously.

13.2 Statistical volatility risk

Assuming that the model choice is right and that estimated values are correct, life tables still just give mean values of death probabilities over a whole population. Therefore, in the case of German data it is suggested to add a gender-specific security margin of 6.26 percent for males and 7.22 percent for females to account for the risk of *random fluctuations* in deaths, approximately at a 95 percent quantile, see Pasdika and J. Wolff [91, Section 2.4.1]. More recently, see the German Actuarial Association (DAV) [28, Section 4.1], this security margin is assumed to be not gender specific due to legal reasons and it is set to 7.4 percent. In particular for small portfolios, this risk can be crucial since the law of large numbers may not apply. In our annuity model this risk is captured automatically. In particular, extreme statistical fluctuations can be found in the tails of the total portfolio loss distribution.

A direct comparison of the suggested security margin of 7.4 percent on death probabilities to an outcome of our annuity model, like certain quantiles in the total loss distribution, is not really meaningful. As a reference, we can use the same approach as given in Chapter 12 to estimate quantiles for death rates via setting $Y_i = 1$ for all $i \in M_{a,g}(T)$. These quantiles then correspond to statistical fluctuations around death probabilities. In particular, in Example 16.8 we roughly observe a deviation from death probability of 8.4 percent for the five percent quantile and of 8.7 percent for the 95 percent quantile of females aged 55 to 60 years in 2002, i.e., these values are in line with a security margin of 7.4 percent.

13.3 Model, selection and parameter risk

Modelling is usually a projection of a sophisticated real world problem on a relatively simple subspace which cannot cover all facets and observations in the data. Therefore, when applying our model to a portfolio of policyholders, we usually find *structural differences* to the data which is used for estimation. There may also be a difference in mortality rates between individual companies since different types of insurance products attract

different types of policyholders with a different individual risk profile. In addition, changes in the structure of future business and of mortality trends cannot be predicted and are therefore subject to uncertainty. Also, the actual data used for estimation may be subject to statistical fluctuations. In Germany, for these risks a minimal security margin of ten percent is suggested, see Pasdika and Wolff [91, Section 2.4.2]. These risks are not directly addressed in our model since they are data-related problems by nature and, thus, they can just be resolved by using portfolio data, see Remark 13.1.

Another major risk are *selection effects*. Observed mortality rates in insurance portfolios often show a completely different structure due to self-selection of policyholders. In particular, for ages around 60, this effect is very strong. In Germany, a security margin for death probabilities of 15 percent is suggested to cover selection effects, see DAV [28, Section 4.2]. In our particular case, to account for this source of risk we can subtract a risk margin of death probabilities before calculating the loss distribution via extended CreditRisk$^+$. Preferably, this risk margin should be based on portfolio data, see Remark 13.1.

The risk of statistical fluctuations in the data pool, i.e., *parameter risk*, which is used for estimation can be captured by our model in two ways. First, using Markov chain Monte Carlo (MCMC) for estimation of model parameters as described in Section 12.4 returns samples of the joint posterior distributions of the estimators. Thus, to account for parameter risk we can derive loss distributions in our annuity model of Definition 11.11 using different parameter samples taken from the MCMC chain. As our proposed extended CreditRisk$^+$ algorithm is numerically very efficient, we can easily run it for several thousand realisations of the MCMC chain. This procedure then yields approximated distributions of quantiles and expected shortfall such that we can a posteriori choose appropriate risk margins to account for parameter risk. Secondly, we may choose an elegant approach where we assume the parameters $q_{a,g}(t)\,w_{a,g,k}(t)$ for $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$, $k \in \{1,\dots,K\}$ and $t \in \{1,\dots,T\}$ to be random rather than fixed. Therefore, assume that risk factors $\Lambda_k(t)$ are gamma distributed with shape parameter α_k, as well as scale parameter β_k[55] and assume that $q_{a,g}(t)\,w_{a,g,k}(t)$ is independent of all other random variables and has a beta distribution with parameters $(\gamma_{a,g,k}(t), \alpha_k - \gamma_{a,g,k}(t))$[56] with $0 < \gamma_{a,g,k}(t) < \alpha_k$. In this case, $q_{a,g}(t)\,w_{a,g,k}(t)\,\Lambda_k(t)$ is again gamma distributed, see Stuart [117], with shape parameter $\gamma_{a,g,k}(t)$ and scale parameter β_k. Assuming a suitable family for $(\gamma_{a,g,k}(t))_{t\in\{1,\dots,T\}}$ such that trends in death probabilities and weights are considered as in Assumption 12.12, we can derive estimates for this modified approach with slightly adapted likelihood and posterior functions, see Sections 12.2 and 12.3. Alternatively, we can estimate parameters $\gamma_{a,g,k}(t)$ with $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$, $k \in \{1,\dots,K\}$ and $t \in \{1,\dots,T\}$ via matching of moments using previously derived estimates for death probabilities and weights from the original

[55] Its density is then given by

$$f_{\Lambda_k}(x) = \begin{cases} \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)}\, e^{-\beta_k\, x}\, x^{\alpha_k - 1} & \text{for } x > 0\,, \\ 0 & \text{for } x \le 0\,, \end{cases}$$

where Γ denotes the gamma function. Note that this distribution coincides with Definition 11.11(b) with $e_k = \alpha_k/\beta_k$ and $\sigma_k^2 = \alpha_k/\beta_k^2$, i.e., when expectations are set to one.

[56] Its density is given by

$$f_{B_{\alpha_k,\gamma_{a,g,k}(t)}}(x) := \begin{cases} \frac{1}{B(\alpha_k,\gamma_{a,g,k}(t))}\, x^{\alpha_k}\,(1-x)^{\gamma_{a,g,k}(t)} & \text{for } x \in [0,1]\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

where $B(y,z) := \int_0^1 t^{y-1}\,(1-t)^{z-1}\,dt$ for all $y, z > 0$ denotes the beta function.

model. In both cases, loss distributions can then be derived with a similar algorithm as described in Lemma 11.19.

Alternatively, in Kainhofer, Predota and Schmock [69, Sections 4.7.1] it is suggested that all these risks are addressed by adding a constant security margin on the trend. This approach has the great conceptional advantage that the security margin is increasing over time and does not diminish as in the case of direct security margins on death probabilities.

Remark 13.1 (Portfolio mortality data and individual information). If suitable company or portfolio data are available, risk margins for selection effects can be estimated as follows. First, calibrate the model using whole population data as previously illustrated and assume that variances of risk factors and weights for all age categories are fixed. Then, estimate death probabilities for company or portfolio data using any estimation procedure outlined in Chapter 12 where we just need to optimise over parameters $\alpha, \beta, \zeta, \eta$. This approach translates into the assumption that the subportfolio has the same characteristics as the whole population portfolio in terms of risk factor changes such as unexpected improvements of treatments and in terms of risk factor weights. Henceforth, this approach leads to risk margins on death probabilities as suggested by the DAV. Alternatively to the last step, recalling the more general model of Section 19.1.1, it is also possible to estimate risk factor means e given portfolio data and keeping all other parameters fixed. If necessary, further parameters such as risk factor variances can also be re-estimated. These procedures can be adapted freely such that individual information—such as smoker/non-smoker or address—can be considered. Various effects can occur when using individual information. Firstly, risk factor weights may shift so that it is necessary to re-estimate weights based on individual information. Secondly, it is possible that information such as address implies social standards which may indicate individual reaction on improvements in medication due high costs and, therefore, results in changed risk factor variances.

Chapter 14

Scenario Analysis

Scenario analysis is widely used in the financial industry to test reactions of portfolios—credit contracts, trading books, annuities—on stress events such as interest rate spikes or stock market drawdowns. In this short chapter we show that our annuity model is capable of testing scenarios of unexpectedly increased or decreased number of deaths due to a certain cause and the impact of it on a portfolio. Such a scenario may be the introduction of a new, very effective treatment or the unexpected outburst of an epidemic.

Definition 14.1 (Scenario)**.** Given the annuity model of Definition 11.11 and data for periods $1, \ldots, T$, a *scenario* is defined as a projected and potentially stressed vector of number of deaths $(N_{a,g,k}(T+1))_{(a,g,k)\in I} = (n^{\text{scen}}_{a,g,k}(T+1))_{(a,g,k)\in I}$ at period $T+1$ for a subset $I = I_A \times I_g \times I_K \subset \{1, \ldots, A\} \times \{\text{f}, \text{m}\} \times \{1, \ldots, K\}$ of age groups I_A, genders I_g and death causes I_K.

Once we are given a scenario, we want to estimate the impact of it on the portfolio for the next period $T+1$. We proceed in three steps:

First, we estimate all model parameters with any of the procedures given in Chapter 12 using data for $1, \ldots, T$ without considering the scenario. This step need not be repeated when other scenarios within the same setting are tested.

Then, in the second step, we estimate realisations of risk factors for the period $T+1$ given our scenario. Therefore, we use a slightly changed version of the maximum a posteriori estimation procedure defined in Section 12.2. More precisely, we use a modelling setup with fixed, previously estimated risk factor variances σ_k^2 and parameter forecasts $q_{a,g}(T+1)$, as well as $w_{a,g,k}(T+1)$ for all $a \in \{1, \ldots, A\}$, $g \in \{\text{f}, \text{m}\}$ and $k \in \{0, \ldots, K\}$, extrapolated from the estimation in the first step. The number of people $m_{a,g}(T+1)$ for all $a \in \{1, \ldots, A\}$ and $g \in \{\text{f}, \text{m}\}$ can also be extrapolated from the data or be derived from population forecasts. Optimisation of (12.35) with respect to $(\lambda_k(T+1))_{k\in I_K}$ gives estimates for risk factor realisations at $T+1$, denoted by $(\hat{\lambda}_k^{\text{MAP}}(T+1))_{k\in I_K}$. Alternatively and more easily, we may use Equation (12.42) to derive estimates for $(\hat{\lambda}_k^{\text{MAP}}(T+1))_{k\in I_K}$. Note that due to the independence of risk factors over time and due to independence amongst them, we just have to consider terms at time $T+1$ and terms within our scenario, i.e., within index set I_K.

In the third step, run the annuity model with extended CreditRisk$^+$ for time $T+1$ with estimated parameter forecasts and the modification that risk factors $(\Lambda_k(T+1))_{k\in I_K}$ are not random but replaced by their estimates $(\hat{\lambda}_k^{\text{MAP}}(T+1))_{k\in I_K}$, i.e., we run the model

given risk factor realisations of our scenario. As outlined in Section 15.2 for a very simple example, all common stochastic risk factors which are replaced by deterministic values $(\hat{\lambda}_k^{\mathrm{MAP}}(T+1))_{k\in I_K}$ can be joined with idiosyncratic risk such that new weights $w_{a,g,k}^{\mathrm{scen}}(T+1)$ become, for all $a \in I_A$ and $g \in I_g$,

$$w_{a,g,0}^{\mathrm{scen}}(T+1) = w_{a,g,0}(T+1) + \sum_{k\in I_K} w_{a,g,k}(T+1)\,\hat{\lambda}_k^{\mathrm{MAP}}(T+1)\,,$$

as well as

$$w_{a,g,k}^{\mathrm{scen}}(T+1) = \begin{cases} 0 & \text{for } k \in I_k\,, \\ w_{a,g,k}(T+1) & \text{for } k \notin I_k\,. \end{cases}$$

Then, weights over all death causes may not sum up to one anymore but Lemma 11.19 still can be applied. Thus, the loss distribution of the annuity portfolio and risk measures will change according to the stated scenario.

Remark 14.2 (Alternative representation of scenarios). Given a subset of stressed groups $I = I_A \times I_g \times I_K \subset \{1,\dots,A\} \times \{\mathrm{f},\mathrm{m}\} \times \{1,\dots,K\}$, a scenario may also be given in the form that certain death rates suddenly decrease by $x \in (-\infty, 100]^{|I|}$ percent within the following period. In that case, recalling Definition 14.1, simply set

$$n_{a,g,k}^{\mathrm{scen}}(T+1) = \left\lfloor \frac{m_{a,g}(T+1)\,q_{a,g}(T+1)}{m_{a,g}(T)\,q_{a,g}(T)} \Big(1 - \frac{x}{100}\Big)\, n_{a,g,k}(T) \right\rfloor, \quad (a,g,k) \in I\,,$$

where the first term above accounts for trends in mortality and population growth. Then, using this data, perform the three steps described before to derive impacts of this scenario.

An application of this approach towards scenario analysis based on Australian data is given in Section 15.2. Since the extended CreditRisk$^+$ algorithm is very fast, many different scenarios can easily be tested as indicated in Example 2.5.

Chapter 15

A Real World Example

As a real-world application for the previously described estimation procedures, in this chapter we take a look at Australian data for the period 1987 to 2011. We estimate parameters of our annuity model with ten non-idiosyncratic risk factors using the matching of moments approach, see Section 12.1, as well as the maximum likelihood approach with Markov chain Monte Carlo (MCMC), see Section 12.4. Then, based on the estimated model, we build a simple annuity portfolio and derive several interesting results including a scenario of reduced mortality due to neoplasms.

15.1 Estimation

As an applied example for estimation in our annuity model from Definition 11.11, as well as for some further applications, we take annual death data from Australia for the period 1987 to 2011. We fit our annuity model using the matching of moments approach as given in Section 12.1, as well as the maximum likelihood approach with Markov chain Monte Carlo (MCMC), see Section 12.4. Data source for historical Australian population, categorised by age and gender, is taken from the Australian Bureau of Statistics and data for the number of deaths categorised by death cause and divided into eight age categories[57] for each gender is taken from the AIHW. The provided death data are divided into 19 different death causes—based on the ICD-9 or ICD-10 classification—where we identify the following ten of them with common non-idiosyncratic risk factors: 'certain infectious and parasitic diseases', 'neoplasms', 'endocrine, nutritional and metabolic diseases', 'mental and behavioural disorders', 'diseases of the nervous system', 'circulatory diseases', 'diseases of the respiratory system', 'diseases of the digestive system', 'external causes of injury and poisoning', 'diseases of the genitourinary system'. We merge the remaining eight death causes to idiosyncratic risk as their individual contributions to overall death counts are small for all categories. Data handling needs some care as there was a change in classification of death data in 1997 as explained at the website of the Australian Bureau of Statistics or as in Magnus and Sadkowsky [82, Appendix A]. Australia introduced the tenth revision of the International Classification of Diseases (ICD-10, following ICD-9) in 1997, with a transition

[57] 50–54 years, 55–59 years, 60–64 years, 65–69 years, 70–74 years, 75–79 years, 80–84 years and 85+ years, denoted by $a_1, \ldots, a_8$, respectively. Younger people are not taken into account in this example as their contribution in annuity portfolios is minor since retirement age is usually above 50. If more age groups are considered, the number of parameters increases and special care has to be taken when tuning MCMC.

period from 1997 to 1998. Within this period, comparability factors[58] were produced as given in Table 15.1. Thus, for the period 1987 to 1996, death counts have to be multiplied by corresponding comparability factors and rounded to the nearest integer in order to avoid data inconsistencies.

Table 15.1: Comparability factors for ICD-9 to ICD-10.

death cause	factor
infectious	1.25
neoplasms	1.00
endocrine	1.01
mental	0.78
nervous	1.20
circulatory	1.00
respiratory	0.91
digestive	1.05
genitourinary	1.14
external	1.06
not elsewhere (idio.)	1.00

Trends are considered via Assumption 12.12 where trend reduction parameters are fixed a priori with values $\zeta_{a_i,g} = \phi_k = 0$ and $\eta_{a_i,g} = \psi_k = \frac{1}{150}$ for all $i \in \{1, \ldots, 8\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $k \in \{0, \ldots, K\}$ with $K = 10$. Thus, within the maximum likelihood framework, we end up with 394 parameters[59] in our annuity model. For the matching of moments approach we follow the approach suggested in Remarks 12.22 and 12.23 to account for trends. Risk factor variances are then estimated via Approximations (12.48) and (12.49) of the maximum a posteriori approach as they give more reliable results than matching of moments. As numerical optimisation for maximum likelihood breaks down due to high dimensionality, we use MCMC in this maximum likelihood setting instead. Assuming constant prior distributions, we use Algorithm 12.1 with single-step proposals taken from truncated normal distributions[60] with suitable bounds. Using joint proposals turns out to be too complicated to tune. Based on 40 000 MCMC steps with burn-in period of 10 000 we are able to derive estimates of all parameters where starting values are taken from matching of moments, as well as (12.48) and (12.49). Tuning parameters are frequently re-evaluated in the burn-in period. As the execution time of our algorithm is roughly seven hours on a standard computer in 'R', several parallel MCMC chains can be run, each with different starting values. With such an approach we can reduce execution times significantly.

Proper tuning of MCMC with real world data is very important as chains may not show a nice stationary behaviour in the case of poor tuning. As an illustration, Figure 15.1 shows MCMC chains of the variance of risk factor[61] for external causes of injury and poisoning

[58] The comparability factor for the idiosyncratic part is set to one here as it cannot be calculated from other given comparability factors.

[59] 362 to be optimised as idiosyncratic weight parameters are fixed, see Remarks 12.12(i).

[60] Using truncated normal distributions for proposals of all parameters translates into the assumption of bounded prior distributions which guarantees a proper posterior distribution. Of course, parameters for variances and death probabilities are unbounded a priori but, with reasonably large bounds, samples of the MCMC chains never come close to these boundaries. Moreover, using normal proposals for these parameters instead does not influence the results significantly. Thus, it is legitimate to use truncated normal distributions for proposals.

[61] The legend for death causes and age categories is given in Table 19.5.

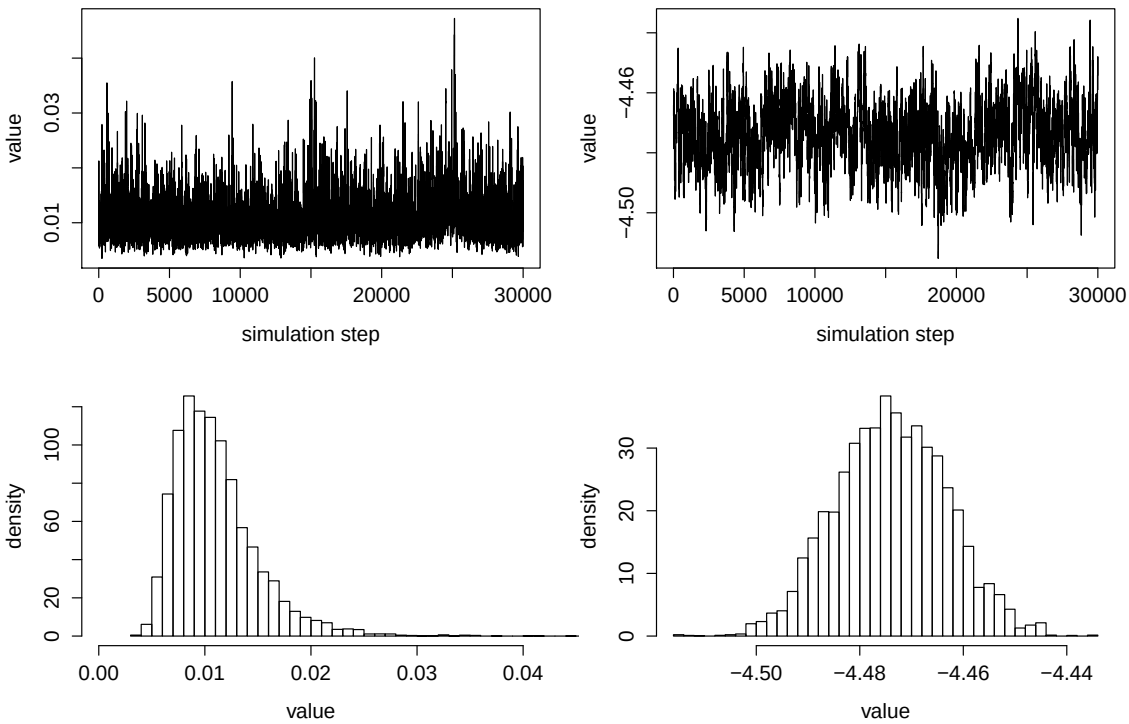


Figure 15.1: MCMC chains and corresponding density histograms (excluding first 10 000 samples) for the variance of risk factor for deaths due to external causes of injury and poisoning σ_9^2 (left), as well as for parameter $\alpha_{2,\mathrm{f}}$ (right), i.e., parameter for logarithmic death probability intercept of females aged 55 to 59 years.

σ_9^2, as well as of the parameter $\alpha_{2,f}$ for logarithmic death probability intercept of females aged 55 to 59 years. As already observed in the density histograms of Figure 12.2, we observe in Figure 15.1 that stationary distributions of MCMC chains for risk factor variances are typically right skewed. This indicates risk which is associated with underestimating variances due to limited observations of tail events.

Table 15.2 shows estimates for risk factor standard deviations using matching of moments, Approximation (12.49), as well as mean estimates of single-step MCMC with corresponding five and 95 percent quantiles, as well as standard errors. First, Table 15.2, as well as the full list in Section 19.3 illustrate that Approximations (12.48) and (12.49) and matching of moments estimates for parameters α, β, u and v are close to mean MCMC estimates. Standard errors, as defined in Shevchenko [114, Section 2.12.2] with block size 50, for corresponding risk factor variances are given in Section 19.3 and are consistently smaller than three percent in our case. Risk factor standard deviations are small but tend to be higher for death causes with just few deaths as statistical fluctuations in the data are higher compared to more frequent death causes. Small risk factor standard deviations support the simplifying assumption of independent risk factors as random effects overlay joint dependencies amongst death causes. See Chapter 17 for further, more rigorous model validation. Solely estimates for the risk factor standard deviation of mental and behavioural disorders give higher values which gets more obvious when looking at realisations of risk factors in Figure 15.2.

Table 15.2: Estimates for risk factor standard deviations $\sigma = (\sigma_1, \ldots, \sigma_{10})$ using matching of moments (MM), approximation as given in (12.49) (appr.) and MCMC mean estimates (mean), as well as corresponding standard deviations (stdev.) and five and 95 percent quantiles (5% and 95%).

	MM	appr.	mean	5%	95%	stdev.
infectious	0.1932	0.0787	0.0812	0.0583	0.1063	0.0147
neoplasms	0.0198	0.0148	0.0173	0.0100	0.0200	0.0029
endocrine	0.0743	0.0340	0.0346	0.0245	0.0469	0.0068
mental	0.1502	0.1357	0.1591	0.1200	0.2052	0.0265
nervous	0.0756	0.0505	0.0557	0.0412	0.0728	0.0098
circulatory	0.0377	0.0243	0.0300	0.0224	0.0387	0.0053
respiratory	0.0712	0.0612	0.0670	0.0510	0.0866	0.0110
digestive	0.0921	0.0645	0.0728	0.0548	0.0943	0.0123
external	0.1044	0.0912	0.1049	0.0787	0.1353	0.0176
genitourinary	0.0535	0.0284	0.0245	0.0141	0.0346	0.0066

As we use the MCMC approach based on maximum likelihood to reduce number of parameters, we do not directly derive estimates for risk factor realisations λ. Instead, we can use Equation (12.42) to derive approximations for risk factor realisation estimates where all required parameters are taken from the MCMC estimation. Results are shown in Figure 15.2. In the top figure we observe a massive jump in the risk factor for mental and behavioural disorders between 2005 to 2006 which is mainly driven by an unexpectedly high increase in deaths due to dementia, also see the report on dementia for Australia of the AIHW.

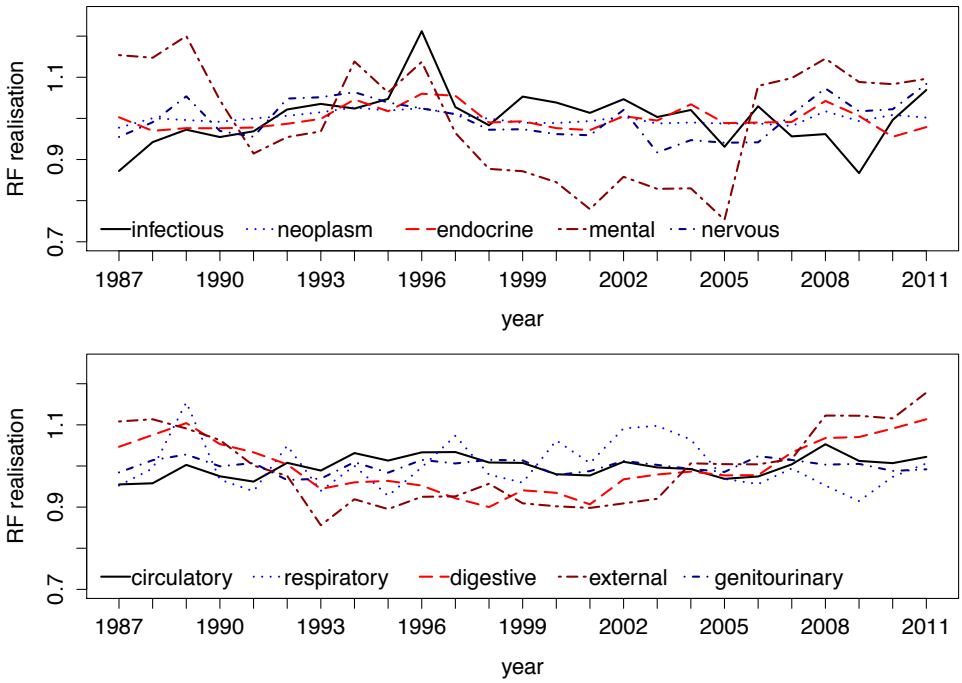


Figure 15.2: Estimated risk factor realisations for all death causes using (12.42) based on estimates taken from MCMC.

In the lower part of Figure 15.2, for example, we observe increased risk factor realisations of diseases of the respiratory system over the years 2002 to 2004. This is mainly driven by many deaths due to influenza and pneumonia during that period. Thus, besides its main purpose to derive loss distributions of annuity portfolios, our model provides a useful tool to detect phenomena in death data.

Table 15.3: Estimated weights for all death causes in years 2011, 2021 and 2031 using (12.14) with MCMC mean estimates for ages 60 to 64 years (left) and 80 to 84 years (right) for both genders. Five and 95 percent quantiles for the year 2031 are given in brackets.

	60 to 64 years			80 to 84 years		
	2011	2021	2031 (quant.)	2011	2021	2031 (quant.)
	male					
neoplasms	0.499	0.531	0.547 $\binom{0.561}{0.531}$	0.324	0.359	0.378 $\binom{0.392}{0.364}$
circulatory	0.228	0.165	0.116 $\binom{0.123}{0.109}$	0.325	0.242	0.173 $\binom{0.181}{0.164}$
external	0.056	0.060	0.062 $\binom{0.073}{0.053}$	0.026	0.028	0.028 $\binom{0.033}{0.024}$
respiratory	0.051	0.043	0.036 $\binom{0.040}{0.032}$	0.106	0.101	0.092 $\binom{0.101}{0.083}$
endocrine	0.044	0.053	0.062 $\binom{0.070}{0.055}$	0.047	0.062	0.077 $\binom{0.084}{0.070}$
digestive	0.041	0.039	0.036 $\binom{0.040}{0.031}$	0.027	0.024	0.020 $\binom{0.023}{0.018}$
nervous	0.029	0.040	0.052 $\binom{0.061}{0.045}$	0.045	0.054	0.061 $\binom{0.068}{0.055}$
not elsewhere (idio.)	0.018	0.023	0.028 $\binom{0.034}{0.023}$	0.015	0.017	0.018 $\binom{0.020}{0.016}$
infectious	0.014	0.019	0.025 $\binom{0.033}{0.020}$	0.015	0.019	0.022 $\binom{0.027}{0.019}$
mental	0.013	0.019	0.027 $\binom{0.036}{0.019}$	0.041	0.068	0.105 $\binom{0.130}{0.078}$
genitourinary	0.008	0.008	0.008 $\binom{0.010}{0.006}$	0.028	0.027	0.025 $\binom{0.028}{0.023}$
	female					
neoplasms	0.592	0.628	0.648 $\binom{0.662}{0.629}$	0.263	0.293	0.303 $\binom{0.319}{0.288}$
circulatory	0.140	0.092	0.060 $\binom{0.065}{0.055}$	0.342	0.233	0.149 $\binom{0.158}{0.140}$
respiratory	0.072	0.071	0.069 $\binom{0.078}{0.060}$	0.100	0.116	0.126 $\binom{0.139}{0.113}$
endocrine	0.038	0.038	0.037 $\binom{0.043}{0.032}$	0.051	0.061	0.068 $\binom{0.074}{0.061}$
nervous	0.036	0.043	0.051 $\binom{0.060}{0.043}$	0.054	0.068	0.080 $\binom{0.089}{0.071}$
external	0.035	0.033	0.032 $\binom{0.038}{0.026}$	0.024	0.025	0.023 $\binom{0.027}{0.020}$
digestive	0.031	0.028	0.024 $\binom{0.029}{0.020}$	0.034	0.029	0.023 $\binom{0.027}{0.020}$
not elsewhere (idio.)	0.022	0.023	0.023 $\binom{0.028}{0.019}$	0.023	0.025	0.024 $\binom{0.027}{0.022}$
infectious	0.014	0.017	0.020 $\binom{0.027}{0.015}$	0.017	0.021	0.024 $\binom{0.028}{0.020}$
mental	0.012	0.019	0.032 $\binom{0.046}{0.021}$	0.062	0.102	0.155 $\binom{0.188}{0.118}$
genitourinary	0.009	0.007	0.005 $\binom{0.006}{0.004}$	0.029	0.028	0.026 $\binom{0.028}{0.023}$

As already presumed in Figure 1.1 in the introduction, our model observes major shifts in weights of certain death causes over previous years as shown in Table 15.3. This table lists weights $w_{\mathrm{a,g},k}(t)$ for all death causes estimated for year 2011, as well as forecasted for years 2021 and 2031 using (12.14) with MCMC mean estimates for ages 60 to 64 years (left) and 80 to 84 years (right). It is obvious that, on top of general reduced mortality, the proportion of deaths for certain certain causes has changed massively over the period 1987 to 2011. Moreover, our model forecasts suggest that if these trends in weight changes persist, then

the future gives a whole new picture of mortality. First, deaths due to circulatory diseases are expected to decrease whilst neoplasms will become the leading death cause over most age categories. Moreover, deaths due to mental and behavioural disorders are expected to rise massively for older ages. This observation nicely illustrates the serial dependence, amongst different death causes[62] captured by our model. High uncertainty in forecasted weights is reflected by wide confidence intervals (values in brackets) for the risk factor of mental and behavioural disorders. These confidence intervals are derived from corresponding MCMC chains and, therefore, solely reflect uncertainty associated with parameter estimation. Note that results for estimated trends depend on the length of the data period as short-term trends might not coincide with mid- to long-term trends.

Table 15.4: Leading death causes with weights in brackets for males of all age categories in years 2011, 2031 and 2051 using (12.14) with MCMC mean estimate.

		male		
		2011	2031	2051
50–54 years	1.	neoplasms (0.385)	neoplasms (0.363)	neoplasms (0.307)
	2.	circulatory (0.223)	external (0.166)	external (0.163)
	3.	external (0.151)	circulatory (0.131)	infectious (0.142)
55–59 years	1.	neoplasms (0.469)	neoplasms (0.498)	neoplasms (0.474)
	2.	circulatory (0.222)	circulatory (0.119)	infectious (0.092)
	3.	external (0.085)	external (0.089)	external (0.083)
60–64 years	1.	neoplasms (0.502)	neoplasms (0.550)	neoplasms (0.535)
	2.	circulatory (0.226)	circulatory (0.114)	nervous (0.077)
	3.	external (0.055)	endocrine (0.061)	endocrine (0.074)
65–69 years	1.	neoplasms (0.505)	neoplasms (0.575)	neoplasms (0.575)
	2.	circulatory (0.226)	circulatory (0.101)	endocrine (0.082)
	3.	respiratory (0.072)	endocrine (0.066)	mental (0.075)
70–74 years	1.	neoplasms (0.474)	neoplasms (0.550)	neoplasms (0.544)
	2.	circulatory (0.241)	circulatory (0.104)	mental (0.111)
	3.	respiratory (0.083)	endocrine (0.074)	endocrine (0.093)
75–79 years	1.	neoplasms (0.405)	neoplasms (0.478)	neoplasms (0.466)
	2.	circulatory (0.277)	circulatory (0.129)	mental (0.185)
	3.	respiratory (0.100)	mental (0.084)	endocrine (0.098)
80–84 years	1.	neoplasms (0.327)	neoplasms (0.385)	neoplasms (0.371)
	2.	circulatory (0.324)	circulatory (0.169)	mental (0.239)
	3.	respiratory (0.106)	mental (0.115)	endocrine (0.092)
85+ years	1.	circulatory (0.395)	circulatory (0.249)	mental (0.329)
	2.	neoplasms (0.217)	neoplasms (0.239)	neoplasms (0.216)
	3.	respiratory (0.115)	mental (0.164)	circulatory (0.133)

Taking a look at projected leading death causes for years 2011, 2031 and 2051 as given in Tables 15.4 and 15.5, we can observe an overall increase in deaths due to neoplasms, as well as mental and behavioural disorders whilst deaths due to circulatory diseases tend to decrease. This potential increase in deaths due to mental and behavioural disorders for older ages will have a massive impact on social systems as, typically, such patients need

[62] If fewer people die from circulatory diseases, the average age will increase. Simultaneously, an increase in (currently) hardly treatable old-age death causes, such as dementia, at some later stage cannot be avoided.

long-term geriatric care.

Table 15.5: Leading death causes with weights in brackets for females of all age categories in years 2011, 2031 and 2051 using (12.14) with MCMC mean estimate.

		female		
		2011	2031	2051
50–54 years	1.	neoplasms (0.576)	neoplasms (0.552)	neoplasms (0.493)
	2.	circulatory (0.118)	external (0.100)	external (0.102)
	3.	external (0.091)	circulatory (0.069)	not elsewhere (0.081)
55–59 years	1.	neoplasms (0.603)	neoplasms (0.615)	neoplasms (0.581)
	2.	circulatory (0.112)	nervous (0.056)	nervous (0.077)
	3.	respiratory (0.058)	respiratory (0.052)	not elsewhere (0.068)
60–64 years	1.	neoplasms (0.597)	neoplasms (0.653)	neoplasms (0.652)
	2.	circulatory (0.141)	respiratory (0.074)	mental (0.071)
	3.	respiratory (0.074)	circulatory (0.059)	respiratory (0.068)
65–69 years	1.	neoplasms (0.551)	neoplasms (0.619)	neoplasms (0.609)
	2.	circulatory (0.162)	respiratory (0.075)	mental (0.112)
	3.	respiratory (0.083)	circulatory (0.060)	nervous (0.065)
70–74 years	1.	neoplasms (0.467)	neoplasms (0.535)	neoplasms (0.522)
	2.	circulatory (0.212)	respiratory (0.103)	mental (0.142)
	3.	respiratory (0.098)	circulatory (0.081)	respiratory (0.092)
75–79 years	1.	neoplasms (0.365)	neoplasms (0.414)	neoplasms (0.378)
	2.	circulatory (0.271)	respiratory (0.117)	mental (0.245)
	3.	respiratory (0.103)	mental (0.116)	respiratory (0.108)
80–84 years	1.	circulatory (0.340)	neoplasms (0.295)	mental (0.324)
	2.	neoplasms (0.263)	mental (0.168)	neoplasms (0.256)
	3.	respiratory (0.101)	circulatory (0.145)	respiratory (0.126)
85+ years	1.	circulatory (0.441)	circulatory (0.273)	mental (0.503)
	2.	neoplasms (0.131)	mental (0.231)	circulatory (0.092)
	3.	mental (0.101)	neoplasms (0.127)	neoplasms (0.090)

15.2 A simple annuity portfolio with applications to parameter risk and scenario analysis

Based on the data and the model estimated in Section 15.1, we now build a simple annuity portfolio. Assume $m = 1\,600$ policyholders which distribute uniformly over all age categories and genders, i.e., each category contains 100 policyholders with corresponding death probabilities, as well as weights as previously estimated and forecasted for 2012,[63] i.e., for the following year after the last data observation. Annuities $X_i = Y_i$ for all $i \in \{1, \dots, m\}$ are paid annually and take deterministic values in $\{11, \dots, 20\}$ such that ten policyholders in each age and gender category share equally high payments. This gives a total of

$$\sum_{i=1}^{m} X_i = 24\,800$$

[63] Forecasted in the sense that we use estimates of α, β, u, v to derive death probabilities and weights as given in Assumption 12.12 at time 2012.

cumulative annuity payments if every policyholder survives. Then, running the extended CreditRisk$^+$ algorithm as given in Lemma 11.19 for the sum

$$S = \sum_{i=1}^{m} \sum_{j=1}^{N_i(2012-t_0)} Y_{i,j} \,,$$

with initial year $t_0 = 1986$ and where $N_i(2012 - t_0)$ denotes the number of deaths of policyholder $i \in \{1, \ldots, m\}$ in 2012 and where $(Y_{i,j})_{j\in\mathbb{N}}$ are independent copies of Y_i, yields the exact loss distribution $L = 24\,800 - S$. This distribution together with 95 and 99 percent quantiles is illustrated in Figure 15.3.

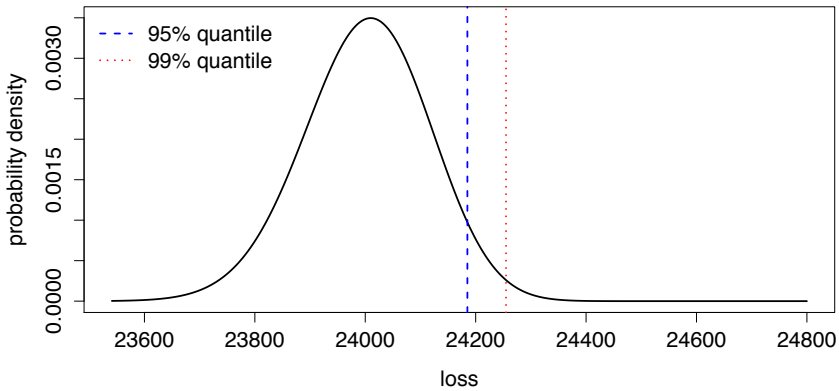


Figure 15.3: Loss distribution of L calculated with extended CreditRisk$^+$ and corresponding 95 and 99 percent quantiles.

More interestingly, we want to show an application of Section 13.3 and quantify parameter risk in our model. More precisely, we want to use the posterior distribution of parameters to account for errors which may occur due to estimation uncertainty, as well as due to statistical fluctuations in Australian data. Therefore, we choose the approach suggested in Section 13.3 where we run extended CreditRisk$^+$ several times to derive the distribution of S for different parameter samples from the MCMC chain. Here, we use 1 000 different samples of the posterior distribution of parameters so that we end up with an empirical distribution of the loss distribution of L. This makes it possible to derive an approximations for distributions of various quantiles of L which is illustrated in Figure 15.4 for the case of 95 and 99 percent quantiles. Obviously, if we believe that MCMC gives a suitable approximation of the posterior distribution of parameters, parameter risk is substantial.

As an application of Chapter 14 we analyse a scenario, indexed by 'scen', where death rates due to neoplasms suddenly decrease. Again, we use previously estimated parameters, forecasted for time 2012, and assume that deaths due to neoplasms are reduced by 25 percent in 2012 over all age categories. More precisely, set $N^{\text{scen}}_{a,g,2}(2012-t_0) = \lfloor 0.75\, N_{a,g,2}(2011-t_0) \rfloor$ as well as $m_{a,g}(2012 - t_0) = m_{a,g}(2011 - t_0)$ for all age categories $a \in \{1, \ldots, A\}$ and both genders $g \in \{\text{m}, \text{f}\}$. Then, given this scenario, we derive risk factor realisation $\hat{\lambda}_2^{\text{MAP}}(2012-t_0)$ using Equation (12.42) which gives $\hat{\lambda}_2^{\text{MAP}}(2012 - t_0) = 0.7991$. The common risk factor for neoplasms is then assumed to be deterministic and can therefore be joined with idiosyncratic risk, i.e., for all $a \in \{1, \ldots, A\}$, as well as $g \in \{\text{m}, \text{f}\}$ we can define new idiosyncratic weights

$$w^{\text{scen}}_{a,g,0}(2012 - t_0) = w_{a,g,0}(2012 - t_0) + w_{a,g,2}(2012 - t_0)\, \hat{\lambda}_2^{\text{MAP}}(2012 - t_0)$$

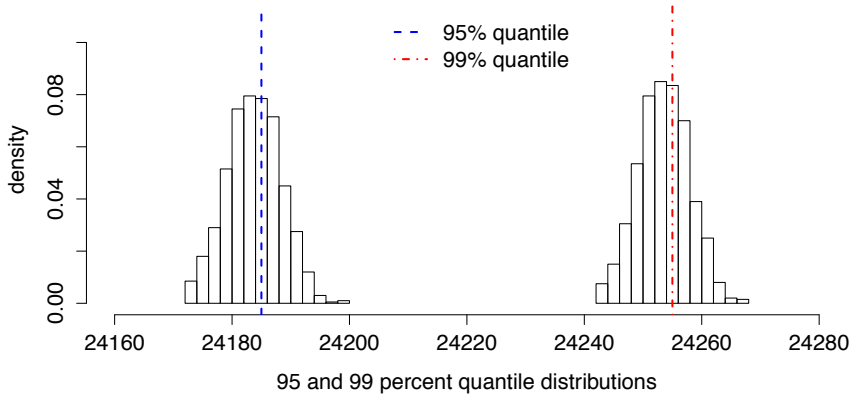


Figure 15.4: Distributions of 95 and 99 percent quantiles based on MCMC chain realisations.

and leave all other weights unchanged except the one for neoplasms

$$w^{\text{scen}}_{a,g,k}(2012-t_0) = \begin{cases} 0 & \text{for } k=2\,, \\ w_{a,g,k}(T+1) & \text{for } k \in \{1,3,4,\ldots,10\}\,. \end{cases}$$

Note that in our scenario, weights do not sum up to one anymore but Algorithm 11.19 still works. Thus, distributions of $S^{\text{scen}} := \sum_{i=1}^{m} \sum_{j=1}^{N_i^{\text{scen}}(2012-t_0)} X_i$ and $L^{\text{scen}} := 24\,800 - S^{\text{scen}}$ can easily be calculated.

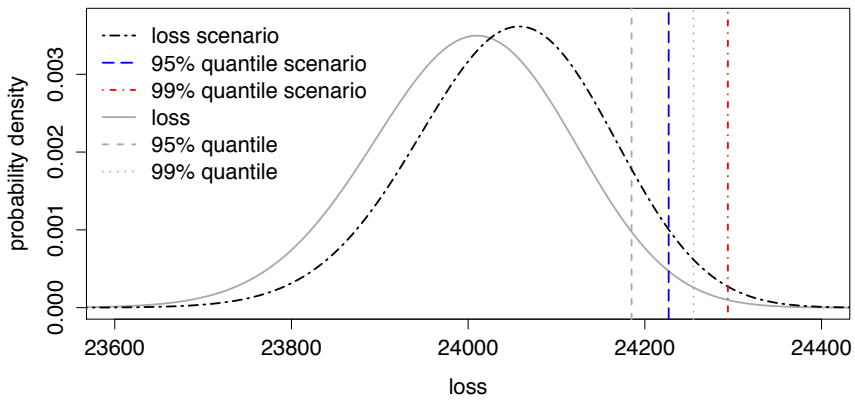


Figure 15.5: Loss distributions of L and L^{scen}, calculated with extended CreditRisk$^+$, as well as corresponding 95 and 99 percent quantiles.

Figure 15.5 shows probability distributions of loss L, as well as of scenario loss L^{neo} with corresponding 95 percent and 99 percent quantiles. Corresponding quantiles of S, S^{scen}, as well as of L and L^{scen} are listed in Table 15.6. There, the main message is that a reduction of 25 percent in cancer death rates leads to a change of roughly six percent in small quantiles of S, i.e., in the dangerous left tail of S corresponding to high losses in L due to just few deaths.

Table 15.6: Value at risk of S and S^{scen} (top), as well as L and L^{scen} (bottom) at different levels δ, i.e., $q_\delta(S)$ and $q_\delta(S^{\text{scen}})$, as well as $q_\delta(L)$ and $q_\delta(L^{\text{scen}})$, using extended CreditRisk$^+$.

level δ	no scenario: S	with scenario: S^{scen}
0.10	654	611
0.05	616	574
0.01	546	507
level δ	no scenario: L	with scenario: L^{scen}
0.90	24 147	24 190
0.95	24 185	24 227
0.99	24 255	24 294

Chapter 16

Stochastic Life Tables and Mortality Forecasts

In this chapter we analyse further applications of our annuity model. First, Section 16.1 gives a short comparison between the one-factor Lee–Carter model and our annuity model with one common stochastic risk factor. Not surprisingly, both models deliver roughly the same interpolation results. But when it comes to prediction, confidence intervals obtained by the Lee-Carter approach seem to overestimate variations of death rates as shown in Example 16.8. Thus, in Section 16.2 we provide an advanced forecasting procedure for death rates and weights within our annuity model which uses multiple common stochastic risk factors giving tighter confidence bands. Using a similar approach, Section 16.3 provides a sophisticated stochastic procedure for population forecasts. Finally, in Section 16.4 we suggest an approach for producing and forecasting life tables using MCMC. This approach then enables us to derive expected future life time where some surprising results occur.

Modelling mortality has a long tradition and, therefore, a vast amount of approaches can be found in the literature. Amongst important achievements in the 19th century we find the famous works of Gompertz [58] and Makeham [83]. Based on the ideas of Gompertz and Makeham, several generalisations and applications can be found in the literature, see, for example, Wetterstrand [127]. Furthermore, many other parametric and non-parametric models have been developed, see Andersen and Vaeth [6] for a comprehensive study of various models. Stochastic mortality models have been introduced in the early 1990s amongst which we find the often-cited Lee–Carter model, see Lee and Carter [78], as well as numerous extensions, see, for example, Brouhns, Denuit and Vermunt [17]. Another important stochastic mortality model, which allows incorporation of cohort effects, was introduced by Cairns, Blake and Dowd [18]. It is a stochastic generalisation of the model introduced by Perks [93]. A quantitative comparison of several important stochastic mortality models can be found in Cairns et al. [19].

As our annuity model primarily deals with mortality, it is perfectly capable of modelling, estimating and forecasting *life tables*, as well as *population counts*. Most importantly, we have tools to estimate model parameters based on publicly available data. Using our annuity model, in particular together with Markov chain Monte Carlo, we can even choose appropriate security margins in the form of quantiles to account for statistical fluctuations, parameter risk and other uncertainties, see Chapter 13 for further discussion on this topic.

16.1 Comparison with Lee–Carter

To show that our multi-factor approach covers traditional models for estimating life tables as well, we compare the annuity model of Definition 11.11 to the elegant *Lee–Carter approach* introduced by Lee and Carter [78]. Given the number of living people $m_{a,g}(t)$, as well as annual deaths $n_{a,g}(t) := \sum_{k=0}^{K} n_{a,g,k}(t)$ for age category $a \in \{1, \dots, A\}$, gender $g \in \{\mathrm{f}, \mathrm{m}\}$ and periods $t \in \{1, \dots, T\}$, the Lee–Carter approach models logarithmic death rates

$$\log r_{a,g}(t) := \log \frac{n_{a,g}(t)}{m_{a,g}(t)}$$

in the form

$$\log r_{a,g}(t) = \mu_{a,g} + \kappa_t \nu_{a,g} + \varepsilon_{a,g,t}\,,$$

with independent normal error terms $\varepsilon_{a,g,t}$ with mean zero and a common time-specific components $(\hat{\kappa}_t)_{t\in\{1,\dots,T\}}$. Hence, death rates are driven by age- and gender-specific parts $\mu_{a,g}, \nu_{a,g}$ and a time component κ_t. Using suitable normalisations, estimates $\hat{\mu}_{a,g}, \hat{\nu}_{a,g}$ and $\hat{\kappa}_t$ for the components $\mu_{a,g}, \nu_{a,g}$ and κ_t for all $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $t \in \{1, \dots, T\}$ may be derived via method of moments and singular value decompositions such that estimated logarithmic death rates are then given

$$\log \hat{r}^{LC}_{a,g}(t) := \hat{\mu}_{a,g} + \hat{\kappa}_t \hat{\nu}_{a,g}\,. \tag{16.1}$$

Note that normalisation approaches for parameters in the Lee–Carter model are not unique throughout the literature, see Kainhofer, Predota and Schmock [69, Section 4.5.1] and Brouhns, Denuit and Vermunt [17]. Since the Lee–Carter method just uses the highest eigenvalue in the singular value decomposition, it is intuitively clear that this approach should coincide with one-factor models. To make this observation more rigorous, consider our annuity model with alternative scaling $\mathbb{E}[N_i(t)] = q_i^*$, see Remark 11.3, and one common stochastic risk factor $\Lambda_1(t)$ with fixed weights $w_{a,g,1}(t) = 1$ for all $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $t \in \{1, \dots, T\}$. Thus, given data $m_{a,g}(t)$ and $n_{a,g}(t) := \sum_{k=0}^{K} n_{a,g,k}(t)$ for $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $t \in \{1, \dots, T\}$, we first have to estimate model parameters and risk factor realisations $(\lambda_1(t))_{t\in\{1,\dots,T\}}$ using the maximum a posteriori approach, see Section 12.2. Since we just use a monotone time trend for death probabilities, see (12.13), we should use realisations $\lambda_1(t)$ of risk factor $\Lambda_1(t)$ to compensate for the variation introduced by κ_t in the Lee–Carter approach. Henceforth, recalling (16.1), we expect

$$\hat{r}^{\mathrm{LC}}_{a,g}(t) \approx \hat{q}^{\mathrm{MAP}}_{a,g}(t)\, \hat{\lambda}^{\mathrm{MAP}}_1(t)\,,$$

for each $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $t \in \{1, \dots, T\}$. Thus, given that $\hat{q}^{\mathrm{MAP}}_{a,g}(t) < 0.5$ for $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $t \in \{1, \dots, T\}$, see Remarks 12.11(a), this conjecture implies that $\hat{\mu}_{a,g} + \log 2 + c \approx \hat{\alpha}^{\mathrm{MAP}}_{a,g}$, as well as $\hat{\kappa}_t \hat{\nu}_{a,g} - c \approx \hat{\beta}^{\mathrm{MAP}}_{a,g} \mathcal{T}_{\hat{\theta}^{\mathrm{MAP}}_{a,g}, \hat{\eta}^{\mathrm{MAP}}_{a,g}}(t) + \log \hat{\lambda}^{\mathrm{MAP}}_1(t)$ with some constant $c \in \mathbb{R}$.

Example 16.2. Using Australian death data from 1987 to 2011 as given in Section 15.1, we compare the outcomes of our annuity model to the Lee–Carter model as described above. Trends are considered via Assumption 12.12 where trend reduction parameters are fixed a priori with values $\zeta_{a_i,g} = \phi_1 = 0$ and $\eta_{a_i,g} = \psi_1 = \frac{1}{150}$ for all $i \in \{1, \dots, 8\}$ and $g \in \{\mathrm{f}, \mathrm{m}\}$. Mean estimates $\hat{q}^{\mathrm{MAP}}_{a,g}(t)$, as well as $\hat{\lambda}^{\mathrm{MAP}}_1(t)$ are derived using MCMC based on the maximum a posteriori approach, see Section 12.2, with 25 000 iterations and a burn-in

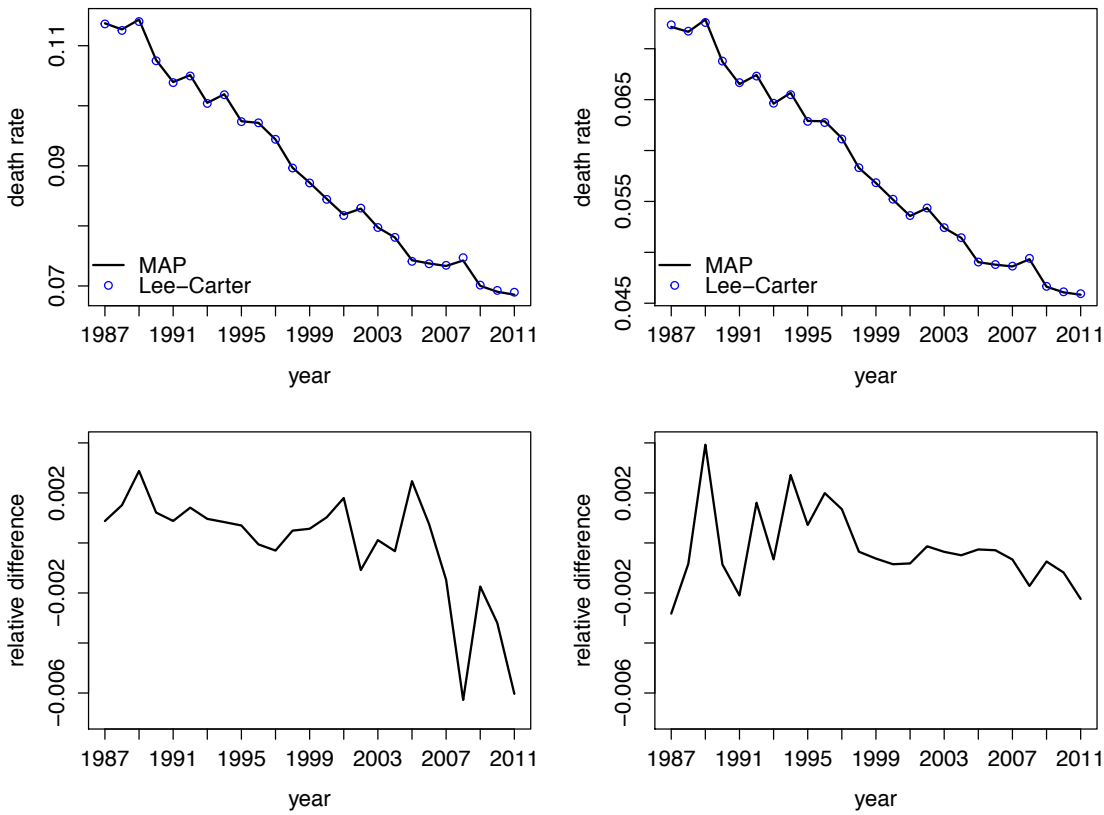


Figure 16.1: Estimated death rates from 1987 to 2011 using the Lee–Carter approach and MCMC with the maximum a posteriori approach (MAP) in our annuity model for males (top left) and females (top right) aged 80 to 84 years. Visually, the results are indistinguishable since relative differences for males (bottom left), as well as females (bottom right) are low.

period of 5 000. Note that we use alternative scaling, as described in Remark 11.3, so that death probabilities equal Poisson intensities, i.e., corresponding means of death indicators. Estimates obtained from the Lee–Carter model can be calculated using the function lca in 'R' of the 'demography' package [101]. Results are shown in Figure 16.1 where we observe the close relationship amongst both interpolation procedures for the age group 80 to 84 years and both genders. Visually, the results for both approaches are indistinguishable since relative differences are less than one percent. Also, for all other age categories, outcomes of both approaches are almost identical.

16.2 Forecasting death rates and rates of different death causes

Using our annuity model of Definition 11.11 and recalling Assumption 12.12, it is straightforward to *forecast death rates*, as well as *rates of different death causes* and to give corresponding confidence intervals. Using death rates, uncertainty in the form of confidence intervals represent statistical fluctuations, as well as random changes in risk factors. Additionally, using results obtained by Markov chain Monte Carlo (MCMC), see Section 12.4, it is even

possible to incorporate parameter uncertainty into predictions. Therefore, for the i-th sample $\theta^i := (\alpha^i, \beta^i, \zeta^i, \eta^i, u^i, v^i, \phi^i, \psi^i, \sigma^i)$ with $i \in \mathbb{N}$ of parameters $\theta = (\alpha, \beta, \zeta, \eta, u, v, \phi, \psi, \sigma)$ of the MCMC chain, i.e., for a realisation of the posterior distribution of parameters, define death probabilities

$$\log \hat{q}^i_{a,g}(t) := \alpha^i_{a,g} + \beta^i_{a,g} \mathcal{T}_{\theta_{a,g},\eta_{a,g}}(t)\,, \tag{16.3}$$

as well as weights

$$\hat{w}^i_{a,g,k}(t) = \frac{\exp\left(u^i_{a,g,k} + v^i_{a,g,k} \mathcal{T}_{\phi_k,\psi_k}(t)\right)}{\sum_{j=0}^{K} \exp\left(u^i_{a,g,j} + v^i_{a,g,j} \mathcal{T}_{\phi_j,\psi_j}(t)\right)}\,, \tag{16.4}$$

for all $a \in \{1, \ldots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $t \in \{T+1, \ldots, S\}$ with some $S \geq T+1$.

First, to forecast death probabilities and weights we may simply use Equations (16.3) and (16.4) for periods $t \in \{T+1, \ldots, S\}$ with some $S \geq T+1$ for various MCMC samples θ^i with $i \in \mathbb{N}$. Hence, this approach gives forecasts for death probabilities and weights where trends and uncertainty associated with parameter risk are included.

Alternatively, if we want to include statistical volatility risk in order to compare forecasts with true death rates and true rates of certain death causes, we suggest the following approach: For a specific age category $a \in \{1, \ldots, A\}$ and gender $g \in \{\mathrm{f}, \mathrm{m}\}$, set $m_{a,g}(t) := m_{a,g}(T)$[64], as well as $Y_j(t) := 1$ for all people $j \in M_{a,g}(T)$ with $|M_{a,g}(T)| = m_{a,g}(T)$ and $t \in \{T+1, \ldots, S\}$. Then, for a single estimate $\hat{\theta}$ of parameter vector θ—for forecasts without parameter uncertainty—or for various MCMC parameter samples $(\hat{\theta}^i)_{i \in N}$ with $N \subset \mathbb{N}$—for forecasts with parameter uncertainty—simply run our annuity model, see Section 11.2, with parameters forecasted for times $t \in \{T+1, \ldots, S\}$ by (16.3) and (16.4). We then obtain the distribution of the total number of deaths $S_{a,g}(t)$ or $S^i_{a,g}(t)$ given $\hat{\theta}$ or $\hat{\theta}^i$, respectively. For the case without parameter uncertainty, forecasted death rate $\hat{r}_{a,g}(t)$ is given by

$$\mathbb{P}\Big(\hat{r}_{a,g}(t) = \frac{n}{m_{a,g}(T)}\Big) = \mathbb{P}(S_{a,g}(t) = n)\,, \quad n \in \mathbb{N}_0\,, \tag{16.5}$$

at time $t \in \{T+1, \ldots, S\}$, for age category $a \in \{1, \ldots, A\}$ and gender $g \in \{\mathrm{f}, \mathrm{m}\}$. Distributions of forecasted death rates $\hat{r}^i_{a,g}(t)$ based on parameter sample $\hat{\theta}^i$ with $i \in N$ are similarly given by

$$\mathbb{P}\Big(\hat{r}^i_{a,g}(t) = \frac{n}{m_{a,g}(T)}\Big) = \mathbb{P}(S^i_{a,g}(t) = n)\,, \quad n \in \mathbb{N}_0\,. \tag{16.6}$$

We can then derive confidence intervals at every desired level where for the latter approach, based on multiple MCMC samples, quantiles of quantiles can be derived to account for parameter uncertainty. Note that, with our Poisson mixture approach, death rates $\hat{r}_{a,g}(t)$ and $\hat{r}^i_{a,g}(t)$ with $i \in N$ can take values greater than one with positive—but very small—probability.

Remark 16.7. The approach described above enables us to forecast death rates where uncertainty associated with random fluctuations and random changes in risk factors is included. In addition, if various MCMC parameter samples are considered, parameter uncertainty can be incorporated. Conversely, possible changes in trends are not captured by this approach, i.e., trends are assumed to be a priori estimated and then fixed. This issue can, for example, be tackled by re-estimating model parameters at each consecutive time for all outcomes of our annuity model.

[64] This assumption is somehow restricting as population does not stay constant over time which then leads to slightly reduced statistical fluctuation. Thus, our forecasted confidence bands will be a bit too wide. To avoid this, we may use population forecasts taken from governmental websites or use the approach outlined in Section 16.3.

Forecasting rates of certain death causes requires a slightly more sophisticated approach as each weight influences all the others within a certain age category $a \in \{1, \dots, A\}$ and gender $g \in \{\mathrm{f}, \mathrm{m}\}$. We suggest the usage of our annuity model where, in addition, losses are allowed to take different values depending on the underlying death cause, see Section 19.1 for the general extended CreditRisk$^+$ model. More precisely, for age category $a \in \{1, \dots, A\}$ and gender $g \in \{\mathrm{f}, \mathrm{m}\}$, we use exactly the same approach as for death rates forecasts where, in addition, the loss of policyholder $j \in M_{a,g}(T)$ due to death cause $k_0 \in \{0, \dots, K\}$ at time $t \in \{T+1, \dots, S\}$ is defined via

$$Y_{j,k}(t) := \begin{cases} (1,0) & \text{for } k \neq k_0\,, \\ (1,1) & \text{for } k = k_0\,. \end{cases}$$

Using the extended CreditRisk$^+$ algorithm with parameters forecasted for the period $t \in \{T+1, \dots, S\}$, see (16.3) and (16.4), based on a single estimate of θ or based on various MCMC parameter samples $(\hat{\theta}^i)_{i \in N}$ with $N \subset \mathbb{N}$—for forecasts with parameter uncertainty—then returns a two-dimensional random vector $S_{a,g,k_0}(t) := (S_{a,g}(t), S_{a,g,k_0}(t))$ or $S^i_{a,g,k_0}(t) := (S^i_{a,g}(t), S^i_{a,g,k_0}(t))$, respectively. In that case, the first component gives the total number of deaths and the second component gives the total number of deaths due to cause k_0. The distribution of forecasted death rate $\hat{r}_{a,g,k_0}(t)$ due to cause $k_0 \in \{0, \dots, K\}$ is then, excluding parameter uncertainty, given by

$$\mathbb{P}(\hat{r}_{a,g,k_0}(t) \in [x,y]) := \sum_{\substack{(n_1,n_2) \in \mathbb{N}_0^2 \\ \frac{n_2}{n_1} \in [x,y]}} \mathbb{P}\big(S_{a,g}(t) = (n_1, n_2)\big)\,,$$

or, including parameter uncertainty via consideration of various MCMC samples, forecasted rate $\hat{r}^i_{a,g,k_0}(t)$ for sample $\hat{\theta}^i$ with $i \in \mathbb{N}$ is given by

$$\mathbb{P}(\hat{r}^i_{a,g,k_0}(t) \in [x,y]) := \sum_{\substack{(n_1,n_2) \in \mathbb{N}_0^2 \\ \frac{n_2}{n_1} \in [x,y]}} \mathbb{P}\big(S^i_{a,g}(t) = (n_1, n_2)\big)\,,$$

for all $x, y \in \mathbb{R}$ with $x \leq y$, at time $t \in \{T+1, \dots, S\}$, for age category $a \in \{1, \dots, A\}$ and gender $g \in \{\mathrm{f}, \mathrm{m}\}$. Again confidence intervals, as well as quantiles of quantiles for the approach based on multiple MCMC samples can easily be derived at every desired level.

Of course, using these forecasts, different health scenarios can be tested and how they influence mortality, as well as how they result in changes of rates for various death causes. As already remarked in Section 11.3 and as suggested in Section 16.4, for the purpose of modelling life tables we can alternatively use Bernoulli mixture models instead of Poisson mixture models to avoid the shortcoming of multiple deaths.

Example 16.8 (Prediction of death rates)**.** As an illustration of the forecasting approaches mentioned above, we predict death rates for Australia and derive confidence intervals in order to compare these results with the Lee–Carter approach, as well as with realised, true death rates. As in Section 15.1, we use Australian death and population data, now for the years 1979 to 2001, to estimate model parameters via MCMC. Again, note that we have to modify the data according to comparability factors given in Table 15.1. Using the mean of 20 000 MCMC samples we forecast death rates and corresponding confidence intervals out of sample for the period 2002 to 2011 via our annuity model, see (16.5).

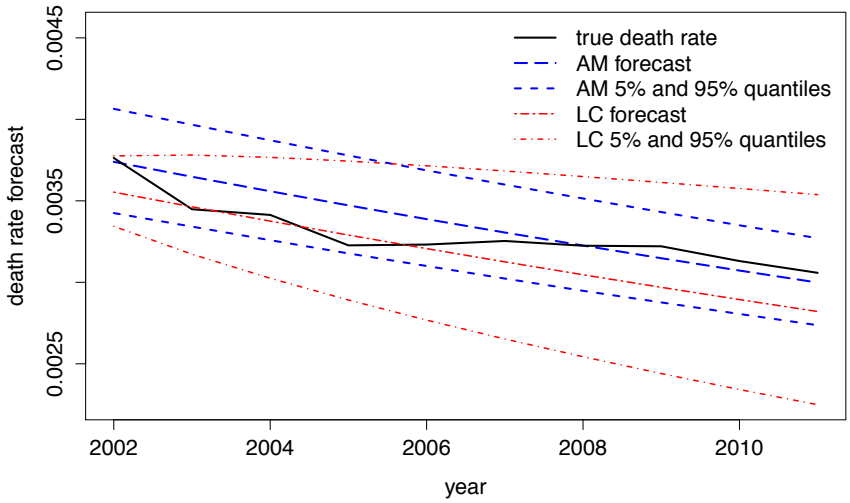


Figure 16.2: Forecasted death rates and 90 percent confidence intervals in Australia for the years 2002 to 2011 of females aged between 55 and 60 years within our annuity model (AM) and the Lee–Carter model (LC), as well as true death rates.

We can then compare these results to realised death rates within the stated period and to forecasts obtained by the Lee–Carter model. Therefore, Figure 16.2 gives death rate forecasts and corresponding five percent and 95 percent quantiles for our approach and the Lee–Carter approach, as well as realised, true death rates obtained in the years 2002 to 2011 in Australia for females aged 55 to 59 years. For our approach, the middle dashed, blue line gives forecasts for death probabilities obtained by mean MCMC estimates, see (12.13), and the upper and lower dashed, blue lines give quantiles for death rates obtained by (16.5). Correspondingly, the dash-dotted, red lines give forecasts using a univariate time series model obtained by coefficients from the fitted Lee–Carter model, see [17, 69, 78]. For the latter, we use the function lca in 'R' of the 'demography' package [101]. We can draw several conclusions from this figure. First, true death rates always fall in the 90 percent confidence band for both procedures. Secondly, confidence intervals obtained from the Lee–Carter approach are mostly wider than confidence intervals obtained by our model. As we assume the trend to be fixed in our model, spreads between five and 95 percent quantiles do not increase significantly over time whereas the autoregressive behaviour within Lee–Carter forecasts leads to growing spreads over time. More precisely for our annuity model, spreads between forecasts for death probabilities and 95 percent quantiles for death rates increase from 8.7 percent in 2002 to 9.1 percent in 2011, as well as spreads between forecasts for death probabilities and 5 percent quantiles for death rates increase from 8.4 percent in 2002 to 8.7 percent in 2011. Note that the confidence bands obtained in our approach roughly correspond to the commonly suggested security margin of 7.4 percent, see Section 13.2, for statistical fluctuations.

As another and more risk-sensitive illustration, Figure 16.3 shows the contribution of parameter uncertainty to quantiles obtained by our annuity model. Out of the 20 000 MCMC samples we take every hundredth sample to forecast quantiles of quantiles of death rates for the years 2002 to 2011 via (16.6). Taking every sample of the MCMC chain would require fairly long execution times. Again, we plot five and 95 percent quantiles of forecasts for death rates within our annuity model, as well as 90 percent confidence bands

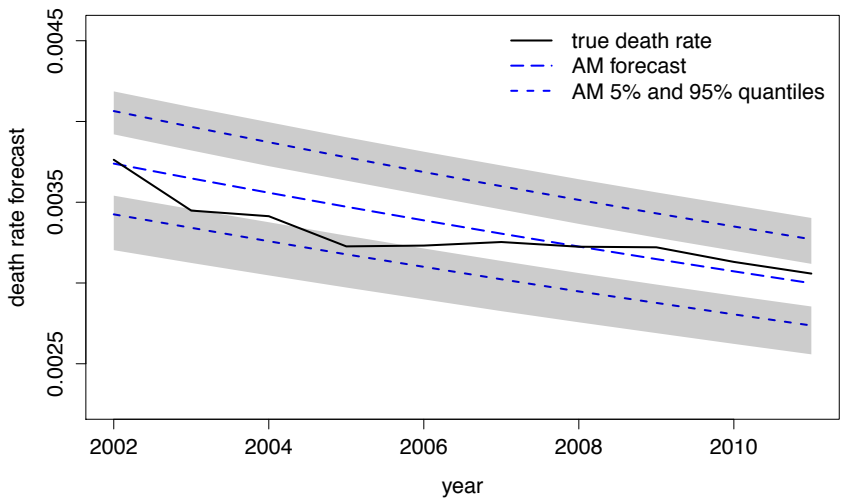


Figure 16.3: Forecasted death rates and 90 percent confidence intervals in Australia for the years 2002 to 2011 of females aged between 55 and 60 years within our annuity model (AM), as well as 90 percent confidence bands for stated quantiles indicating parameter uncertainty (shaded area).

based on these 200 MCMC samples, given as shaded areas. These shaded areas translate into uncertainty associated with parameter risk as MCMC samples from the posterior distribution of parameters. When using forecasts solely based on MCMC mean estimates, uncertainty purely comes from statistical fluctuations and random changes in risk factors. The uncertainty associated with parameter risk[65] is not negligible and increases over time from 9.8 percent in 2002 to 10.8 percent in 2011 for the five percent quantile and from 6.5 percent in 2002 to 8.7 percent in 2011 for the 95 percent quantile.

Remark 16.9 (Numerical underflow). Note that one has to be careful regarding numerical underflow as $\mathbb{P}(S_{a,g}(t) = 0)$, see Lemma 11.19, can become very low. To fix this problem, a suitable positive constant c^* can be added to parameter λ. Then, Recursion (11.20) yields the $\exp(c^*)$-fold of $\mathbb{P}(S_{a,g}(t) = \nu)$ for all $\nu \in \mathbb{N}_0$.

16.3 Population forecasts

As we are able to estimate trends for death probabilities and trends for weights in our model based on data for the periods $1, \ldots, T$, see Chapter 12, we can project them into the future and derive *population forecasts* and their distributions for $t \in \{T+1, \ldots, S\}$ with some $S \geq T+1$.

It is straight-forward to derive population forecasts for the next period as it just requires the usage of our annuity model of Definition 11.11 with $Y_i(T+1) := 1$ for all people $i \in \{1, \ldots, m_{T+1}\}$ living at time $T+1$. In that case all deaths of people are aggregated. For further periods the problem becomes more involved as information about age and gender is not preserved under the aggregation of deaths within extended CreditRisk$^+$ and, therefore, necessary categorisation of people is not possible anymore. This problem can be

[65] Width of the confidence band (shaded area), divided by quantiles of forecasts based on mean MCMC estimates (blue dashed line).

tackled by changing to a setup with higher dimensions where additional dimensions carry information about age and gender. Then, unfortunately, the algorithm quickly becomes computationally very expensive as the summation in (11.20) has to be performed over each point in a high-dimensional grid.

We illustrate this idea with a forecast for the group of females who will be aged 60 to 64 years in ten years. This results in a two-dimensional setting. Extensions to other, or more age categories and periods are straight-forward to obtain. The basic idea is that we choose two-dimensional losses where the second coordinate indicates if a person transits to the higher age group in the following year, given survival. Starting at time T, we want to predict the number of deaths for the period $T+1$ for females aged 50 to 54 years as these people will become the relevant age category in ten years from now. Assume that we have population counts m_T for females aged 50 to 54 years at time T. Then, we set $Y_i(T+1) = (1,0)$ if $i \in \{1, \ldots, \lfloor \frac{4}{5} m_T \rfloor\}$ or $Y_i(T+1) = (0,1)$ otherwise. This corresponds to the simplifying assumption that at each time step a certain percentage of people moves to the higher age category. For i.i.d. copies $(Y_{i,j}(T+1))_{j\in\mathbb{N}}$ of $Y_i(T+1)$, define

$$S(T+1, m_T) := \sum_{i=1}^{m_T} \sum_{j=1}^{N_i(T+1)} Y_{i,j}(T+1)\,.$$

Let $\pi_i \colon \mathbb{R}^2 \to \mathbb{R}$ for $i \in \{1,2\}$ denote the projection on the i-th coordinate. Then, note that $\pi_1(S(T+1, m_T))$ and $\pi_2(S(T+1, m_T))$ count the total number of deaths within the period $T+1$ of people who would and would not, respectively, change into the higher age category if they survived. Then, the number of people living in age category 50 to 54 years at $T+1$ is obtained by

$$M_1(T+1, m_T) := \left\lfloor \frac{4}{5} m_T \right\rfloor - \pi_1(S(T+1, m_T)) + \varepsilon_{1,T+1}$$

and, for age category 55 to 59 years, by

$$M_2(T+1, m_T) := \left\lceil \frac{1}{5} m_T \right\rceil - \pi_2(S(T+1, m_T)) + \varepsilon_{2,T+1}$$

where $\varepsilon_{1,T+1}$ and $\varepsilon_{2,T+1}$ denote exogenous correction terms for migration[66]. As a convention, we set negative values of $M_i(t+1, m_t)$ equal to zero. Using previously estimated model parameters as described in Chapter 12, we can derive the joint distribution of the random sum $S(T+1, m_T)$ using extended CreditRisk$^+$.

In the next step, this procedure can simultaneously be performed for period $T+2$ for each $m_{1,T+1}$ and $m_{2,T+1}$ in the support of $M_1(T+1, m_T)$ and $M_2(T+1, m_T)$, respectively, using parameter forecasts of death probabilities and weights for time $T+2$ as outlined in Remark 12.12. Therefore, set $m_{T+1} := m_{1,T+1} + m_{2,T+1}$ and note that all people who now belong to $m_{2,T+1}$, have a changed death probability and changed weightings due to the transition into the older age category 55 to 59 years. We then set $Y_i(T+2) = (1,0)$ if $i \in \{1, \ldots, \lfloor \frac{3}{4} m_{1,T+1} \rfloor\}$, or $Y_i(T+1) = (0,1)$ otherwise. Again, we use the simplifying assumption that a certain percentage of people moves to the higher age category. Similarly

[66] Immigration rates and forecasts, as well as fertility rates can mostly be found on governmental web sites. In the case of Australia these data can be found at website of the Department of Immigration and Border Protection.

as before, we can derive

$$S(T+2, m_{T+1}) := \sum_{i=1}^{m_{T+1}} \sum_{j=1}^{N_i(T+2)} Y_{i,j}(T+2)\,,$$

as well as the number of people living in age category 50 to 54 years at $T+2$

$$M_1(T+2, m_{1,T+1}, m_{2,T+1}) := \left\lfloor \frac{3}{4} m_{1,T+1} \right\rfloor - \pi_1(S(T+2, m_{T+1})) + \varepsilon_{1,T+2} \tag{16.10}$$

and, for age category 55 to 59 years,

$$\begin{aligned} M_2(T+2, &m_{1,T+1}, m_{2,T+1}) \\ &:= \left\lceil \frac{m_{1,T+1}}{4} \right\rceil + m_{2,T+1} - \pi_2(S(T+2, m_{T+1})) + \varepsilon_{2,T+2} \end{aligned} \tag{16.11}$$

where, again, $\varepsilon_{1,T+2}$ and $\varepsilon_{2,T+2}$ denote exogenous correction terms for migration. Final distributions of $M_i(T+2, M_1(T+1, m_T), M_2(T+1, m_T))$ for $i \in \{1,2\}$ are then obtained by mixing (16.10) and (16.11) with distributions of $M_1(T+1, m_T)$ and $M_2(T+1, m_T)$. This approach can then be iterated analogously up to period $T+10$ to finally derive a stochastic population forecast for females aged 60 to 64 years in ten years time. From a computational point of view, it is advisable to discretise distributions using the method of stochastic rounding, see Schmock [111].

Obviously, the major problem is to transfer the prior information about age and gender of each policyholder to the distribution of S a posteriori. In theory, this is easy as we can switch to a high-dimensional setting but, for applications, this becomes a burden in terms of computational complexity.

16.4 Death probability forecasts using Markov chain Monte Carlo

Using our annuity model of Section 11.2 with just idiosyncratic risk, i.e., $K = 0$, it is straight-forward to *derive* and *forecast death probabilities*. Prediction of death rates and corresponding confidence intervals work similarly as outlined in the more advanced approach[67] of Section 16.2. Alternatively, one common stochastic risk factor with a weight of one, i.e., no idiosyncratic risk, can be assumed in which case any of the estimation procedures provided in Chapter 12 can be used. However, relying solely on idiosyncratic risk, it is possible to assume death indicators to be independent and Bernoulli distributed instead of Poisson distributed. Then, recalling Assumption 12.5 and following the approach provided in Lemma 12.51, the likelihood function for parameters $(\alpha, \beta, \zeta, \eta)$ given data n is given by

$$\ell^{\mathrm{B}}(n \,|\, \alpha, \beta, \zeta, \eta) = \prod_{t=1}^{T} \prod_{a=1}^{A} \prod_{g \in \{\mathrm{f,m}\}} \binom{m_{a,g}(t)}{n_{a,g,0}(t)} q_{a,g}(t)^{n_{a,g,0}(t)} \left(1 - q_{a,g}(t)\right)^{m_{a,g}(t) - n_{a,g,0}(t)}\,, \tag{16.12}$$

with $0 \le n_{a,g,0}(t) \le m_{a,g}(t)$ for all $a \in \{1, \ldots, A\}$, $g \in \{\mathrm{f,m}\}$ and $t \in \{1, \ldots, T\}$. Using death data for dates $\{1, \ldots, T\}$, this likelihood function can then be used to estimate

[67] For a derivation of life tables based on multiple risk factors we could not find sufficiently rich data which is why we stick with a simpler approach in this section.

parameters and corresponding forecasts for dates $t \in \{T+1, T+2, \dots\}$ via the parameter family given in (12.13), i.e.,

$$q_{a,g}(t) = F^{\mathrm{Lap}}\big(\alpha_{a,g} + \beta_{a,g}\,\mathcal{T}_{\zeta_{a,g},\eta_{a,g}}(t)\big)\,. \tag{16.13}$$

Similarly, of course, matching of moments or maximum a posteriori estimation can be used instead.

Remark 16.14 (Trend reduction). Using either simple linear regression of logarithmic Australian death rates or, otherwise, the stated approach, both for the years 1974 to 2013 grouped into four ten-year-blocks, we can actually observe a depreciating trend parameter $\beta_{a,\mathrm{f}}$ over most ages for females. For males, we do not see such a clear trend at the moment. Nevertheless, we can assume that it is legitimate to incorporate trend reduction for long-term forecasts in Australia, also see Remarks 12.15(b). Unsuccessfully, we have also tried to find cohort effects in Australian data such as patterns in trend reduction which shift along the time path of certain generations. Solely, the generation born between 1945 and 1955 shows some trend acceleration.

Furthermore, it is of major interest to derive expected future life time for each age category once death probabilities have been estimated. To be consistent concerning longevity risk, mortality trends should be included in the derivation of expected future life time as a 60-year-old today with probably not have as good medication as a 60-year-old in several decades. However, it seems that this is not the standard approach in the literature. In particular in the case of Australia, life tables obtained by the Australian Bureau of Statistics for the years 2011 to 2013 are simply derived by taking death rates for the years 2011 to 2013 without considering mortality trends. However, based on the definitions given, for example, in Kainhofer, Predota and Schmock [69, Section 5.4], we define expected future life time[68] of a person in age category $a \in \{1, \dots, A\}$, of gender $g \in \{\mathrm{f}, \mathrm{m}\}$, at date T by

$$e_{a,g}(T) = \mathbb{E}[K_{a,g}(T)] = \sum_{k=1}^{\infty} {}_k p_{a,g}(T) \tag{16.15}$$

where survival probabilities[69] over $k \in \mathbb{N}$ years are given by

$${}_k p_{a,g}(T) := \prod_{j=0}^{k-1} \big(1 - q_{a+j,g}(T+j)\big)$$

and where $K_{a,g}(T)$ denotes the number of completed future years lived by a person of particular age and gender at time T. Note that the series above can be assumed to be a finite sum as survival probabilities are set to zero above some maximum age, e.g., 120 years. Correspondingly, by a simple change of order of summation, the standard deviation of $K_{a,g}(T)$ for each $a \in \{1, \dots, A\}$ and $g \in \{\mathrm{f}, \mathrm{m}\}$, at date T is given by

$$s_{a,g}(T) = \sqrt{\mathrm{Var}(K_{a,g}(T))} = \sqrt{2\sum_{k=1}^{\infty} k\,{}_k p_{a,g}(T) - e_{a,g}(T) - e_{a,g}(T)^2}\,. \tag{16.16}$$

[68] More precisely, we should be talking about expected curtate future life time as we calculate the expected number of completed future years lived by a person, i.e., we ignore the fraction of the death year when this person dies. For a formula for true expected future life time with a yearly constant force of mortality see, for example, [69, Section 5.4].

[69] The definition of survival probabilities over $k \in \mathbb{N}$ years as a product over one-year survival probabilities corresponds to the classical approach in life insurance. In our case, the product consists of one-year survival probabilities over transiting age groups with corresponding projections into the future.

In the following example we use this approach to derive the 2013 life table and expected future life time for Australians based on data for the years 1971 to 2013 where mortality trends are considered. In this example we see a remarkable jump in life expectancy when considering mortality trends which implies a massive impact on social budgets, pension funds, as well as insurance companies.

Example 16.17 (Prediction of death probabilities)**.** As an illustration of this simple interpolating and forecasting approach mentioned above, we derive the 2013 Australian life table and corresponding expected future life times based on publicly available death data[70] for the years 1971 to 2013. A complete list of results is given in Section 19.4. Death data are divided into 100 one-year age groups and a group 100+. As in Chapter 15, we use likelihood function (16.12) to estimate model parameters via MCMC. For stable estimation we fix $\zeta_{a,g} = 0$ for all ages $a \in \{0, \ldots, 100+\}$ and both genders $g \in \{\mathrm{f}, \mathrm{m}\}$.

Based on 35 000 MCMC samples with a burn-in period of 15 000, we get death probabilities, as well as mortality trends for each age and gender by inserting means over sample chains into (16.13). Note that MCMC provides samples from the posterior distribution of parameters such that parameter uncertainty can be estimated via confidence intervals on death probabilities. In this example, we observe negligible parameter uncertainty due to a long period of data which is why MCMC quantiles are not listed in Section 19.4, such as 90 percent confidence intervals for expected future life times are consistently smaller than 0.4 years.

Once parameters α, β and η have been estimated, we smooth all components along ages for each gender separately using function `cobs` in 'R' of the 'cobs' package [87] with 25 knots so that noise for forecasted death probabilities is reduced. These smoothing splines are very flexible as various constraints such as positivity or fixed knots can be imposed. Note that the less knots we use, the rougher the smoothing gets. Even more restricting, Kainhofer, Predota and Schmock [69, Section 4.7.3] make trends monotonic by partial linear interpolation to avoid non-monotone death probabilities. Logarithmic death probabilities $\log q_{a,g}(t)$ with corresponding forecasts, see (16.13), mortality trends $\beta_{a,g}$, as well as trend reduction parameters $\eta_{a,g}$ for males (left) and females (right) are provided in Figure 16.4. Recall that $1/\eta_{a,g}$ gives the time when initial trends are halved. We can draw some immediate, well-known conclusions. First we see an overall improvement in mortality over all ages where the trend is particularly strong for young ages and ages between 50 and 80 whereas the trend vanishes and even gets negative towards the age of 100 implying a natural barrier for life expectancy. Furthermore, we see the classical hump of increased mortality driven by accidents around the age of 20 which is more developed for males. Long-term forecasts of death probabilities can show non-monotonic behaviour due to the massive hump in the mortality trend, as well as slow trend reduction around the age of 50 to 60. This may imply that the strong effect on mortality improvement is just temporary Trend reduction parameters $\eta_{a,g}$ show similar patterns for males and females where reduction is stronger for females. This indicates a convergence of male death probabilities towards female death probabilities which can also be observed in expected future life times, see Section 19.4. Moreover, estimation results get noisy for very high ages due to sparse data.

For the derivation of expected future life times, we assume a death probability of zero for ages 121+ and constant parameters $\alpha_{a,g}, \beta_{a,g}$ and $\eta_{a,g}$ (before smoothing) for ages $a \in \{101, \ldots, 120\}$ given by previously estimated, corresponding parameters for group 100+.

[70] Death counts and population counts for each age and gender are taken from the Australian Bureau of Statistics and again the Australian Bureau of Statistics, respectively.

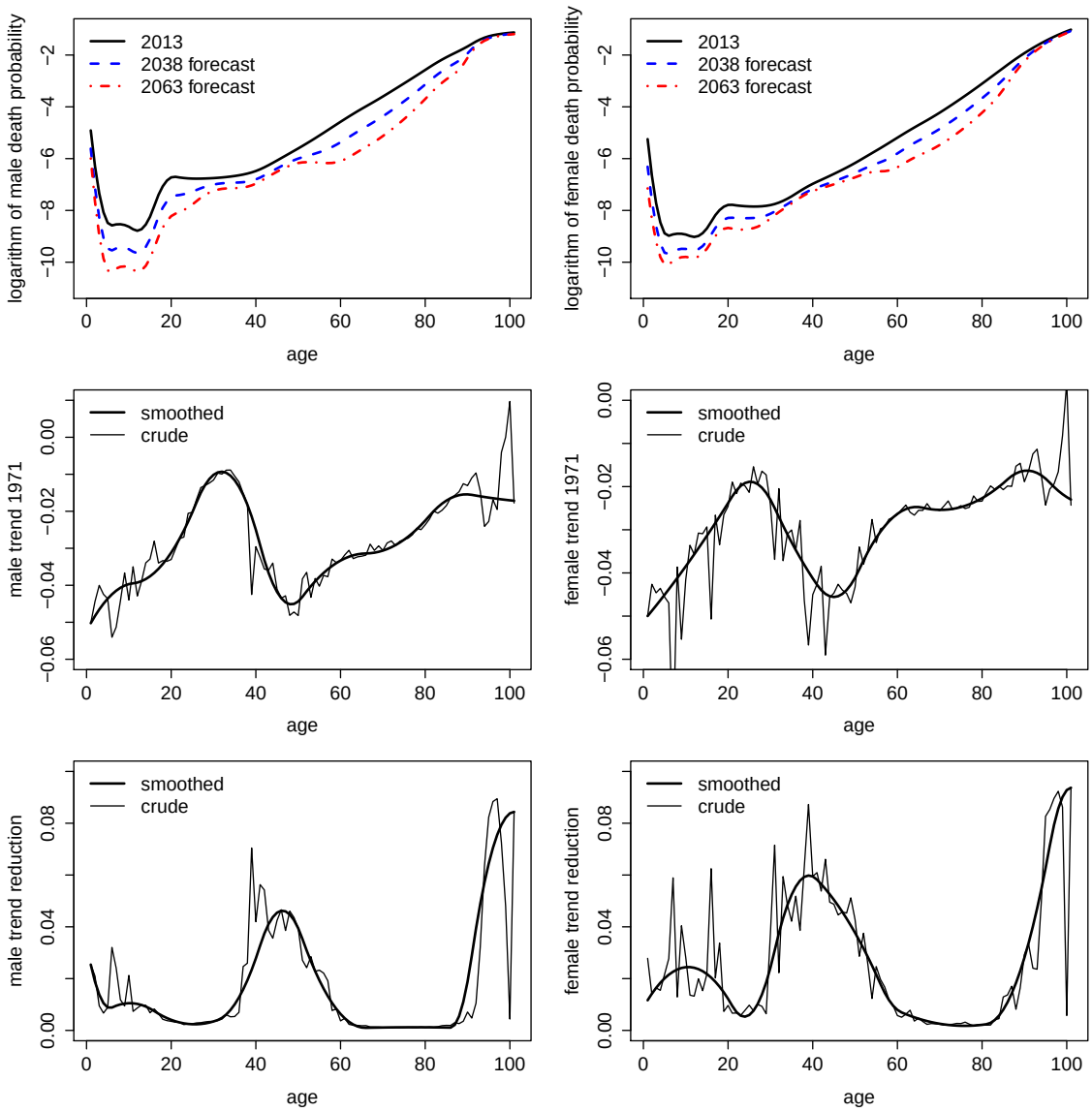


Figure 16.4: Australian logarithmic death probabilities (top) for 2013, as well as forecasts for 2063 and 2113, i.e., $(\log q_{a,g}(43), \log q_{a,g}(68), \log q_{a,g}(93))_{a\in\{1,\dots,A\}}$*, based on our annuity model using Australian mortality data from 1971 to 2013, as well as corresponding mortality trends (middle), i.e.,* $(\beta_{a,g})_{a\in\{1,\dots,A\}}$*, and trend reduction (bottom), i.e.,* $(\eta_{a,g})_{a\in\{1,\dots,A\}}$*, for males (left) and females (right).*

This approach does certainly not reflect real world observations but, nevertheless, it is used due to non-available data for older ages and minor impact on final results as just few people get older than 100 years. Kainhofer, Predota and Schmock [69, Section 4.7.2] provide a more sophisticated approach towards this issue. We can draw a remarkable conclusion from the results provided in Section 19.4. Whilst the Australian Bureau of Statistics made a press release in late October 2014 saying that 'Aussie men now expected to live past 80', our model states that Australian men, born in 2013, are expected to live 87.95 years. This divergence arises as we consider mortality trends, even including trend reduction. If no trend is considered in our model, we end up with a life expectancy of roughly 80 years

for Australian men, born in 2013, coinciding with the result published by the Australian Bureau of Statistics. Thus, there is a gap of eight years in life expectancy for males and almost six years for women between our model with trend and without trend. It is thus highly recommended for every company or organisation which is exposed to longevity risk to consider mortality trends when modelling life insurance contracts or annuities as a gap of several years can have a massive impact on long-term liabilities.

Chapter 17

Model Validation and Model Selection

In this chapter we propose several validation techniques in order to check whether our annuity model of Definition 12.1 fits the given data or not. Results for Australian data, as given in Chapter 15, strongly suggest that the proposed model is suitable. Furthermore, we recall some classical model selection approaches.

For the following sections in this chapter assume that we are given data as described in Assumptions 12.3 and 12.5 and assume that model parameters have been estimated by means of a method described in Section 12. If any of the following validation approaches suggests misspecification in the model or if parameter estimation does not seem to be accurate, one possibility to tackle these problems is to reduce risk factors, i.e., merge death causes. A more formal approach would be a reduction of risk factors via principal component analysis or independent component analysis, see, e.g., Hyvärinen, Karhunen and Oja [66]. That being said, we would unfortunately lose the direct interpretation of risk factors as death causes.

17.1 Validation via cross-covariance

Having estimated all model parameters in our annuity model of Definition 11.11, transform the data as described in Remark 12.23 such that we may assume that sequence of number of deaths $(N'_{a,g,k}(t))_{t\in\{1,\dots,T\}}$ of age category $a \in \{1,\dots,A\}$, gender $g \in \{\mathrm{f},\mathrm{m}\}$ and cause $k \in \{0,\dots,K\}$ is i.i.d. over time. Therefore, we may drop time parameter t for notational convenience. Based on Equation (11.17) and Assumption 12.5, for every $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$, as well as $k \in \{0,\dots,K\}$, we may deduce that

$$\mathrm{Var}(N'_{a,g,k}) = \begin{cases} m_{a,g}\, q_{a,g}\, w_{a,g,0} & \text{for } k = 0\,, \\ m_{a,g}\, q_{a,g}\, w_{a,g,k} + (m_{a,g}\, q_{a,g}\, w_{a,g,k})^2 \sigma_k^2 & \text{else}\,, \end{cases} \tag{17.1}$$

and, for all $a' \in \{1,\dots,A\}$, $g' \in \{\mathrm{f},\mathrm{m}\}$ with $a \neq a'$ or $g \neq g'$ and $k \in \{1,\dots,K\}$, based on Equation (11.17) we get

$$\begin{aligned} \mathrm{Cov}(N'_{a,g,k}, N'_{a',g',k}) &= \sum_{i\in M_{a,g}} \sum_{i'\in M_{a',g'}} \mathrm{Cov}(N'_{i,k}, N'_{i',k}) \\ &= m_{a,g}\, m_{a',g'}\, q_{a,g}\, q_{a',g'}\, w_{a,g,k}\, w_{a',g',k}\, \sigma_k^2\,, \end{aligned} \tag{17.2}$$

where $|M_{a,g}| = m_{a,g}$ and $|M_{a',g'}| = m_{a',g'}$. When death probabilities, weights, and risk factor variances are derived via Markov chain Monte Carlo (MCMC), see Section 12.4, we can use the samples from the Markov chains to derive quantiles, e.g., five and 95 percent quantiles, of (17.1) and (17.2). Then, these bounds can be compared to corresponding sample variances

$$s_{a,g,k} := \frac{1}{T-1} \sum_{t=1}^{T} (n'_{a,g,k}(t) - \bar{n}'_{a,g,k})^2 \,,$$

where $\bar{n}'_{a,g,k} := 1/T \sum_{s=1}^{T} n'_{a,g,k}(s)$, and to corresponding sample covariances

$$q_{a,g,a',g',k} := \frac{1}{T-1} \sum_{t=1}^{T} (n'_{a,g,k}(t) - \bar{n}'_{a,g,k})\,(n'_{a',g',k}(t) - \bar{n}'_{a',g',k}) \,,$$

for all $a, a' \in \{1, \ldots, A\}$ and $g, g' \in \{\mathrm{f}, \mathrm{m}\}$, as well as $k \in \{0, \ldots, K\}$ with $a \neq a'$ or $g \neq g'$. Note that estimators corresponding to these estimates for variances and covariances are unbiased.

Remark 17.3 (Example of Chapter 15, continued). Applying the validation procedure for cross-covariances as described above to the example of Chapter 15, we get that 45.9 percent of all sample variances and covariances lie within the five and 95 percent quantiles of (17.1) and (17.2), respectively, based on our derived MCMC chain. Thus, roughly half of all variances $\mathrm{Var}(N'_{a,g,k})$ and covariances $\mathrm{Cov}(N'_{a,g,k}, N'_{a',g',k})$ are accepted on a 10 percent significance level.

17.2 Validation via independence

One major outcome of our modelling approach is that death counts for different death cause intensities are independent as independent risk factors are assumed. Thus, for all $a, a' \in \{1, \ldots, A\}$ and $g, g' \in \{\mathrm{f}, \mathrm{m}\}$, as well as $k, k' \in \{0, \ldots, K\}$ with $k \neq k'$ and $t \in \{1, \ldots, T\}$, we have

$$\mathrm{Cov}(N_{a,g,k}(t), N_{a',g',k'}(t)) = 0 \,.$$

Having estimated all model parameters in our annuity model of Definition 11.11 by means of Chapter 12, transform the data as described in Remark 12.20 and subsequently normalise the transformed data, given $\mathrm{Var}(N'_{a,g,k}(t)|\Lambda_k(t)) > 0$ a.s., as follows:

$$N^*_{a,g,k}(t) := \frac{N'_{a,g,k}(t) - \mathbb{E}[N'_{a,g,k}(t)\,|\,\Lambda_k(t)]}{\sqrt{\mathrm{Var}(N'_{a,g,k}(t)|\Lambda_k(t))}} = \frac{N'_{a,g,k}(t) - m_{a,g}\, q_{a,g}\, w_{a,g,k}\, \Lambda_k(t)}{\sqrt{m_{a,g}\, q_{a,g}\, w_{a,g,k}\, \Lambda_k(t)}} \,,$$

for $a \in \{1, \ldots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$, $k \in \{0, \ldots, K\}$ and $t \in \{1, \ldots, T\}$ with $\Lambda_0(t) := 1$. Using the conditional central limit theorem as discussed in Grzenda and Zięba [59], we have $N^*_{a,g,k}(t) \to N(0,1)$ in distribution as $m_{a,g}(t) \to \infty$ where $N(0,1)$ denotes the standard normal distribution. Thus, using estimates for parameters $\alpha, \beta, \zeta, \eta, u, v, \phi, \psi$ and λ, indicated by a hat and obtained by one of the methods described in Chapter 12, we get estimates for death probabilities and weights via Assumption 12.12 such that normalised death counts $n^*_{a,g,k}(t)$ are given by

$$n^*_{a,g,k}(t) = \frac{n'_{a,g,k}(t) - m_{a,g}\, \hat{q}_{a,g}\, \hat{w}_{a,g,k}\, \hat{\lambda}_k(t)}{\sqrt{m_{a,g}\, \hat{q}_{a,g}\, \hat{w}_{a,g,k}\, \hat{\lambda}_k(t)}} \,,$$

for all $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$, $k \in \{0, \dots, K\}$ and $t \in \{1, \dots, T\}$ with $\hat{\lambda}_0(t) := 1$. Then, assuming that each pair $(N^*_{a,g,k}(t), N^*_{a',g',k'}(t))$ for $a, a' \in \{1, \dots, A\}$ and $g, g' \in \{\mathrm{f}, \mathrm{m}\}$, as well as $k, k' \in \{0, \dots, K\}$ with $k \neq k'$ and $t \in \{1, \dots, T\}$ has a joint normal distribution with some correlation coefficient ρ and standard normal marginals, we may derive the sample correlation coefficient

$$R_{a,g,a',g',k,k'} := \frac{\sum_{t=1}^{T} (N^*_{a,g,k}(t) - \bar{N}^*_{a,g,k})(N^*_{a',g',k'}(t) - \bar{N}^*_{a',g',k'})}{\sqrt{\sum_{t=1}^{T} (N^*_{a,g,k}(t) - \bar{N}^*_{a,g,k})^2 \sum_{t=1}^{T} (N^*_{a',g',k'}(t) - \bar{N}^*_{a',g',k'})^2}} ,$$

where $\bar{N}^*_{a,g,k} := 1/T \sum_{s=1}^{T} N^*_{a,g,k}(s)$. Then, the test of the null hypothesis $\rho = 0$ against the alternative hypothesis $\rho \neq 0$ rejects the null hypothesis at an δ-percent level, see, for example, Lehmann and Romano [79, Chapter 5.13], when

$$\frac{|R_{a,g,a',g',k,k'}|}{\sqrt{(1 - R^2_{a,g,a',g',k,k'})/(T-2)}} > K_{\delta,T} , \tag{17.4}$$

with $K_{\delta,T}$ such that $\int_{K_{\delta,T}}^{\infty} t_{T-2}(y)\, dy = \delta/2$ where t_{T-2} denotes the density of a t-distribution with $(T-2)$ degrees of freedom. Note that in this case we test for correlation. If a significant correlation is present, one can always merge several death causes and look whether the model fits better afterwards. Various other non-parametric tests for independence are mentioned in the literature where rank tests are the most popular ones. See Feuerverger [47], as well as Lehmann and Romano [79, Chapter 6.8] for details about rank tests and further references.

Remark 17.5 (Example of Chapter 15, continued). Applying the validation procedure for independence as described above to the example of Chapter 15, we get that 88.9 percent of all independence tests, see (17.4), are accepted at a five percent significance level. Thus, we may assume that our model fits the data suitably with respect to independence amongst death counts due to different causes.

17.3 Validation via serial correlation

Note that Assumption 12.1(b) guarantees that the sequence $(N_{a,g,k}(t))_{t \in \{1,\dots,T\}}$ is independent and thus uncorrelated for all $a \in \{1, \dots, A\}$, $g \in \{\mathrm{f}, \mathrm{m}\}$ and $k \in \{0, \dots, K\}$. Using the same data transformation and normalisation as in Section 17.2, we may assume that random variables $(N^*_{a,g,k}(t))_{t \in \{1,\dots,T\}}$ are identically and standard normally distributed. Then, we can check for serial dependence and autocorrelation in the data. If we find such dependence structures, then the model specifications will probably not fit the data. Many tests are available most of which assume an autoregressive model with normal errors such as the Breusch–Godfrey test, see Godfrey [56]. For the Breusch–Godfrey test a linear model is fitted to the data where the residuals are assumed to follow an autoregressive process of length $p \in \mathbb{N}$. Then, $(T-p)R^2$ asymptotically follows a χ^2 distribution with p degrees of freedom under the null hypothesis that there is no autocorrelation. In 'R', an implementation of the Breusch–Godfrey is available within the function `bgtest` in the 'lmtest' package, see [133].

Remark 17.6 (Example of Chapter 15, continued). Applying the validation procedure for serial correlation based on the Breusch–Godfrey test as described above to the example of Chapter 15, the null hypothesis, i.e., that there is no serial correlation of order $1, 2, \dots, 10$, is not rejected at a five percent level in 93.8 percent of all cases. Again, this is an indicator that our model with trends for weights and death probabilities fits the data suitably

Beyond serial correlation, it may be interesting to look at serial effects over different death causes and age categories as there may be causalities between a reduction in deaths due to certain death causes and a possibly lagged increase in different ones. Note that we already remove a lot of dependence via time-dependent weights and death probabilities. As illustrated in Figure 1.1, such serial effects are visible in the case of mental and behavioural disorders and circulatory diseases. In the context of health care systems it is crucial to pay attention to such dependent developments as, for example, duration and costs of geriatric care heavily depend on underlying diseases.

17.4 Validation via risk factor realisations

In our annuity model, risk factors Λ are assumed to be independent and identically gamma distributed with mean one and variance σ_k^2 for every $k \in \{1, \dots, K\}$, $\Lambda_k(1), \dots, \Lambda_k(t)$. Based on these assumptions, we can use estimates for risk factor realisations λ to judge whether our model adequately fits the data. These estimates can either be obtained via MCMC based on the maximum a posteriori setting or by Equations (12.42) or (12.48). Given estimates $\hat{\lambda} := \hat{\lambda}^{\mathrm{MAP}}$ of risk factor realisations λ we may use following two different approaches.

First, for each $k \in \{1, \dots, K\}$, we may check whether estimates $\hat{\lambda}_k(1), \dots, \hat{\lambda}_k(T)$ suggest a rejection of the null hypothesis that they are sampled from a gamma distribution with mean one and variance σ_k^2. The classical way is to use the Kolmogorov–Smirnov test, see, for example, Lehmann and Romano [79, Chapter 6.13] and the references therein, as well as Footnote 71. In 'R' an implementation of this test is provided by the `ks.test` function, see [99]. The null hypotheses is rejected as soon as the test statistic $\sup_{x \in \mathbb{R}} |F_T(x) - F(x)|$ exceeds the corresponding critical value where F_T denotes the empirical distribution function of samples $\hat{\lambda}_k(1), \dots, \hat{\lambda}_k(T)$ and where F denotes the gamma distribution function with mean one and variance σ_k^2. Secondly, we can test whether the independence assumption amongst $\Lambda_1(t), \dots, \Lambda_K(t)$ for each $t \in \{1, \dots, T\}$ can be accepted via some non-parametric test for independence as, for example, shown in Lehmann and Romano [79, Chapter 6.8] and Feuerverger [47].

Remark 17.7 (Example of Chapter 15, continued). Testing whether risk factor realisations are sampled from a gamma distribution via the Kolmogorov–Smirnov test as described above gives acceptance of the null hypothesis for all ten risk factors on all suitable levels of significance. Note that, as we fit risk factors to given data, it is not surprising that all null hypotheses are accepted.

17.5 Model selection

As briefly discussed in Remark 12.18, our proposed setup may lead to models with several hundred parameters and may therefore be over-parametrised. Nevertheless, the modelling setup always depends on the ultimate goal. For example, if the development of all death causes is of interest, then a reduction of risk factors is not wanted. On the contrary, in the context of annuity portfolios several risk factors may be merged to one risk factor as their contributions to the risk of the total portfolio are small. To decide which model to use, model choice criteria, as described below, should be used. In our case, we have the problem that a reduction in risk factors leads to a different data structure and, therefore, information criteria cannot be applied straight away. In this section we describe some approaches how the problem of selecting a suitable model can be addressed.

First we give a short recap of some major information criteria for given model parameters θ with likelihood function $\ell(n\,|\,\theta)$ and with corresponding maximum likelihood estimate $\hat{\theta} := \hat{\theta}^{\mathrm{MLE}}$. Upper case Θ denotes estimators in a Bayesian setting corresponding to θ. The classical *Akaike information criterion (AIC)*, see Akaike [5], based on the Kullbeck–Leibler mean information, is given by

$$\mathrm{AIC} := 2q - 2\log \ell\big(n\,\big|\,\hat{\theta}\big)$$

where $q = K + 4A + 4AK$ is the number of model parameters in our setup and $\ell(n\,|\,\theta)$ is the likelihood function given by Equation (12.52), evaluated at its maximum $\hat{\theta}$. The *Bayesian information criterion (BIC)*, also called Schwarz' information criterion, gives another asymptotic criterion for model selection. Invoking Robert [102, Section 7], it is defined as

$$\mathrm{BIC} := q(\log T + \log 2\pi) - 2\log \ell\big(n\,\big|\,\hat{\theta}\big)$$

where, in addition, T denotes the sample size of of the data n. A Bayesian alternative to the two criteria described above is the *deviance information criterion (DIC)*. It is defined as, see Robert [102, Section 7] and the references therein,

$$\mathrm{DIC} := 2\left(\mathbb{E}[D(\Theta)\,|\,N = n] - D(\mathbb{E}[\Theta\,|\,N = n])\right) + D(\mathbb{E}[\Theta\,|\,N = n])\,, \tag{17.8}$$

given data $N = n$ and deviance

$$D(z) := -2\log \ell(n\,|\,z) + C\,, \quad z \in \mathbb{R}^k\,,$$

with C being a constant, common to all candidate models, which may therefore be chosen arbitrarily. The close relationship of DIC to AIC is obvious. Expectations in (17.8) can be approximated using MCMC samples from parameter estimation. All the above information criteria have a term which penalises a higher number of model parameters, i.e., a measure of complexity, and they have a term which rewards for high values of the likelihood function, i.e., a measure of goodness of fit. Finally, we choose the model specification with lowest AIC, BIC or DIC. Using these approaches and the likelihood function given in (12.52), we can now select amongst different parameter families of weights and death probabilities, see Remark 12.12, for which AIC, BIC or DIC is minimised. We can also use AIC, BIC and DIC to select an optimal number of risk factors in our model under the premise that we are just interested in the number of people dying and not in the development of certain death causes. Therefore, for notational convenience, define $N^* := (N^*_{a,g}(t)) \in \mathbb{N}_0^{A\times 2\times T}$ with

$$N^*_{a,g}(t) = \sum_{k=1}^{K} N_{a,g,k}(t)\,,$$

for all $a \in \{1,\dots,A\}$, $g \in \{\mathrm{f},\mathrm{m}\}$, $k \in \{1,\dots,K\}$ and $t \in \{1,\dots,T\}$. We can then derive the likelihood function corresponding to N^* via convolution of the joint likelihood function (12.52). In applications we suggest to use fast Fourier transform (FFT) methods to derive these convolutions. Again, we may choose the number of risk factors such that AIC, BIC or DIC is minimised.

Chapter 18

Conclusion to Risk Aggregation with Applications to Credit and Life Insurance

We develop a model, based on the collective risk model extended CreditRisk$^+$, to derive loss distributions of annuity and life insurance portfolios over one period. Death probabilities are incorporated stochastically into the model and dependence is introduced via common stochastic risk factors. Yet, there exists a fast and numerically stable algorithm to derive loss distributions exactly, even for large portfolios. Furthermore, it is possible to derive various risk measures of the total loss distribution exactly, including value at risk and expected shortfall. Such a risk management tool is required by many regulators in the financial industry. We provide various estimation procedures based on publicly available data. Methods range from matching of moments, maximum a posteriori to maximum likelihood. The latter two require the use of Markov chain Monte Carlo due to high dimensionality in common settings. We briefly analyse different sources of risk which are captured by our model. Based on Australian mortality data from 1997 to 2011, we provide a real world example with corresponding estimation results. Furthermore, we show more applications of our annuity model including scenario analysis, as well as mortality forecasts and population forecasts as otherwise mortality is overestimated. In particular, we see that it is crucial to consider mortality trends in our annuity model when considering long-term forecasts. For completeness, we give different model validation techniques and briefly recall some model selection tools. Model validation techniques suggest that our model suitably fits Australian data. In the appendix we give scope to the most general version of our model with multi-level dependence structures where estimation procedures are subject to current research.

To summarise, our approach provides a useful risk management tool to analyse annuity and life insurance portfolios where mortality is modelled stochastically. The model allows for various other applications, including forecasts. Besides its complexity, it is still easy to interpret and easy to explain—also to non-mathematicians—as the concept of risk factors is common in economics.

Chapter 19

Appendix to Risk Aggregation with Applications to Credit and Life Insurance

This chapter deals with several issues. In Section 19.1 we provide the most general form of extended CreditRisk$^+$ and recall the corresponding algorithm for deriving loss distributions exactly. In Section 19.2 we give an introductory example which should convince the reader that multiple deaths in our annuity model are not a major issue. In Section 19.3 we list all estimates of the real world example given in Chapter 15. Finally, in Section 19.4 we provide Australian male and female life tables for 2013 based on Example 16.17.

19.1 Extended CreditRisk$^+$

19.1.1 General model and Panjer's recursion

As mentioned in Section 11.3, there exist several extensions to our annuity model. Therefore, we shortly introduce the most general case of extended CreditRisk$^+$ based on the results in Schmock [111, Section 6]. Estimation procedures for the general model are subject to future research. Besides risk groups and dependence scenarios, interpretations of all quantities stay the same as in Chapter 11. Risk groups are used to model joint deaths of several policyholders simultaneously. Further dependence is introduced via a linear dependence structure amongst death causes and via dependence scenarios. For notational purposes, note that the usage of the term risk factor slightly changes in the general case in contrast to the independent case, see Section 11.2. We provide definitions for a portfolio of annuity payments which need not be paid in the case of death to get the loss S' corresponding to the sum S in Definition 11.8.

Definition 19.1 (Extended annuity model, quantities)**.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that all following random variables are well-defined. Then, we assume the following:

(a) Let $1, \ldots, m$, with $m \in \mathbb{N}$, denote all policyholders and let the collection G denote non-empty subsets of all policyholders which are subject to joint death where, for each $g \in G$, the death probability is given by q_g^* and set $q_g := -\log(1 - q_g^*)$.

(b) Consider $C \in \mathbb{N}$ non-idiosyncratic death cause intensities $\Lambda_1, \dots, \Lambda_C$ as well as $K \in \mathbb{N}$ risk factors $R_1, \dots, R_K$.

(c) Consider a non-empty finite set $\mathcal{J}$ of dependence scenarios and a probability distribution on $\mathcal{J}$ with corresponding random variable J.

(d) For each dependence scenario $j \in \mathcal{J}$, consider a $(C+1) \times (K+1)$-dimensional matrix $A_j = (a^j_{c,k})_{c \in \{0,\dots,C\}, k \in \{0,\dots,K\}}$ with non-negative entries where $a^j_{0,k} = 0$ for all $j \in \mathcal{J}$ and $k \in \{1, \dots, K\}$.

(e) For the random matrix $A_J := \sum_{j \in \mathcal{J}} A_j 1_{\{J=j\}}$, non-negative death cause intensities $\Lambda_1, \dots, \Lambda_C$ are given by

$$\Lambda_c = a^J_{c,0} + \sum_{k=1}^{K} a^J_{c,k} R_k \,, \quad c \in \{1, \dots, C\} \,.$$

(f) Correspondingly, for all $g \in G$ and $j \in \mathcal{J}$, let idiosyncratic weights be denoted by $w_{0,g,j} \in [0,1]$ and let non-idiosyncratic weights be denoted by $w_{c,g,j} \in [0,1]$, for $c \in \{1, \dots, C\}$ such that $\sum_{c=0}^{C} w_{c,g,j} = 1$.

(g) Let the $\mathbb{N}_0$-valued random variables $N_{c,g,j}$ denote the number of deaths due to death cause $c \in \{0, \dots, C\}$ of risk group $g \in G$, as well as scenario $j \in \mathcal{J}$ and define

$$N_{c,g} := \sum_{j \in \mathcal{J}} N_{c,g,j} 1_{\{J=j\}} \,.$$

(h) For every group $g \in G$, every death cause $c \in \{0, \dots, C\}$, every dependence scenario $j \in \mathcal{J}$ and dimension $d \in \mathbb{N}$, let the $(\mathbb{N}_0^d)^g$-valued i.i.d. sequence $(Y_{c,g,i,j,n})_{i \in g}$ with $n \in \mathbb{N}$ denote the annuity payments to the respective policyholder in the following period which need not be paid due to death of death cause c. They are independent of all other random variables and they may also include the discounted actuarial reserve, as well as different lines of business, see Remark 11.6.

(i) The cumulative loss is then given by

$$S' := \sum_{j \in \mathcal{J}} 1_{\{J=j\}} \sum_{g \in G} \sum_{c=0}^{C} \sum_{n=1}^{N_{c,g,j}} \sum_{i \in g} Y_{c,g,i,j,n} \,.$$

Based on these quantities, we consider some probabilistic assumptions to guarantee the existence of a stable numerical algorithm to derive the loss distribution of S' exactly.

Definition 19.2 (Generalised annuity model)**.** Given Definition 19.1, we assume the following:

(a) Conditioned on J, the $\mathbb{N}_0$-valued random variables $(N_{0,g})_{g \in G}$ are independent from one another and from all other random variables and their joint distribution is given by

$$\mathbb{P}\Big(\bigcap_{g \in G} \{N_{0,g} = n_{0,g}\} \,\Big|\, J \Big) = \prod_{g \in G} e^{-q_g w_{0,g,J} a^J_{0,0}} \frac{(q_g w_{0,g,J} a^J_{0,0})^{n_{0,g}}}{n_{0,g}!} \quad \text{a.s.},$$

for all $n_{0,g} \in \mathbb{N}_0$ and $g \in G$.

(b) Risk factors $R_1, \dots, R_K$ are independent and gamma distributed with mean $e_k > 0$ and variance $\sigma_k^2 > 0$. For all $k \in \{1, \dots, K\}$, the degenerate case $\sigma_k^2 = 0$ is also allowed.

(c) Conditioned on the random variables $J, R_1, \dots, R_K$, the $\mathbb{N}_0$-valued random variables $(N_{c,g})_{c\in\{1,\dots,C\},g\in G}$ are independent and their joint distribution is given by

$$\mathbb{P}\Big(\bigcap_{g\in G}\bigcap_{c=1}^{C}\{N_{c,g} = n_{c,g}\}\,\Big|\,J, R_1, \dots, R_C\Big) = \prod_{g\in G}\prod_{c=1}^{C} e^{-q_g\, w_{c,g,J}\,\Lambda_c}\,\frac{(q_g\, w_{c,g,J}\,\Lambda_c)^{n_{c,g}}}{n_{c,g}!} \quad \text{a.s.},$$

for all $n_{c,g} \in \mathbb{N}_0$ with $c \in \{1, \dots, C\}$ and $g \in G$.

(d) The random variables $R_1, \dots, R_K$ and the scenario random variable J are independent.

(e) $\mathbb{E}[w_{0,g,J}\, a_{0,0}^J + \sum_{c=1}^{C} w_{c,g,J}\,\Lambda_c] = 1$.

With such a setting, death cause intensities can be dependent by means of a linear dependence structure and dependence scenarios. In particular, many correlation structures amongst death causes are possible to achieve. Also, in this more general case, there exists a numerically stable algorithm which is based on iterated Panjer's recursion to derive loss distributions exactly, similarly as in Lemma 11.19. For further details see Schmock [111].

Definition 19.3. Given the generalised annuity model of Definitions 19.1 and 19.2, for notational convenience in the next lemma, first define probability distributions of group losses

$$q_{c,g,j,\nu} := \sum_{\substack{\mu=(\mu_i)_{i\in g}\in(\mathbb{N}_0^d)^g \\ \sum_{i\in g}\mu_i=\nu}} \mathbb{P}(Y_{c,g,i,j,1} = \mu_i \text{ for all } i \in g)\,,$$

for all $c \in \{0, \dots, C\}$, $g \in G$, $j \in \mathcal{J}$ and $\nu \in \mathbb{N}_0^d$. Then, define the cumulative Poisson intensity

$$\lambda_{j,k,\nu} := \sum_{g\in G} q_g^* \sum_{c=0}^{C} w_{c,g,j}\, a_{c,k}^j\, q_{c,g,j,\nu}\,,$$

for loss size $\nu \in \mathbb{N}_0^d \setminus \{0\}$ due to risk factor $k \in \{0, \dots, K\}$ and dependence scenario $j \in \mathcal{J}$, as well as, correspondingly, the cumulative Poisson intensity for non-zero losses

$$\bar{\lambda}_{j,k} := \sum_{\nu\in\mathcal{S}_{j,k}} \lambda_{j,k,\nu}\,,$$

where $\mathcal{S}_{j,k} := \{\nu \in \mathbb{N}_0^d \setminus \{0\} \,|\, \lambda_{j,k,\nu} > 0\}$. If $\bar{\lambda}_{j,k} > 0$ for dependence scenario $j \in \mathcal{J}$ and $k \in \{0, \dots, K\}$, define

$$q_{j,k,\nu} := \begin{cases} \lambda_{j,k,\nu}/\bar{\lambda}_{j,k} & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\}\,, \\ 0 & \text{for } \nu = 0 \in \mathbb{N}_0^d\,, \end{cases}$$

as well as, if $\bar{\lambda}_k = 0$,

$$q_{j,k,\nu} := \begin{cases} 0 & \text{for all } \nu \in \mathbb{N}_0^d \setminus \{0\}\,, \\ 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d\,. \end{cases}$$

Finally, for all $j \in \mathcal{J}$ and $k \in \{0, \dots, K\}$, $p_{j,k} := \bar{\lambda}_{j,k}\sigma_k^2/(e_k + \bar{\lambda}_{j,k}\sigma_k^2) \in [0,1)$ as well as

$$\lambda_j := \bar{\lambda}_{j,0} + \sum_{k=1}^{K} \bar{\lambda}_{j,k} \frac{e_k^2}{e_k + \bar{\lambda}_{j,k}\sigma_k^2} c(p_{j,k}) \,,$$

where

$$c(p) := \sum_{n\in\mathbb{N}} \frac{p^{n-1}}{n} = \begin{cases} -\frac{\log(1-p)}{p} & \text{for } p \in (0,1)\,, \\ 1 & \text{for } p = 1\,. \end{cases}$$

Note that all definitions also work in the degenerate case $\sigma_k^2 = 0$ for all $k \in \{1, \dots, K\}$ as well.

Using Definitions 19.1 and 19.2, we can obtain a generalisation of Lemma 11.19 where an iterated Panjer algorithm is used to derive loss distributions of our annuity model.

Lemma 19.4. *Given the annuity model of Definitions 19.1 and 19.2, there exists a numerically stable algorithm based on iterated Panjer's recursion which allows an exact computation of the probability distribution of S' up to every desired cumulative probability. More precisely, $\mathbb{P}(S' = 0) = \sum_{j\in\mathcal{J}} \exp\big(\lambda_j (c_{j,0} - 1)\big)\mathbb{P}(J = j)$ and, recursively,*

$$\mathbb{P}(S' = \nu) = \sum_{j\in\mathcal{J}} d_{j,\nu}\mathbb{P}(J = j)\,, \quad \nu = (\nu_1, \dots, \nu_d) \in \mathbb{N}_0^d \setminus \{0\}\,, \tag{19.5}$$

with $d_{j,0} = \exp\big(\lambda_j (c_{j,0} - 1)\big)$ and

$$d_{j,\nu} = \frac{\lambda_j}{\nu_i} \sum_{\substack{n\in\mathbb{N}_0^d \\ 0<n\leq\nu}} n_i c_{j,n} d_{j,\nu-n}\,, \quad \nu \in \mathbb{N}_0^d \setminus \{0\}\,, \tag{19.6}$$

where $i \in \{1, \dots, d\}$ can be chosen arbitrarily such that $\nu_i \neq 0$ and where $0 < n \leq \nu$ is meant in the sense of Footnote 38. Moreover, if $\lambda_j > 0$ for scenario $j \in \mathcal{J}$,

$$c_{j,\nu} = \frac{1}{\lambda_j}\left(\bar{\lambda}_{j,0} q_{j,0,\nu} + \sum_{k=1}^{K} b_{j,k,\nu}\bar{\lambda}_{j,k} \frac{e_k^2}{e_k + \bar{\lambda}_{j,k}\sigma_k^2} c(p_{j,k})\right), \quad \nu \in \mathbb{N}_0^d\,, \tag{19.7}$$

where, for all $k \in \{1, \dots, K\}$, $b_{j,k,0} = q_{j,k,0}\, c(p_{j,k} q_{j,k,0})/c(p_{j,k})$ and

$$b_{j,k,\nu} = \frac{1}{1 - p_{j,k} q_{j,k,0}}\left(\frac{q_{j,k,\nu}}{c(p_{j,k})} + \frac{p_{j,k}}{\nu_i} \sum_{\substack{n\in\mathcal{S}_{j,k}, \\ n\leq\nu}} (\nu_i - n_i) q_{j,k,n} b_{j,k,\nu-n}\right), \tag{19.8}$$

for all $\nu \in \mathbb{N}_0^d \setminus \{0\}$, where $i \in \{1, \dots, d\}$ is chosen such that $\nu_i \neq 0$. Conversely, if $\lambda_j = 0$ for dependence scenario $j \in \mathcal{J}$,

$$c_{j,\nu} = \begin{cases} 0 & \text{for } \nu \in \mathbb{N}_0^d \setminus \{0\}\,, \\ 1 & \text{for } \nu = 0 \in \mathbb{N}_0^d\,. \end{cases}$$

Proof. A detailed derivation is given in Schmock [111, Section 6.6]. □

19.1.2 Pseudo implementation of the algorithm

In this section we provide a pseudo implementation of the iterated Panjer algorithm within CreditRisk$^+$ which is described in Lemma 19.4. The algorithm for the simple annuity model works simultaneously, see Lemma 11.19, in which case we have $G = \{\{1\}, \dots, \{m\}\}$, one dependence scenario j, as well as $K = C$ with $a^J_{c,k} = 1_{\{c\}}(k)$ for all $c, k \in \{1, \dots, C\}$.

Input: Quantities given in Definitions 19.1 and 19.2.
Output: Exact distribution of S' up to value $\nu^* \in \mathbb{N}_0^d$ exceeding some cumulative level $\delta \in (0,1)^d$, i.e., $\mathbb{P}(S' \leq \nu^*) \geq \delta$.

```
for c ∈ {0, ..., C}, g ∈ G, j ∈ 𝒥 and ν ∈ ℕ_0^d, see Definition 19.3, do
    derive q_{c,g,j,ν};
    derive λ_{j,k,ν};
    derive λ̄_{j,k}, q_{j,k,ν};
    derive p_{j,k};
    derive λ_j;
end
for j ∈ 𝒥 and k ∈ {0, ..., K} do
    initialise b_{j,k,0} = q_{j,k,0} c(p_{j,k} q_{j,k,0}) / c(p_{j,k});
    initialise c_{j,0}, see Equation (19.7);
    initialise d_{j,0} = exp(λ_j (c_{j,0} − 1));
end
initialise ℙ(S' = 0) = Σ_{j∈𝒥} exp(λ_j (c_{j,0} − 1)) ℙ(J = j);
for ν = 0 to ν*, see Footnote 38, do
    for j ∈ 𝒥 and k ∈ {0, ..., K} do
        derive b_{j,k,ν}, as well as c_{j,ν} and store them;
        derive d_{j,ν};
    end
    derive ℙ(S' = ν) = Σ_{j∈𝒥} d_{j,ν} ℙ(J = j) and go to next ν;
end
```

Algorithm 19.1: Extended CreditRisk$^+$ algorithm

Remarks 19.9. (Some notes on Algorithm 19.1).

(a) Note that for distributions $Y_{c,g,i,j,1}$ with infinite support, calculation of $q_{c,g,j,\nu}$ has to be stopped at some suitable level of approximation.

(b) For the derivation of $b_{j,k,\nu}$ and $d_{j,\nu}$ we have to recall previous values $b_{j,k,n}$, $c_{j,n}$ and $d_{j,n}$ for certain $0 \leq n \leq \nu$. Thus, these values have to be stored for further usage in this recursive procedure.

(c) In the one-dimensional case when going to the next $\nu \in \mathbb{N}_0$, we just take the consecutive integer. In the multi-dimensional case when going to the next $\nu \in \mathbb{N}_0^d$, one has to go through the space $\mathbb{N}_0^d$ such that no values required for the recursions in (19.6) and (19.8) are left out. Of course, for $d \geq 2$, there are multiple possibilities of how to jump through $\mathbb{N}_0^d$. Note that this procedure quickly becomes time-consuming for higher dimensions which is why we suggest to choose a maximum dimension of $d = 3$.

19.2 Introductory example justifying multiple deaths

The main purpose of this simple example is to convince the reader that the setup with multiple deaths is suitable for large portfolios and that it gives accurate results, combined with an highly efficient algorithm. We use notation and assumptions as introduced in Chapter 11. Consider a portfolio of $m = 1\,000$ policyholders with annuity payments made continuously over time. In the case of survival, independent annuity payment $X_1, \ldots, X_m$ follow a log-normal distribution with parameters $\mu = 4$ and $\sigma = 0.5$, i.e., roughly with mean 61.87 and standard deviation 42.52. The independent random variables $Y_1, \ldots, Y_m$, i.e., the amounts which need not be paid in the case of death, take the form $Y_i = U_i X_{i,1}$ for all $i \in \{1, \ldots, m\}$ where $X_{i,1}$ and X_i share the same distribution, as well as where U_i is uniformly distributed on $(0, 1]$ and independent of all other random variables, also see Remark 11.6(c). Moreover, all policyholders $i \in \{1, \ldots, m\}$ share an annual death probability of $q^* = 0.05$.

Table 19.1: Value at risk (top) and expected shortfall (bottom) for various levels δ of gains S^* and S in our example with N_i^* being Bernoulli distributed and N_i being Poisson distributed with $\mathbb{P}(N_i = 0) = 1 - q^*$ (Poisson) and $\mathbb{E}[N_i] = q^*$ (Poisson'), indicated by prime, based on 10 000 simulations. 95 percent binomial confidence intervals are given in brackets. Bounds are given for S^* with $\mathbb{P}(N_i = 0) = 1 - q^*$ based on the Kolmogorov–Smirnov distance and the Wasserstein metric.

	Bernoulli	Poisson	Poisson'	lower bound	upper bound
level δ	value at risk $q_\delta(\cdot)$				
0.01	1 007.87 (−19.1;+12.2)	1 005.16 (−16.3;+21.5)	985.88 (−16.6;+15.4)	827.59	1 061.85
0.05	1 174.09 (−9.6;+10.1)	1 170.17 (−9.4;+9.9)	1 138.79 (−9.8;+8.9)	1 151.59	1 197.17
0.15	1 325.84 (−7.3;+8.3)	1 324.91 (−7.0;+7.4)	1 292.84 (−7.0;+7.8)	1 315.73	1 334.80
0.85	1 922.95 (−10.0;+11.0)	1 925.94 (−7.9;+9.7)	1 881.84 (−9.2;+12.0)	1 911.71	1 934.97
0.95	2 114.15 (−11.9;+14.7)	2 139.00 (−13.8;+16.9)	2 077.45 (−15.5;+16.0)	2 089.59	2 144.82
0.99	2 333.72 (−18.0;+22.8)	2 373.00 (−19.2;+30.3)	2 300.47 (−20.9;+31.7)	2 257.98	2619.87
level δ	expected shortfall $\mathrm{ES}_\delta[\cdot]$				
0.01	1 631.45	1 634.92	1 592.58	1 624.19	1 638.71
0.05	1 653.46	1 657.15	1 614.30	1 645.89	1 661.02
0.15	1 699.61	1 704.01	1 660.16	1 691.15	1 708.07
0.85	2 090.60	2 105.27	2 051.30	2 042.66	2 138.53
0.95	2 257.61	2 285.17	2 222.59	2 113.82	2 401.41
0.99	2 483.49	2 512.10	2 452.86	1 764.52	3 202.45

Based on Monte Carlo, our aim is to compare empirical distributions of the following two models: For the first model, i.e., the exact model, $N_1^*, \ldots, N_m^*$ denote Bernoulli random variables with $\mathbb{P}(N_i^* = 1) = 1 - \mathbb{P}(N_i^* = 0) = q^*$ and where $X_i = X_{i,1}$ a.s. for all policyholders $i \in \{1, \ldots, m\}$. Thus, we are interested in the sums $S^* = \sum_{i=1}^m \sum_{j=1}^{N_i^*} Y_i = \sum_{i=1}^m N_i^* Y_i$ and

$L^* = \sum_{i=1}^m X_i - S^*$. Note that, for sums and death counts, an asterisk indicates the model with Bernoulli distributed deaths. For the second model, number of deaths $N_1, \ldots, N_m$ are Poisson random variables with $\mathbb{P}(N_i = 0) = 1 - q^*$ for all policyholders $i \in \{1, \ldots, m\}$ and we are interested in the sums $S = \sum_{i=1}^m \sum_{j=1}^{N_i} Y_{i,j}$ and $L = \sum_{i=1}^m X_i - S$ where $(Y_{i,j})_{j \in \mathbb{N}}$ are independent copies of Y_i. Table 19.1 lists quantiles for S^* and S using 10 000 simulations with the given model specifications. Quantiles for the model with Poisson distributed deaths with specification $\mathbb{E}[N_i] = q^*$, indicated by a prime, are also listed. Obviously, the latter specification does not show a good fit for quantiles in the left tail of S which is why we suggest to use specification $\mathbb{P}(N_i = 0) = 1 - q^*$ for most applications. In brackets, based on our simulation, conservative 95 percent binomial confidence intervals for value at risk estimates are given, i.e., intervals such that with a probability of at least 95 percent the true value of value at risk is in this interval. The method to calculate these confidence intervals can be found in Shevchenko [114, Section 3.2.1]. Based on the empirical distributions of S^* and S, the estimated Kolmogorov–Smirnov distance is 0.0089 and the estimated Wasserstein distance with Euclidean metric is 7.1897.[71] Based on these distances we can derive bounds for value at risk and expected shortfall for the model with Poisson distributed deaths as shown in Tables 19.1 and 19.3, see Schmock [111, Section 7]. Note that these bounds are just estimates as we use simulation. As we can see in Table 19.1, the fit of our model with multiple deaths is good on all levels for value at risk, as well as expected shortfall given scaling $\mathbb{P}(N_i = 0) = 1 - q^*$ of Definition 11.2.

Table 19.2: Kolmogorov–Smirnov and Wasserstein distance between empirical loss distributions L^* and L based on 10 000 simulations with Bernoulli distributed deaths N_i^* and Poisson distributed deaths N_i with $\mathbb{P}(N_i = 0) = 1 - q^*$ using three different dependence assumptions.

	Poisson indep.	Poisson comon.	Poisson countermon.
Kolmogorov–Smirnov distance	0.0111	0.0587	0.0850
Wasserstein distance	17.2716	220.6057	244.2475

Calculating the total loss $L = \sum_{i=1}^m X_i - S$ now raises the question which form of dependence we should assume between $\sum_{i=1}^m X_i$ and S. We try three types of dependence: independence, comonotonicity and countermonotonicity. For the notions of comonotonicity and countermonotonicity and their applications in risk management see, for example, McNeil, Frey and Embrechts [85]. To achieve the results for comonotonicity and countermonotonicity, we order all simulations of $\sum_{i=1}^m X_i$ and S and then simply subtract them. For comonotonicity both simulations are ordered ascending and for countermonotonicity one of them has to be ordered in a descending manner. For the results based on independence, we simply add the two empirical distributions of $\sum_{i=1}^m X_i$ and S as they are simulated independently in the case of Poisson distributed deaths. Table 19.2 illustrates the fit for each form of dependence between the empirical distributions of $\sum_{i=1}^m X_i$ and S. As already illustrated earlier, it

[71] For information and definitions of Kolmogorov-Smirnov and Wasserstein distances—or metrics, more precisely—see Schmock [111]. For the derivation of the Kolmogorov–Smirnov distance we use the function ks.test in 'R' of the 'stats' package [99] and for the derivation of the Wasserstein distance we use the function emd in 'R' of the 'emdist' package [122]. The latter calculates the so-called earth mover's distance which is equivalent to the Wasserstein distance. The emd-function struggles with the high number of simulations which is why we derive an estimate with just half the simulation points.

shows that the fit of our model with multiple deaths compared to the exact model with Bernoulli distributed deaths is good, in particular under the assumption of independence between total loss $\sum_{i=1}^{m} X_i$ and S with scaling $\mathbb{P}(N_i = 0) = 1 - q^*$ for all $i \in \{1, \ldots, m\}$.

Table 19.3: Value at risk, expected shortfall for various levels for L^* and L with Bernoulli distributed deaths N_i^* and Poisson distributed deaths N_i with with scaling $\mathbb{P}(N_i = 0) = 1 - q^*$ and the independence assumption based on 10 000 simulations. 95 percent binomial confidence intervals are given in brackets. Bounds are given for L^* based on the Kolmogorov–Smirnov distance and the Wasserstein metric.

	Bernoulli	Poisson indep.	lower bound	upper bound
level δ	value at risk $q_\delta(\cdot)$			
0.01	57 779.56 (−72.1;+81.0)	57 652.29 (−97.3;+94.1)	na	58 086.67
0.05	58 488.73 (−46.4;+48.9)	58 446.92 (−51.1;+45.6)	58 368.24	58 602.89
0.15	59 147.69 (−29.6;+27.3)	59 104.49 (−24.1;+32.0)	59 093.03	59 195.15
0.85	61 335.62 (−34.2;+30.1)	61 346.01 (−29.3;+37.2)	61 282.05	61 381.12
0.95	61 968.83 (−39.7;+44.6)	62 010.18 (−35.4;+46.6)	61 868.21	62 106.98
0.99	62 716.55 (−63.7;+76.5)	62 771.25 (−67.8;+71.5)	62 375.86	na
level δ	expected shortfall $\mathrm{ES}_\delta[\cdot]$			
0.01	60 271.07	60 268.95	60 253.62	60 288.52
0.05	60 357.74	60 359.50	60 339.55	60 375.92
0.15	60 533.86	60 539.55	60 513.54	60 554.18
0.85	61 921.29	61 958.56	61 806.14	62 036.43
0.95	62 545.90	62 604.26	62 200.47	62 891.33
0.99	63 695.59	63 781.38	61 968.44	65 422.75

Obviously, the model where $\sum_{i=1}^{m} X_i$ and S are independent fits better compared to the models using comonotonicity and countermonotonicity. In Table 19.3 we therefore list quantiles for the simulated total portfolio loss L with Poisson distributed deaths N_i under the assumption of independence against the simulated portfolio loss L^* with Bernoulli distributed deaths N_i^*. As in Table 19.1, conservative 95 percent binomial confidence intervals for value at risk estimates are given in brackets. Estimated Kolmogorov–Smirnov distance and Wasserstein distance with Euclidean metric between L and L^* are then used to drive bounds for quantiles and expected shortfall of the loss with Poisson distributed deaths as given in Schmock [111, Section 7].

One possible justification why independence fits very well is the comparison of standard deviations of the different approaches. Table 19.4 shows standard deviations of our model with different dependence assumptions based on our simulations, as well as true standard deviations, given by Formulas (19.10) and (19.11). The cases of comonotonicity and countermonotonicity between $\sum_{i=1}^{m} X_i$ and S always give lower and upper bounds, respectively, for the variance of its sum, see for example Cheung and Vanduffel [26]. Note the almost perfect fit between Monte Carlo standard deviations and true standard deviations. To derive true

standard deviations of the total loss, recall that L^* denotes the total portfolio loss under the assumption that N_i^* is Bernoulli distributed and that L denotes the total portfolio loss where N_i is Poisson distributed with $\mathbb{P}(N_i = 0) = 1 - q^*$ for all $i \in \{1, \dots, m\}$. If $\sum_{i=1}^m X_i$ and S are independent, then straight-forward calculation yields

$$\begin{aligned}\operatorname{Var}(L^*) &= \operatorname{Var}\left(\sum_{i=1}^m X_i - \sum_{i=1}^m N_i U_i X_i\right) = \sum_{i=1}^m \operatorname{Var}(X_i(1 - N_i U_i)) \\ &= \sum_{i=1}^m \left(\mathbb{E}\big[X_i^2\big]\, \mathbb{E}\big[(1 - N_i U_i)^2\big] - \mathbb{E}[X_i]^2 \left(1 - \frac{q^*}{2}\right)^2 \right).\end{aligned}$$

As $\mathbb{E}[(1 - N_i U_i)^2] = 1 - q^* + \mathbb{E}[N_i^2]\, \mathbb{E}[U_i^2] = 1 - 2q^*/3$ for all $i \in \{1, \dots, m\}$, we finally get

$$\operatorname{Var}(L^*) = \sum_{i=1}^m \left(\mathbb{E}[X_i^2] \left(1 - \frac{2q^*}{3}\right) - \mathbb{E}[X_i]^2 \left(1 - \frac{q^*}{2}\right)^2 \right). \tag{19.10}$$

Correspondingly, for the case with Poisson distributed deaths, we have

$$\operatorname{Var}(L) = \operatorname{Var}\left(\sum_{i=1}^m X_i - \sum_{i=1}^m \sum_{j=1}^{N_i} Y_{i,j}\right) = \sum_{i=1}^m \left(\operatorname{Var}(X_i) - \operatorname{Var}\left(\sum_{j=1}^{N_i} Y_{i,j}\right)\right).$$

As $\mathbb{E}[N_i] = \operatorname{Var}(X_i) = -\log(1 - q^*) =: q$ for all $i \in \{1, \dots, m\}$, Wald's formula for random sums gives, see, for example, Schmock [111, Section 4.7.1],

$$\begin{aligned}\operatorname{Var}\left(\sum_{j=1}^{N_i} Y_{i,j}\right) &= \mathbb{E}[N_i]\, \operatorname{Var}(Y_{i,j}) + \mathbb{E}[Y_{i,j}]^2 \operatorname{Var}(N_i) \\ &= q\left(\operatorname{Var}(U_i X_{i,1}) + \mathbb{E}[U_i X_{i,1}]^2 \right) = \frac{q}{3}\, \mathbb{E}\big[X_i^2\big]\ ,\end{aligned}$$

which, substituted in the previous equation, implies

$$\operatorname{Var}(L) = \sum_{i=1}^m \left(\mathbb{E}[X_i^2] \left(1 - \frac{q}{3}\right) - \mathbb{E}[X_i]^2 \right). \tag{19.11}$$

Table 19.4: Empirical standard deviations (stdev.) of L^* and L for different dependence assumptions based on 10 000 simulations, as well as true standard deviations.

	Bernoulli exact	Poisson indep.	Poisson comon.	Poisson counter.
empirical stdev. of L^* and L	1 055.78	1 085.92	1 338.46	754.68
true stdev. of L^* and L	1 054.66	1 082.21	na	na

19.3 Real world MCMC estimation results

In this section we give estimation results for the real world example given in Chapter 15. We assume a setup with ten common stochastic risk factors and eight age groups for each

gender which gives 362 model parameters to be optimised. These parameters are estimated from given Australian death and population data taken from governmental websites, also see Chapter 15. Table 19.6 gives matching of moments estimates (Match. moments), as well as, Markov chain Monte Carlo mean estimates (MCMC mean) based on 30 000 samples within the maximum likelihood setting, standard deviations (Standard dev.), five percent and 95 percent quantiles (5% quantile, 95% quantile), mean acceptance probabilities (Accept. prob.) and standard errors (Standard error)[72] for all parameters. Mode estimates, i.e., parameters which give the highest value of the log-likelihood function, are not given as they are dominated by mean estimates. Note that risk factor variances for the matching of moments approach are estimated via (12.48) and (12.49). Evaluation time is roughly seven hours. If parallelised,

Table 19.5: Legend for age categories and death causes.

index	age category	death cause
0		not elsewhere (idio.)
1	50–54 years	infectious
2	55–59 years	neoplasms
3	60–64 years	endocrine
4	65–69 years	mental
5	70–74 years	nervous
6	75–79 years	circulatory
7	80–84 years	respiratory
8	85+ years	digestive
9		external
10		genitourinary

evaluation times of less than 20 minutes can be achieved. For notational convenience, we identify age categories and death causes with numbers as given in Table 19.5. Parameters for males are denoted by 'm' and for females by 'f'.

There are two remarkable observations. First, matching of moment estimates show very good results while being easy and quick to calculate. Secondly, risk factor variances are small which gives indication that our model and families for trends describe the given data reasonably well. For further discussion on adequacy of our model see Chapter 17.

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$\alpha_{1,\mathrm{m}}$	−4.4442	−4.4345	−4.4521	−4.4166	0.2367	0.0108	0.000454
$\alpha_{2,\mathrm{m}}$	−3.8659	−3.8523	−3.8667	−3.8374	0.2097	0.0090	0.000386
$\alpha_{3,\mathrm{m}}$	−3.3069	−3.2973	−3.3093	−3.2847	0.2277	0.0075	0.000315
$\alpha_{4,\mathrm{m}}$	−2.8063	−2.7997	−2.8117	−2.7879	0.2325	0.0072	0.000317
$\alpha_{5,\mathrm{m}}$	−2.3290	−2.3220	−2.3344	−2.3100	0.2215	0.0072	0.000325
$\alpha_{6,\mathrm{m}}$	−1.8701	−1.8615	−1.8735	−1.8496	0.2462	0.0073	0.000331
$\alpha_{7,\mathrm{m}}$	−1.4386	−1.4277	−1.4403	−1.4148	0.2152	0.0077	0.000352
$\alpha_{8,\mathrm{m}}$	−0.9461	−0.9353	−0.9485	−0.9229	0.2272	0.0079	0.000367
$\alpha_{1,\mathrm{f}}$	−4.9756	−4.9726	−4.9933	−4.9514	0.2190	0.0129	0.000527

[72] Defined as in Shevchenko [114, Section 2.12.2] with block size 50.

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$\alpha_{2,\mathrm{f}}$	−4.4794	−4.4739	−4.4910	−4.4562	0.2062	0.0107	0.000431
$\alpha_{3,\mathrm{f}}$	−4.0017	−3.9956	−4.0091	−3.9812	0.2078	0.0087	0.000346
$\alpha_{4,\mathrm{f}}$	−3.5073	−3.5017	−3.5147	−3.4883	0.2948	0.0079	0.000324
$\alpha_{5,\mathrm{f}}$	−2.9817	−2.9733	−2.9861	−2.9608	0.2097	0.0077	0.000341
$\alpha_{6,\mathrm{f}}$	−2.4342	−2.4248	−2.4375	−2.4124	0.2376	0.0076	0.000339
$\alpha_{7,\mathrm{f}}$	−1.8936	−1.8817	−1.8953	−1.8681	0.2206	0.0083	0.000387
$\alpha_{8,\mathrm{f}}$	−1.1795	−1.1678	−1.1820	−1.1535	0.2325	0.0087	0.000420
$\beta_{1,\mathrm{m}}$	−0.0256	−0.0263	−0.0275	−0.0251	0.2221	0.0007	0.000032
$\beta_{2,\mathrm{m}}$	−0.0322	−0.0331	−0.0341	−0.0321	0.2472	0.0006	0.000026
$\beta_{3,\mathrm{m}}$	−0.0361	−0.0367	−0.0375	−0.0358	0.1987	0.0005	0.000021
$\beta_{4,\mathrm{m}}$	−0.0358	−0.0362	−0.0370	−0.0354	0.2221	0.0005	0.000021
$\beta_{5,\mathrm{m}}$	−0.0338	−0.0342	−0.0350	−0.0334	0.2378	0.0005	0.000021
$\beta_{6,\mathrm{m}}$	−0.0293	−0.0298	−0.0306	−0.0291	0.2490	0.0005	0.000022
$\beta_{7,\mathrm{m}}$	−0.0229	−0.0236	−0.0245	−0.0228	0.2711	0.0005	0.000023
$\beta_{8,\mathrm{m}}$	−0.0104	−0.0111	−0.0120	−0.0102	0.3262	0.0005	0.000025
$\beta_{1,\mathrm{f}}$	−0.0224	−0.0226	−0.0241	−0.0212	0.1980	0.0009	0.000036
$\beta_{2,\mathrm{f}}$	−0.0265	−0.0269	−0.0281	−0.0257	0.2359	0.0007	0.000029
$\beta_{3,\mathrm{f}}$	−0.0275	−0.0278	−0.0288	−0.0269	0.2238	0.0006	0.000024
$\beta_{4,\mathrm{f}}$	−0.0286	−0.0289	−0.0298	−0.0280	0.2152	0.0005	0.000022
$\beta_{5,\mathrm{f}}$	−0.0278	−0.0283	−0.0292	−0.0274	0.1822	0.0005	0.000023
$\beta_{6,\mathrm{f}}$	−0.0265	−0.0271	−0.0279	−0.0263	0.3295	0.0005	0.000022
$\beta_{7,\mathrm{f}}$	−0.0209	−0.0216	−0.0225	−0.0208	0.2901	0.0005	0.000025
$\beta_{8,\mathrm{f}}$	−0.0077	−0.0084	−0.0095	−0.0075	0.2143	0.0006	0.000028
$u_{1,\mathrm{m},0}$	−4.0278	−4.0278	na	na	na	na	na
$u_{2,\mathrm{m},0}$	−4.3493	−4.3493	na	na	na	na	na
$u_{3,\mathrm{m},0}$	−4.7611	−4.7611	na	na	na	na	na
$u_{4,\mathrm{m},0}$	−4.8991	−4.8991	na	na	na	na	na
$u_{5,\mathrm{m},0}$	−4.7991	−4.7991	na	na	na	na	na
$u_{6,\mathrm{m},0}$	−4.7306	−4.7306	na	na	na	na	na
$u_{7,\mathrm{m},0}$	−4.6540	−4.6540	na	na	na	na	na
$u_{8,\mathrm{m},0}$	−4.2780	−4.2780	na	na	na	na	na
$u_{1,\mathrm{f},0}$	−4.1212	−4.1212	na	na	na	na	na
$u_{2,\mathrm{f},0}$	−4.3228	−4.3228	na	na	na	na	na
$u_{3,\mathrm{f},0}$	−4.1519	−4.1519	na	na	na	na	na
$u_{4,\mathrm{f},0}$	−4.2093	−4.2093	na	na	na	na	na
$u_{5,\mathrm{f},0}$	−4.2212	−4.2212	na	na	na	na	na
$u_{6,\mathrm{f},0}$	−4.2234	−4.2234	na	na	na	na	na
$u_{7,\mathrm{f},0}$	−4.1600	−4.1600	na	na	na	na	na
$u_{8,\mathrm{f},0}$	−3.9748	−3.9748	na	na	na	na	na
$u_{1,\mathrm{m},1}$	−4.8179	−4.8553	−5.0244	−4.6845	0.2260	0.1051	0.004719
$u_{2,\mathrm{m},1}$	−5.1691	−5.1501	−5.3070	−5.0039	0.1919	0.0929	0.004050
$u_{3,\mathrm{m},1}$	−5.2899	−5.2447	−5.4068	−5.0741	0.2383	0.0997	0.004436
$u_{4,\mathrm{m},1}$	−5.2882	−5.2653	−5.4115	−5.1127	0.2380	0.0898	0.004061
$u_{5,\mathrm{m},1}$	−5.2476	−5.2205	−5.3438	−5.1013	0.2327	0.0751	0.003330
$u_{6,\mathrm{m},1}$	−4.9477	−4.9002	−5.0082	−4.7819	0.2392	0.0681	0.003077
$u_{7,\mathrm{m},1}$	−5.0017	−4.8799	−4.9971	−4.7674	0.2412	0.0674	0.003006
$u_{8,\mathrm{m},1}$	−4.8736	−4.8231	−4.9188	−4.7221	0.2608	0.0598	0.002636
$u_{1,\mathrm{f},1}$	−5.3563	−5.2237	−5.4843	−4.9573	0.2286	0.1599	0.006951
$u_{2,\mathrm{f},1}$	−5.0857	−5.1281	−5.3450	−4.8967	0.2322	0.1356	0.005786
$u_{3,\mathrm{f},1}$	−4.9586	−4.9784	−5.1500	−4.8117	0.2922	0.1035	0.004239
$u_{4,\mathrm{f},1}$	−4.8918	−4.8716	−5.0059	−4.7326	0.2619	0.0837	0.003467
$u_{5,\mathrm{f},1}$	−5.0575	−5.0612	−5.1909	−4.9340	0.2321	0.0776	0.003347
$u_{6,\mathrm{f},1}$	−4.8571	−4.8297	−4.9449	−4.7115	0.2367	0.0712	0.003188

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$u_{7,\mathrm{f},1}$	−4.9602	−4.9158	−5.0160	−4.8045	0.1800	0.0657	0.003019
$u_{8,\mathrm{f},1}$	−5.0033	−4.9478	−5.0298	−4.8684	0.2185	0.0491	0.002216
$u_{1,\mathrm{m},2}$	−1.0292	−1.0775	−1.1960	−0.9680	0.2992	0.0688	0.003361
$u_{2,\mathrm{m},2}$	−0.9854	−0.9897	−1.0869	−0.8973	0.2319	0.0596	0.002924
$u_{3,\mathrm{m},2}$	−0.9965	−1.0066	−1.1109	−0.9114	0.2858	0.0603	0.002970
$u_{4,\mathrm{m},2}$	−1.0867	−1.0856	−1.1780	−0.9898	0.2290	0.0555	0.002746
$u_{5,\mathrm{m},2}$	−1.2188	−1.1876	−1.2520	−1.1278	0.2803	0.0390	0.001908
$u_{6,\mathrm{m},2}$	−1.3845	−1.3770	−1.4451	−1.2964	0.2371	0.0439	0.002168
$u_{7,\mathrm{m},2}$	−1.5613	−1.5232	−1.5883	−1.4720	0.2188	0.0351	0.001717
$u_{8,\mathrm{m},2}$	−1.8489	−1.8306	−1.8855	−1.7770	0.2347	0.0327	0.001589
$u_{1,\mathrm{f},2}$	−0.5675	−0.5878	−0.7389	−0.4536	0.2505	0.0877	0.004291
$u_{2,\mathrm{f},2}$	−0.6817	−0.7234	−0.8570	−0.5954	0.2631	0.0798	0.003909
$u_{3,\mathrm{f},2}$	−0.8091	−0.8383	−0.9323	−0.7539	0.2289	0.0551	0.002672
$u_{4,\mathrm{f},2}$	−1.0027	−1.0111	−1.0764	−0.9419	0.2397	0.0408	0.001964
$u_{5,\mathrm{f},2}$	−1.2437	−1.2437	−1.3018	−1.1838	0.2141	0.0357	0.001725
$u_{6,\mathrm{f},2}$	−1.5434	−1.5404	−1.6066	−1.4814	0.1797	0.0373	0.001826
$u_{7,\mathrm{f},2}$	−1.8917	−1.8879	−1.9419	−1.8144	0.2039	0.0376	0.001844
$u_{8,\mathrm{f},2}$	−2.3319	−2.3249	−2.3661	−2.2898	0.2179	0.0233	0.001116
$u_{1,\mathrm{m},3}$	−3.8678	−3.8715	−4.0227	−3.7333	0.2147	0.0888	0.004097
$u_{2,\mathrm{m},3}$	−3.7550	−3.7332	−3.8547	−3.6155	0.2248	0.0745	0.003428
$u_{3,\mathrm{m},3}$	−3.7914	−3.7814	−3.9030	−3.6703	0.2476	0.0697	0.003273
$u_{4,\mathrm{m},3}$	−3.8240	−3.8074	−3.9192	−3.6947	0.2713	0.0665	0.003145
$u_{5,\mathrm{m},3}$	−3.8762	−3.8239	−3.9087	−3.7420	0.2648	0.0514	0.002394
$u_{6,\mathrm{m},3}$	−3.9027	−3.8939	−3.9786	−3.8031	0.2714	0.0524	0.002465
$u_{7,\mathrm{m},3}$	−3.9225	−3.8876	−3.9632	−3.8168	0.2453	0.0450	0.002074
$u_{8,\mathrm{m},3}$	−3.9504	−3.9359	−4.0070	−3.8625	0.2320	0.0445	0.002054
$u_{1,\mathrm{f},3}$	−3.8332	−3.8248	−4.0130	−3.6441	0.2319	0.1111	0.005027
$u_{2,\mathrm{f},3}$	−3.5723	−3.6020	−3.7589	−3.4490	0.2536	0.0958	0.004351
$u_{3,\mathrm{f},3}$	−3.4476	−3.4735	−3.5883	−3.3621	0.2096	0.0697	0.003154
$u_{4,\mathrm{f},3}$	−3.4451	−3.4433	−3.5313	−3.3547	0.2152	0.0532	0.002370
$u_{5,\mathrm{f},3}$	−3.5066	−3.4954	−3.5760	−3.4109	0.2107	0.0496	0.002250
$u_{6,\mathrm{f},3}$	−3.6029	−3.6011	−3.6849	−3.5244	0.2339	0.0482	0.002246
$u_{7,\mathrm{f},3}$	−3.6820	−3.6867	−3.7549	−3.6078	0.2748	0.0449	0.002093
$u_{8,\mathrm{f},3}$	−3.9607	−3.9794	−4.0350	−3.9253	0.2167	0.0333	0.001539
$u_{1,\mathrm{m},4}$	−5.1455	−5.1647	−5.4041	−4.9502	0.2374	0.1394	0.006321
$u_{2,\mathrm{m},4}$	−5.3045	−5.3221	−5.5242	−5.1031	0.2037	0.1311	0.005958
$u_{3,\mathrm{m},4}$	−5.4969	−5.5057	−5.6652	−5.3128	0.2261	0.1073	0.004867
$u_{4,\mathrm{m},4}$	−5.5890	−5.5850	−5.7591	−5.3227	0.2363	0.1277	0.006056
$u_{5,\mathrm{m},4}$	−5.3995	−5.3637	−5.4994	−5.2037	0.2484	0.0905	0.004204
$u_{6,\mathrm{m},4}$	−5.0384	−5.0703	−5.2116	−4.8793	0.2613	0.0945	0.004540
$u_{7,\mathrm{m},4}$	−4.5697	−4.6006	−4.7179	−4.4272	0.2302	0.0823	0.003987
$u_{8,\mathrm{m},4}$	−4.1185	−4.1452	−4.2689	−3.9523	0.3544	0.0941	0.004617
$u_{1,\mathrm{f},4}$	−5.6623	−5.6527	−5.9731	−5.3410	0.3014	0.1934	0.008239
$u_{2,\mathrm{f},4}$	−5.9077	−5.8989	−6.2070	−5.5720	0.2272	0.1881	0.008269
$u_{3,\mathrm{f},4}$	−5.9563	−5.9763	−6.2388	−5.7106	0.2132	0.1573	0.006985
$u_{4,\mathrm{f},4}$	−5.7354	−5.7436	−5.9311	−5.5508	0.2156	0.1161	0.005157
$u_{5,\mathrm{f},4}$	−5.3013	−5.3512	−5.4926	−5.1601	0.2564	0.0980	0.004505
$u_{6,\mathrm{f},4}$	−4.7881	−4.8340	−4.9727	−4.6545	0.2354	0.0955	0.004609
$u_{7,\mathrm{f},4}$	−4.3328	−4.3555	−4.4905	−4.2183	0.1859	0.0813	0.003970
$u_{8,\mathrm{f},4}$	−3.8327	−3.8401	−3.9486	−3.6873	0.3190	0.0765	0.003764
$u_{1,\mathrm{m},5}$	−4.1793	−4.2062	−4.3557	−4.0537	0.2139	0.0930	0.004171
$u_{2,\mathrm{m},5}$	−4.4335	−4.4144	−4.5440	−4.2809	0.2237	0.0785	0.003473
$u_{3,\mathrm{m},5}$	−4.5237	−4.5192	−4.6556	−4.3843	0.2213	0.0837	0.003889

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$u_{4,\mathrm{m},5}$	−4.3262	−4.3196	−4.4223	−4.2046	0.2682	0.0664	0.003020
$u_{5,\mathrm{m},5}$	−4.2258	−4.1940	−4.2812	−4.1062	0.2371	0.0530	0.002412
$u_{6,\mathrm{m},5}$	−3.8915	−3.8894	−3.9841	−3.7965	0.1796	0.0575	0.002746
$u_{7,\mathrm{m},5}$	−3.7354	−3.6921	−3.7759	−3.6132	0.2029	0.0493	0.002330
$u_{8,\mathrm{m},5}$	−3.7863	−3.7304	−3.8061	−3.6567	0.2270	0.0457	0.002145
$u_{1,\mathrm{f},5}$	−3.9329	−3.9125	−4.1054	−3.7320	0.2527	0.1144	0.005164
$u_{2,\mathrm{f},5}$	−3.9556	−3.9833	−4.1570	−3.8115	0.2125	0.1050	0.004874
$u_{3,\mathrm{f},5}$	−3.9582	−3.9717	−4.0955	−3.8505	0.2504	0.0744	0.003292
$u_{4,\mathrm{f},5}$	−4.0062	−4.0081	−4.1100	−3.9092	0.2225	0.0610	0.002699
$u_{5,\mathrm{f},5}$	−3.9202	−3.9104	−3.9982	−3.8238	0.3031	0.0531	0.002366
$u_{6,\mathrm{f},5}$	−3.8197	−3.8220	−3.9063	−3.7392	0.2341	0.0503	0.002323
$u_{7,\mathrm{f},5}$	−3.7925	−3.7788	−3.8567	−3.6899	0.2293	0.0494	0.002329
$u_{8,\mathrm{f},5}$	−3.8113	−3.7403	−3.7957	−3.6816	0.2670	0.0343	0.001585
$u_{1,\mathrm{m},6}$	−0.9566	−0.9804	−1.0917	−0.8726	0.2515	0.0685	0.003346
$u_{2,\mathrm{m},6}$	−0.8749	−0.8487	−0.9421	−0.7527	0.2201	0.0582	0.002850
$u_{3,\mathrm{m},6}$	−0.8081	−0.7880	−0.8953	−0.6963	0.2861	0.0602	0.002969
$u_{4,\mathrm{m},6}$	−0.7093	−0.6825	−0.7788	−0.5796	0.2367	0.0578	0.002856
$u_{5,\mathrm{m},6}$	−0.6315	−0.5780	−0.6428	−0.5139	0.2564	0.0406	0.001992
$u_{6,\mathrm{m},6}$	−0.5843	−0.5568	−0.6254	−0.4814	0.2274	0.0432	0.002137
$u_{7,\mathrm{m},6}$	−0.5571	−0.4893	−0.5569	−0.4304	0.2029	0.0365	0.001797
$u_{8,\mathrm{m},6}$	−0.5478	−0.5023	−0.5554	−0.4445	0.2421	0.0332	0.001625
$u_{1,\mathrm{f},6}$	−1.5642	−1.5562	−1.7120	−1.4141	0.2287	0.0900	0.004358
$u_{2,\mathrm{f},6}$	−1.2591	−1.2692	−1.4181	−1.1379	0.2361	0.0847	0.004137
$u_{3,\mathrm{f},6}$	−1.0506	−1.0524	−1.1484	−0.9668	0.2330	0.0560	0.002698
$u_{4,\mathrm{f},6}$	−0.8093	−0.7928	−0.8591	−0.7251	0.2345	0.0404	0.001931
$u_{5,\mathrm{f},6}$	−0.6237	−0.5986	−0.6576	−0.5386	0.2300	0.0368	0.001783
$u_{6,\mathrm{f},6}$	−0.4799	−0.4551	−0.5286	−0.3939	0.2264	0.0397	0.001952
$u_{7,\mathrm{f},6}$	−0.3826	−0.3526	−0.4081	−0.2784	0.3206	0.0377	0.001852
$u_{8,\mathrm{f},6}$	−0.3575	−0.3256	−0.3639	−0.2881	0.2564	0.0236	0.001150
$u_{1,\mathrm{m},7}$	−3.3985	−3.4525	−3.6010	−3.2975	0.2229	0.0900	0.004188
$u_{2,\mathrm{m},7}$	−3.0424	−3.0442	−3.1663	−2.9314	0.2576	0.0722	0.003363
$u_{3,\mathrm{m},7}$	−2.6988	−2.7125	−2.8054	−2.6119	0.2470	0.0612	0.002925
$u_{4,\mathrm{m},7}$	−2.5249	−2.5298	−2.6366	−2.4299	0.2465	0.0605	0.002939
$u_{5,\mathrm{m},7}$	−2.3436	−2.3176	−2.3838	−2.2488	0.2088	0.0415	0.001991
$u_{6,\mathrm{m},7}$	−2.3163	−2.3120	−2.3916	−2.2318	0.2397	0.0490	0.002391
$u_{7,\mathrm{m},7}$	−2.2705	−2.2348	−2.2949	−2.1790	0.2257	0.0350	0.001672
$u_{8,\mathrm{m},7}$	−2.2026	−2.1998	−2.2732	−2.1339	0.2340	0.0415	0.002020
$u_{1,\mathrm{f},7}$	−3.0609	−3.0834	−3.2625	−2.9135	0.2316	0.1072	0.005015
$u_{2,\mathrm{f},7}$	−2.8145	−2.8617	−3.0339	−2.7134	0.2219	0.0961	0.004567
$u_{3,\mathrm{f},7}$	−2.7219	−2.7566	−2.8782	−2.6464	0.2211	0.0699	0.003324
$u_{4,\mathrm{f},7}$	−2.6422	−2.6527	−2.7357	−2.5682	0.2892	0.0522	0.002433
$u_{5,\mathrm{f},7}$	−2.6713	−2.6708	−2.7335	−2.6047	0.2379	0.0399	0.001851
$u_{6,\mathrm{f},7}$	−2.8352	−2.8309	−2.8972	−2.7546	0.2580	0.0433	0.002063
$u_{7,\mathrm{f},7}$	−2.9907	−2.9742	−3.0455	−2.8948	0.2387	0.0458	0.002210
$u_{8,\mathrm{f},7}$	−2.8523	−2.8268	−2.8744	−2.7754	0.2673	0.0305	0.001452
$u_{1,\mathrm{m},8}$	−2.9601	−2.9858	−3.1218	−2.8557	0.2609	0.0813	0.003841
$u_{2,\mathrm{m},8}$	−3.0740	−3.0558	−3.1829	−2.9382	0.2467	0.0741	0.003518
$u_{3,\mathrm{m},8}$	−3.1986	−3.1997	−3.3154	−3.0870	0.2457	0.0688	0.003272
$u_{4,\mathrm{m},8}$	−3.3983	−3.3839	−3.4828	−3.2759	0.2564	0.0622	0.002954
$u_{5,\mathrm{m},8}$	−3.6247	−3.5810	−3.6714	−3.4900	0.2233	0.0549	0.002581
$u_{6,\mathrm{m},8}$	−3.5991	−3.5724	−3.6693	−3.4720	0.2317	0.0580	0.002755
$u_{7,\mathrm{m},8}$	−3.5277	−3.4695	−3.5561	−3.3912	0.2391	0.0490	0.002270
$u_{8,\mathrm{m},8}$	−3.3220	−3.2830	−3.3560	−3.2152	0.3311	0.0422	0.001927

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$u_{1,\mathrm{f},8}$	−3.3783	−3.3883	−3.5769	−3.2257	0.2171	0.1067	0.004948
$u_{2,\mathrm{f},8}$	−3.3880	−3.4104	−3.5683	−3.2551	0.2756	0.0936	0.004328
$u_{3,\mathrm{f},8}$	−3.3496	−3.3640	−3.4852	−3.2506	0.2079	0.0714	0.003231
$u_{4,\mathrm{f},8}$	−3.4046	−3.4028	−3.4966	−3.3049	0.2557	0.0578	0.002574
$u_{5,\mathrm{f},8}$	−3.4292	−3.4140	−3.4935	−3.3266	0.2396	0.0505	0.002267
$u_{6,\mathrm{f},8}$	−3.3828	−3.3680	−3.4476	−3.2858	0.2033	0.0481	0.002239
$u_{7,\mathrm{f},8}$	−3.2865	−3.2632	−3.3411	−3.1798	0.2466	0.0479	0.002244
$u_{8,\mathrm{f},8}$	−3.2082	−3.1864	−3.2420	−3.1335	0.2155	0.0331	0.001545
$u_{1,\mathrm{m},9}$	−2.2003	−2.2020	−2.3407	−2.0780	0.2276	0.0798	0.003864
$u_{2,\mathrm{m},9}$	−2.7208	−2.6882	−2.8233	−2.5740	0.2474	0.0734	0.003515
$u_{3,\mathrm{m},9}$	−3.2461	−3.2150	−3.3452	−3.0842	0.1930	0.0779	0.003771
$u_{4,\mathrm{m},9}$	−3.6122	−3.5774	−3.6949	−3.4694	0.2467	0.0696	0.003310
$u_{5,\mathrm{m},9}$	−3.8529	−3.7818	−3.8914	−3.6732	0.2602	0.0662	0.003133
$u_{6,\mathrm{m},9}$	−3.9732	−3.9205	−4.0207	−3.8017	0.2019	0.0658	0.003142
$u_{7,\mathrm{m},9}$	−4.0407	−3.9623	−4.0623	−3.8662	0.2394	0.0589	0.002756
$u_{8,\mathrm{m},9}$	−3.8570	−3.7774	−3.8687	−3.6838	0.2730	0.0564	0.002657
$u_{1,\mathrm{f},9}$	−2.7315	−2.7057	−2.8797	−2.5456	0.2310	0.1007	0.004816
$u_{2,\mathrm{f},9}$	−3.1048	−3.1116	−3.2833	−2.9578	0.2341	0.0979	0.004642
$u_{3,\mathrm{f},9}$	−3.4553	−3.4459	−3.5793	−3.3136	0.2919	0.0806	0.003678
$u_{4,\mathrm{f},9}$	−3.7185	−3.6809	−3.7966	−3.5657	0.2302	0.0694	0.003136
$u_{5,\mathrm{f},9}$	−3.8805	−3.8411	−3.9497	−3.7291	0.2349	0.0667	0.003032
$u_{6,\mathrm{f},9}$	−4.0297	−3.9778	−4.0836	−3.8839	0.2030	0.0606	0.002800
$u_{7,\mathrm{f},9}$	−4.0687	−4.0182	−4.1148	−3.8941	0.2426	0.0649	0.003065
$u_{8,\mathrm{f},9}$	−3.9386	−3.8811	−3.9631	−3.8024	0.2981	0.0472	0.002239
$u_{1,\mathrm{m},10}$	−5.5548	−5.5932	−5.8239	−5.3630	0.3007	0.1405	0.005491
$u_{2,\mathrm{m},10}$	−5.3966	−5.3705	−5.5633	−5.1878	0.2103	0.1147	0.004823
$u_{3,\mathrm{m},10}$	−5.0330	−5.0177	−5.1657	−4.8690	0.1979	0.0884	0.003843
$u_{4,\mathrm{m},10}$	−4.8532	−4.8223	−4.9448	−4.6969	0.2476	0.0748	0.003229
$u_{5,\mathrm{m},10}$	−4.6336	−4.6026	−4.6962	−4.5068	0.2491	0.0575	0.002514
$u_{6,\mathrm{m},10}$	−4.0052	−3.9809	−4.0655	−3.8944	0.2550	0.0521	0.002393
$u_{7,\mathrm{m},10}$	−3.6663	−3.6070	−3.6783	−3.5429	0.2388	0.0415	0.001867
$u_{8,\mathrm{m},10}$	−3.3733	−3.3561	−3.4225	−3.2917	0.2878	0.0393	0.001788
$u_{1,\mathrm{f},10}$	−4.1922	−4.1902	−4.4143	−3.9875	0.3203	0.1309	0.005367
$u_{2,\mathrm{f},10}$	−4.0403	−3.9939	−4.1807	−3.8139	0.2475	0.1108	0.004739
$u_{3,\mathrm{f},10}$	−3.9698	−3.9700	−4.1085	−3.8375	0.2579	0.0820	0.003407
$u_{4,\mathrm{f},10}$	−3.9376	−3.9321	−4.0351	−3.8318	0.2295	0.0616	0.002538
$u_{5,\mathrm{f},10}$	−3.9756	−3.9593	−4.0448	−3.8727	0.1920	0.0527	0.002250
$u_{6,\mathrm{f},10}$	−3.8109	−3.8004	−3.8826	−3.7228	0.2543	0.0480	0.002145
$u_{7,\mathrm{f},10}$	−3.7326	−3.7050	−3.7769	−3.6232	0.2083	0.0458	0.002117
$u_{8,\mathrm{f},10}$	−3.6701	−3.6511	−3.7005	−3.6066	0.2497	0.0288	0.001296
$v_{1,\mathrm{m},0}$	0.0034	0.0034	na	na	na	na	na
$v_{2,\mathrm{m},0}$	0.0163	0.0163	na	na	na	na	na
$v_{3,\mathrm{m},0}$	0.0302	0.0302	na	na	na	na	na
$v_{4,\mathrm{m},0}$	0.0297	0.0297	na	na	na	na	na
$v_{5,\mathrm{m},0}$	0.0225	0.0225	na	na	na	na	na
$v_{6,\mathrm{m},0}$	0.0177	0.0177	na	na	na	na	na
$v_{7,\mathrm{m},0}$	0.0168	0.0168	na	na	na	na	na
$v_{8,\mathrm{m},0}$	0.0107	0.0107	na	na	na	na	na
$v_{1,\mathrm{f},0}$	0.0240	0.0240	na	na	na	na	na
$v_{2,\mathrm{f},0}$	0.0314	0.0314	na	na	na	na	na
$v_{3,\mathrm{f},0}$	0.0132	0.0132	na	na	na	na	na
$v_{4,\mathrm{f},0}$	0.0192	0.0192	na	na	na	na	na
$v_{5,\mathrm{f},0}$	0.0206	0.0206	na	na	na	na	na

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$v_{6,\mathrm{f},0}$	0.0193	0.0193	na	na	na	na	na
$v_{7,\mathrm{f},0}$	0.0169	0.0169	na	na	na	na	na
$v_{8,\mathrm{f},0}$	0.0175	0.0175	na	na	na	na	na
$v_{1,\mathrm{m},1}$	0.0527	0.0534	0.0425	0.0641	0.2318	0.0065	0.000298
$v_{2,\mathrm{m},1}$	0.0537	0.0534	0.0440	0.0634	0.2714	0.0058	0.000255
$v_{3,\mathrm{m},1}$	0.0417	0.0391	0.0283	0.0493	0.1969	0.0064	0.000288
$v_{4,\mathrm{m},1}$	0.0412	0.0396	0.0295	0.0494	0.2816	0.0060	0.000269
$v_{5,\mathrm{m},1}$	0.0386	0.0362	0.0284	0.0445	0.2175	0.0048	0.000215
$v_{6,\mathrm{m},1}$	0.0296	0.0261	0.0186	0.0331	0.2145	0.0045	0.000203
$v_{7,\mathrm{m},1}$	0.0352	0.0270	0.0200	0.0341	0.2468	0.0043	0.000193
$v_{8,\mathrm{m},1}$	0.0292	0.0254	0.0193	0.0311	0.2544	0.0036	0.000164
$v_{1,\mathrm{f},1}$	0.0554	0.0485	0.0328	0.0641	0.3094	0.0096	0.000417
$v_{2,\mathrm{f},1}$	0.0331	0.0365	0.0223	0.0498	0.2526	0.0083	0.000356
$v_{3,\mathrm{f},1}$	0.0263	0.0285	0.0178	0.0395	0.2707	0.0066	0.000272
$v_{4,\mathrm{f},1}$	0.0238	0.0228	0.0134	0.0317	0.2473	0.0056	0.000232
$v_{5,\mathrm{f},1}$	0.0352	0.0350	0.0263	0.0432	0.2534	0.0051	0.000221
$v_{6,\mathrm{f},1}$	0.0312	0.0297	0.0217	0.0373	0.2540	0.0048	0.000218
$v_{7,\mathrm{f},1}$	0.0366	0.0337	0.0267	0.0400	0.1844	0.0042	0.000194
$v_{8,\mathrm{f},1}$	0.0336	0.0298	0.0247	0.0350	0.2107	0.0031	0.000141
$v_{1,\mathrm{m},2}$	0.0028	0.0031	−0.0044	0.0111	0.2136	0.0047	0.000231
$v_{2,\mathrm{m},2}$	0.0093	0.0086	0.0026	0.0151	0.1798	0.0039	0.000193
$v_{3,\mathrm{m},2}$	0.0126	0.0124	0.0062	0.0188	0.2813	0.0038	0.000187
$v_{4,\mathrm{m},2}$	0.0170	0.0164	0.0106	0.0224	0.2616	0.0036	0.000177
$v_{5,\mathrm{m},2}$	0.0198	0.0172	0.0133	0.0214	0.2178	0.0025	0.000124
$v_{6,\mathrm{m},2}$	0.0200	0.0189	0.0143	0.0234	0.2811	0.0028	0.000136
$v_{7,\mathrm{m},2}$	0.0178	0.0148	0.0116	0.0189	0.3027	0.0021	0.000105
$v_{8,\mathrm{m},2}$	0.0129	0.0113	0.0080	0.0144	0.2302	0.0019	0.000094
$v_{1,\mathrm{f},2}$	0.0004	0.0005	−0.0081	0.0101	0.2313	0.0054	0.000266
$v_{2,\mathrm{f},2}$	0.0076	0.0094	0.0013	0.0175	0.2585	0.0049	0.000239
$v_{3,\mathrm{f},2}$	0.0121	0.0133	0.0079	0.0195	0.2965	0.0036	0.000173
$v_{4,\mathrm{f},2}$	0.0170	0.0168	0.0119	0.0211	0.2766	0.0028	0.000135
$v_{5,\mathrm{f},2}$	0.0204	0.0201	0.0161	0.0240	0.3025	0.0024	0.000116
$v_{6,\mathrm{f},2}$	0.0228	0.0225	0.0187	0.0269	0.2393	0.0025	0.000120
$v_{7,\mathrm{f},2}$	0.0233	0.0229	0.0183	0.0263	0.2287	0.0023	0.000112
$v_{8,\mathrm{f},2}$	0.0124	0.0118	0.0098	0.0142	0.2968	0.0013	0.000064
$v_{1,\mathrm{m},3}$	0.0261	0.0244	0.0150	0.0341	0.1823	0.0058	0.000272
$v_{2,\mathrm{m},3}$	0.0216	0.0192	0.0113	0.0273	0.2104	0.0048	0.000222
$v_{3,\mathrm{m},3}$	0.0278	0.0261	0.0188	0.0336	0.2078	0.0044	0.000207
$v_{4,\mathrm{m},3}$	0.0322	0.0307	0.0236	0.0380	0.2463	0.0043	0.000205
$v_{5,\mathrm{m},3}$	0.0366	0.0330	0.0276	0.0385	0.2218	0.0033	0.000156
$v_{6,\mathrm{m},3}$	0.0354	0.0342	0.0288	0.0397	0.2505	0.0033	0.000154
$v_{7,\mathrm{m},3}$	0.0351	0.0325	0.0280	0.0373	0.2489	0.0028	0.000130
$v_{8,\mathrm{m},3}$	0.0295	0.0282	0.0238	0.0325	0.2252	0.0026	0.000123
$v_{1,\mathrm{f},3}$	0.0134	0.0129	0.0018	0.0247	0.1969	0.0070	0.000316
$v_{2,\mathrm{f},3}$	0.0066	0.0081	−0.0014	0.0180	0.2354	0.0060	0.000270
$v_{3,\mathrm{f},3}$	0.0073	0.0083	0.0010	0.0159	0.2162	0.0045	0.000204
$v_{4,\mathrm{f},3}$	0.0156	0.0149	0.0087	0.0207	0.2399	0.0036	0.000161
$v_{5,\mathrm{f},3}$	0.0205	0.0197	0.0139	0.0252	0.2806	0.0034	0.000151
$v_{6,\mathrm{f},3}$	0.0276	0.0274	0.0223	0.0329	0.2306	0.0032	0.000149
$v_{7,\mathrm{f},3}$	0.0293	0.0297	0.0249	0.0340	0.2165	0.0028	0.000129
$v_{8,\mathrm{f},3}$	0.0326	0.0339	0.0306	0.0372	0.2385	0.0020	0.000092
$v_{1,\mathrm{m},4}$	0.0389	0.0402	0.0262	0.0554	0.1943	0.0088	0.000406
$v_{2,\mathrm{m},4}$	0.0359	0.0377	0.0241	0.0509	0.2259	0.0084	0.000384

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$v_{3,\mathrm{m},4}$	0.0465	0.0464	0.0336	0.0562	0.1947	0.0067	0.000307
$v_{4,\mathrm{m},4}$	0.0531	0.0529	0.0360	0.0646	0.2297	0.0083	0.000395
$v_{5,\mathrm{m},4}$	0.0572	0.0554	0.0443	0.0641	0.2547	0.0061	0.000287
$v_{6,\mathrm{m},4}$	0.0580	0.0609	0.0473	0.0700	0.1847	0.0061	0.000297
$v_{7,\mathrm{m},4}$	0.0525	0.0559	0.0445	0.0634	0.2319	0.0053	0.000261
$v_{8,\mathrm{m},4}$	0.0517	0.0544	0.0412	0.0620	0.1911	0.0061	0.000302
$v_{1,\mathrm{f},4}$	0.0448	0.0457	0.0265	0.0650	0.2568	0.0117	0.000507
$v_{2,\mathrm{f},4}$	0.0514	0.0534	0.0339	0.0719	0.2339	0.0115	0.000511
$v_{3,\mathrm{f},4}$	0.0584	0.0617	0.0452	0.0772	0.3193	0.0098	0.000439
$v_{4,\mathrm{f},4}$	0.0607	0.0627	0.0496	0.0747	0.2287	0.0074	0.000331
$v_{5,\mathrm{f},4}$	0.0587	0.0633	0.0506	0.0721	0.2247	0.0063	0.000293
$v_{6,\mathrm{f},4}$	0.0596	0.0641	0.0518	0.0738	0.2123	0.0063	0.000309
$v_{7,\mathrm{f},4}$	0.0612	0.0640	0.0547	0.0727	0.2808	0.0052	0.000256
$v_{8,\mathrm{f},4}$	0.0590	0.0606	0.0502	0.0681	0.2066	0.0051	0.000253
$v_{1,\mathrm{m},5}$	0.0268	0.0261	0.0161	0.0366	0.2143	0.0062	0.000282
$v_{2,\mathrm{m},5}$	0.0358	0.0341	0.0256	0.0423	0.2605	0.0051	0.000227
$v_{3,\mathrm{m},5}$	0.0401	0.0391	0.0305	0.0477	0.2094	0.0053	0.000247
$v_{4,\mathrm{m},5}$	0.0305	0.0300	0.0225	0.0367	0.2584	0.0043	0.000194
$v_{5,\mathrm{m},5}$	0.0357	0.0332	0.0275	0.0394	0.2391	0.0036	0.000163
$v_{6,\mathrm{m},5}$	0.0281	0.0274	0.0216	0.0336	0.2909	0.0036	0.000173
$v_{7,\mathrm{m},5}$	0.0258	0.0228	0.0178	0.0281	0.2198	0.0032	0.000150
$v_{8,\mathrm{m},5}$	0.0251	0.0211	0.0166	0.0257	0.1882	0.0027	0.000130
$v_{1,\mathrm{f},5}$	0.0246	0.0228	0.0116	0.0344	0.2263	0.0070	0.000316
$v_{2,\mathrm{f},5}$	0.0277	0.0291	0.0181	0.0398	0.1928	0.0065	0.000302
$v_{3,\mathrm{f},5}$	0.0264	0.0268	0.0193	0.0347	0.1999	0.0047	0.000208
$v_{4,\mathrm{f},5}$	0.0305	0.0299	0.0233	0.0365	0.2414	0.0040	0.000181
$v_{5,\mathrm{f},5}$	0.0328	0.0318	0.0263	0.0375	0.2346	0.0035	0.000156
$v_{6,\mathrm{f},5}$	0.0345	0.0345	0.0292	0.0402	0.2790	0.0033	0.000152
$v_{7,\mathrm{f},5}$	0.0366	0.0356	0.0304	0.0405	0.2338	0.0031	0.000145
$v_{8,\mathrm{f},5}$	0.0371	0.0322	0.0285	0.0356	0.2444	0.0021	0.000099
$v_{1,\mathrm{m},6}$	−0.0209	−0.0225	−0.0298	−0.0146	0.2469	0.0047	0.000229
$v_{2,\mathrm{m},6}$	−0.0236	−0.0264	−0.0328	−0.0204	0.2957	0.0038	0.000186
$v_{3,\mathrm{m},6}$	−0.0257	−0.0281	−0.0345	−0.0211	0.2271	0.0039	0.000190
$v_{4,\mathrm{m},6}$	−0.0296	−0.0321	−0.0385	−0.0257	0.2398	0.0037	0.000183
$v_{5,\mathrm{m},6}$	−0.0302	−0.0344	−0.0388	−0.0299	0.2350	0.0027	0.000133
$v_{6,\mathrm{m},6}$	−0.0268	−0.0293	−0.0338	−0.0248	0.2133	0.0028	0.000136
$v_{7,\mathrm{m},6}$	−0.0216	−0.0267	−0.0303	−0.0226	0.2042	0.0023	0.000111
$v_{8,\mathrm{m},6}$	−0.0144	−0.0180	−0.0213	−0.0147	0.2809	0.0019	0.000094
$v_{1,\mathrm{f},6}$	−0.0216	−0.0235	−0.0326	−0.0138	0.2182	0.0056	0.000272
$v_{2,\mathrm{f},6}$	−0.0347	−0.0350	−0.0436	−0.0256	0.2692	0.0053	0.000256
$v_{3,\mathrm{f},6}$	−0.0356	−0.0363	−0.0419	−0.0298	0.2393	0.0037	0.000176
$v_{4,\mathrm{f},6}$	−0.0388	−0.0409	−0.0457	−0.0363	0.3020	0.0029	0.000136
$v_{5,\mathrm{f},6}$	−0.0355	−0.0377	−0.0419	−0.0336	0.2121	0.0025	0.000120
$v_{6,\mathrm{f},6}$	−0.0312	−0.0332	−0.0372	−0.0283	0.2928	0.0026	0.000128
$v_{7,\mathrm{f},6}$	−0.0262	−0.0285	−0.0331	−0.0250	0.2111	0.0023	0.000114
$v_{8,\mathrm{f},6}$	−0.0170	−0.0194	−0.0216	−0.0173	0.2492	0.0014	0.000067
$v_{1,\mathrm{m},7}$	−0.0013	−0.0002	−0.0103	0.0099	0.2447	0.0061	0.000284
$v_{2,\mathrm{m},7}$	−0.0091	−0.0097	−0.0170	−0.0015	0.2840	0.0047	0.000216
$v_{3,\mathrm{m},7}$	−0.0111	−0.0109	−0.0177	−0.0045	0.2132	0.0040	0.000189
$v_{4,\mathrm{m},7}$	−0.0049	−0.0049	−0.0113	0.0021	0.2370	0.0039	0.000191
$v_{5,\mathrm{m},7}$	−0.0049	−0.0071	−0.0116	−0.0028	0.2896	0.0027	0.000126
$v_{6,\mathrm{m},7}$	0.0009	0.0000	−0.0047	0.0053	0.2130	0.0031	0.000149
$v_{7,\mathrm{m},7}$	0.0014	−0.0013	−0.0051	0.0027	0.2261	0.0023	0.000111

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$v_{8,\mathrm{m},7}$	0.0017	0.0013	−0.0028	0.0054	0.3368	0.0025	0.000122
$v_{1,\mathrm{f},7}$	−0.0072	−0.0059	−0.0168	0.0055	0.2738	0.0068	0.000315
$v_{2,\mathrm{f},7}$	−0.0029	−0.0004	−0.0098	0.0101	0.3121	0.0061	0.000287
$v_{3,\mathrm{f},7}$	0.0042	0.0059	−0.0016	0.0139	0.2164	0.0047	0.000222
$v_{4,\mathrm{f},7}$	0.0083	0.0085	0.0024	0.0141	0.2138	0.0037	0.000174
$v_{5,\mathrm{f},7}$	0.0151	0.0150	0.0104	0.0192	0.2540	0.0027	0.000125
$v_{6,\mathrm{f},7}$	0.0235	0.0233	0.0184	0.0278	0.2389	0.0029	0.000138
$v_{7,\mathrm{f},7}$	0.0288	0.0277	0.0228	0.0323	0.2233	0.0029	0.000140
$v_{8,\mathrm{f},7}$	0.0190	0.0175	0.0141	0.0208	0.2546	0.0020	0.000096
$v_{1,\mathrm{m},8}$	0.0108	0.0100	0.0013	0.0191	0.2760	0.0056	0.000265
$v_{2,\mathrm{m},8}$	0.0070	0.0054	−0.0024	0.0138	0.2198	0.0049	0.000235
$v_{3,\mathrm{m},8}$	0.0009	0.0000	−0.0076	0.0070	0.2442	0.0044	0.000210
$v_{4,\mathrm{m},8}$	0.0010	−0.0005	−0.0073	0.0063	0.2320	0.0041	0.000192
$v_{5,\mathrm{m},8}$	0.0071	0.0036	−0.0022	0.0096	0.2112	0.0036	0.000169
$v_{6,\mathrm{m},8}$	0.0013	−0.0010	−0.0070	0.0052	0.2371	0.0037	0.000175
$v_{7,\mathrm{m},8}$	−0.0032	−0.0076	−0.0128	−0.0023	0.2386	0.0032	0.000146
$v_{8,\mathrm{m},8}$	−0.0106	−0.0137	−0.0181	−0.0095	0.2200	0.0026	0.000119
$v_{1,\mathrm{f},8}$	0.0098	0.0101	0.0001	0.0216	0.2914	0.0066	0.000302
$v_{2,\mathrm{f},8}$	0.0035	0.0043	−0.0056	0.0137	0.2298	0.0058	0.000267
$v_{3,\mathrm{f},8}$	−0.0042	−0.0040	−0.0116	0.0040	0.2396	0.0047	0.000209
$v_{4,\mathrm{f},8}$	−0.0024	−0.0033	−0.0103	0.0031	0.2472	0.0040	0.000178
$v_{5,\mathrm{f},8}$	−0.0007	−0.0020	−0.0081	0.0034	0.2628	0.0035	0.000156
$v_{6,\mathrm{f},8}$	−0.0006	−0.0017	−0.0070	0.0040	0.1912	0.0033	0.000152
$v_{7,\mathrm{f},8}$	−0.0027	−0.0043	−0.0093	0.0008	0.2309	0.0031	0.000142
$v_{8,\mathrm{f},8}$	−0.0062	−0.0077	−0.0111	−0.0043	0.2394	0.0021	0.000096
$v_{1,\mathrm{m},9}$	0.0139	0.0112	0.0025	0.0208	0.2627	0.0055	0.000267
$v_{2,\mathrm{m},9}$	0.0120	0.0091	0.0017	0.0178	0.2395	0.0048	0.000233
$v_{3,\mathrm{m},9}$	0.0161	0.0131	0.0051	0.0214	0.2716	0.0050	0.000239
$v_{4,\mathrm{m},9}$	0.0128	0.0100	0.0030	0.0177	0.2350	0.0045	0.000213
$v_{5,\mathrm{m},9}$	0.0096	0.0044	−0.0028	0.0111	0.2476	0.0043	0.000205
$v_{6,\mathrm{m},9}$	0.0105	0.0066	−0.0009	0.0132	0.2814	0.0042	0.000201
$v_{7,\mathrm{m},9}$	0.0167	0.0114	0.0056	0.0175	0.2785	0.0037	0.000173
$v_{8,\mathrm{m},9}$	0.0174	0.0116	0.0063	0.0173	0.2454	0.0034	0.000161
$v_{1,\mathrm{f},9}$	0.0155	0.0131	0.0029	0.0239	0.1895	0.0063	0.000303
$v_{2,\mathrm{f},9}$	0.0110	0.0110	0.0011	0.0216	0.2451	0.0061	0.000290
$v_{3,\mathrm{f},9}$	0.0051	0.0039	−0.0050	0.0127	0.2274	0.0053	0.000240
$v_{4,\mathrm{f},9}$	0.0032	0.0001	−0.0075	0.0084	0.2464	0.0048	0.000216
$v_{5,\mathrm{f},9}$	0.0042	0.0015	−0.0063	0.0086	0.2768	0.0045	0.000203
$v_{6,\mathrm{f},9}$	0.0095	0.0062	−0.0002	0.0130	0.2353	0.0040	0.000185
$v_{7,\mathrm{f},9}$	0.0161	0.0128	0.0053	0.0191	0.2874	0.0041	0.000195
$v_{8,\mathrm{f},9}$	0.0167	0.0130	0.0080	0.0180	0.2506	0.0031	0.000147
$v_{1,\mathrm{m},10}$	0.0237	0.0259	0.0112	0.0404	0.2748	0.0089	0.000355
$v_{2,\mathrm{m},10}$	0.0226	0.0215	0.0096	0.0338	0.2555	0.0073	0.000307
$v_{3,\mathrm{m},10}$	0.0073	0.0075	−0.0024	0.0174	0.2363	0.0059	0.000253
$v_{4,\mathrm{m},10}$	0.0126	0.0110	0.0029	0.0191	0.2407	0.0049	0.000211
$v_{5,\mathrm{m},10}$	0.0181	0.0157	0.0093	0.0220	0.2074	0.0038	0.000169
$v_{6,\mathrm{m},10}$	0.0094	0.0073	0.0020	0.0131	0.2192	0.0034	0.000154
$v_{7,\mathrm{m},10}$	0.0051	0.0006	−0.0034	0.0053	0.2105	0.0026	0.000119
$v_{8,\mathrm{m},10}$	0.0029	0.0013	−0.0026	0.0050	0.2753	0.0023	0.000105
$v_{1,\mathrm{f},10}$	−0.0282	−0.0280	−0.0422	−0.0130	0.1962	0.0089	0.000360
$v_{2,\mathrm{f},10}$	−0.0351	−0.0371	−0.0496	−0.0243	0.2415	0.0076	0.000313
$v_{3,\mathrm{f},10}$	−0.0274	−0.0274	−0.0368	−0.0175	0.2359	0.0058	0.000236
$v_{4,\mathrm{f},10}$	−0.0144	−0.0148	−0.0222	−0.0076	0.2114	0.0045	0.000181

Table 19.6: Estimates for our annuity model based on Australian data from 1987 to 2011 using matching of moments, as well as MCMC with 30 000 steps.

parameter	match. moments	MCMC mean	5% quantile	95% quantile	accept. prob.	standard dev.	standard error
$v_{5,\mathrm{f},10}$	0.0023	0.0009	−0.0052	0.0068	0.2641	0.0037	0.000154
$v_{6,\mathrm{f},10}$	0.0064	0.0060	0.0008	0.0114	0.2224	0.0032	0.000143
$v_{7,\mathrm{f},10}$	0.0097	0.0079	0.0030	0.0124	0.2361	0.0028	0.000130
$v_{8,\mathrm{f},10}$	0.0141	0.0127	0.0101	0.0156	0.2184	0.0017	0.000076
σ_1^2	0.0373	0.0066	0.0034	0.0113	0.2513	0.0025	0.000057
σ_2^2	0.0004	0.0003	0.0001	0.0004	0.2289	0.0001	0.000002
σ_3^2	0.0055	0.0012	0.0006	0.0022	0.2327	0.0005	0.000012
σ_4^2	0.0225	0.0253	0.0144	0.0421	0.2278	0.0089	0.000221
σ_5^2	0.0057	0.0031	0.0017	0.0053	0.2107	0.0012	0.000026
σ_6^2	0.0014	0.0009	0.0005	0.0015	0.2324	0.0003	0.000010
σ_7^2	0.0051	0.0045	0.0026	0.0075	0.2700	0.0016	0.000032
σ_8^2	0.0085	0.0053	0.0030	0.0089	0.2505	0.0019	0.000041
σ_9^2	0.0109	0.0110	0.0062	0.0183	0.2308	0.0039	0.000100
σ_{10}^2	0.0029	0.0006	0.0002	0.0012	0.2429	0.0003	0.000007

19.4 Australian life tables 2013

Below, based on Example 16.17 and using MCMC with 20 000 samples, Australian male and female life tables for 2013 are. For notational purposes we leave out time parameter $T = 43$. Moreover, for a closer link to traditional notation, age categories are denoted by x for males and by y for females and gender variables are left out. For each age $x \in \{0, 1, \ldots, 100+\}$, and correspondingly for y, the table gives annual death probabilities q_x for males aged between x and $x + 1$ in 2013, as well as survivors l_x based on a starting value of 100 000 people. Furthermore, smoothed mortality parameters α_x, as well as smoothed trend parameters β_x and η_x are provided, as well as expected future life times (EFLT) with and without trends. Expected future life time with trend is given by Equation (16.15) whereas expected future life time without trend is simply calculated by using the 2013 period life table, i.e.,

$$e_x^* = \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} \left(1 - q_{x+j}\right), \quad x \in \{0, 1, \ldots, 100+\}.$$

Table 19.7: 2013 Australian male life table.

age	death prob.	surv. up to x	intercept param.	trend param.	trend reduc.	EFLT old	EFLT new	std. dev.
x	q_x	l_x	α_x	β_x	η_x	e_x^*	e_x	s_x
0	0.004063	100000	−3.1721	−0.0503	0.0254	79.41	87.95	13.53
1	0.000923	99594	−4.5478	−0.0483	0.0188	78.74	87.24	12.40
2	0.000301	99502	−5.6108	−0.0466	0.0138	77.81	86.25	12.18
3	0.000140	99472	−6.3611	−0.0450	0.0105	76.83	85.21	12.15
4	0.000093	99458	−6.7986	−0.0436	0.0089	75.84	84.14	12.17
5	0.000086	99449	−6.9235	−0.0425	0.0088	74.85	83.08	12.20
6	0.000092	99440	−6.9072	−0.0415	0.0095	73.86	82.01	12.24

Table 19.7: 2013 Australian male life table.

age	death prob.	surv. up to x	intercept param.	trend param.	trend reduc.	EFLT old	EFLT new	std. dev.
x	q_x	l_x	α_x	β_x	η_x	e_x^*	e_x	s_x
7	0.000095	99431	−6.9210	−0.0407	0.0100	72.87	80.94	12.28
8	0.000093	99422	−6.9652	−0.0401	0.0104	71.87	79.87	12.31
9	0.000088	99412	−7.0396	−0.0397	0.0105	70.88	78.80	12.35
10	0.000080	99403	−7.1442	−0.0395	0.0105	69.88	77.72	12.39
11	0.000076	99396	−7.2061	−0.0393	0.0102	68.89	76.65	12.44
12	0.000080	99388	−7.1525	−0.0389	0.0098	67.90	75.57	12.48
13	0.000097	99380	−6.9833	−0.0383	0.0092	66.90	74.49	12.53
14	0.000131	99370	−6.6985	−0.0376	0.0084	65.91	73.41	12.57
15	0.000200	99357	−6.2982	−0.0367	0.0075	64.92	72.33	12.61
16	0.000310	99338	−5.8932	−0.0356	0.0064	63.93	71.25	12.63
17	0.000438	99307	−5.5944	−0.0343	0.0055	62.95	70.19	12.63
18	0.000560	99263	−5.4018	−0.0328	0.0048	61.98	69.12	12.61
19	0.000652	99208	−5.3155	−0.0312	0.0041	61.01	68.07	12.56
20	0.000688	99143	−5.3354	−0.0294	0.0035	60.05	67.02	12.50
21	0.000696	99075	−5.4075	−0.0274	0.0031	59.09	65.97	12.44
22	0.000711	99006	−5.4778	−0.0252	0.0027	58.13	64.92	12.37
23	0.000734	98935	−5.5461	−0.0228	0.0025	57.18	63.87	12.31
24	0.000766	98863	−5.6126	−0.0203	0.0024	56.22	62.82	12.24
25	0.000807	98787	−5.6772	−0.0176	0.0024	55.26	61.78	12.17
26	0.000851	98707	−5.7342	−0.0150	0.0025	54.30	60.73	12.10
27	0.000891	98623	−5.7780	−0.0129	0.0027	53.35	59.68	12.02
28	0.000926	98535	−5.8087	−0.0113	0.0030	52.40	58.63	11.94
29	0.000956	98444	−5.8261	−0.0101	0.0035	51.45	57.58	11.86
30	0.000981	98350	−5.8303	−0.0095	0.0040	50.50	56.53	11.78
31	0.000999	98254	−5.8212	−0.0093	0.0048	49.55	55.48	11.70
32	0.001012	98155	−5.7990	−0.0096	0.0061	48.60	54.43	11.62
33	0.001021	98056	−5.7635	−0.0104	0.0079	47.64	53.37	11.55
34	0.001029	97956	−5.7149	−0.0116	0.0101	46.69	52.31	11.48
35	0.001037	97855	−5.6530	−0.0133	0.0127	45.74	51.25	11.41
36	0.001051	97754	−5.5779	−0.0155	0.0158	44.79	50.19	11.34
37	0.001074	97651	−5.4896	−0.0182	0.0194	43.84	49.13	11.28
38	0.001111	97546	−5.3881	−0.0214	0.0234	42.88	48.06	11.22
39	0.001168	97438	−5.2734	−0.0250	0.0278	41.93	46.99	11.17
40	0.001247	97324	−5.1454	−0.0291	0.0327	40.98	45.92	11.11
41	0.001352	97202	−5.0127	−0.0332	0.0374	40.03	44.85	11.05
42	0.001475	97071	−4.8838	−0.0366	0.0410	39.09	43.79	10.98
43	0.001614	96928	−4.7585	−0.0395	0.0437	38.14	42.72	10.91
44	0.001765	96771	−4.6369	−0.0418	0.0455	37.20	41.66	10.84
45	0.001927	96601	−4.5191	−0.0435	0.0462	36.27	40.59	10.76
46	0.002096	96415	−4.4050	−0.0446	0.0460	35.34	39.54	10.68
47	0.002271	96212	−4.2946	−0.0451	0.0448	34.41	38.48	10.59
48	0.002446	95994	−4.1879	−0.0450	0.0426	33.49	37.43	10.50
49	0.002617	95759	−4.0849	−0.0444	0.0394	32.58	36.37	10.40
50	0.002779	95509	−3.9856	−0.0431	0.0353	31.66	35.32	10.30
51	0.002940	95243	−3.8884	−0.0416	0.0309	30.75	34.27	10.21
52	0.003115	94963	−3.7916	−0.0402	0.0267	29.84	33.22	10.11
53	0.003309	94667	−3.6951	−0.0389	0.0229	28.93	32.18	10.01
54	0.003531	94354	−3.5990	−0.0377	0.0193	28.03	31.13	9.91
55	0.003792	94021	−3.5034	−0.0367	0.0161	27.13	30.09	9.81
56	0.004103	93664	−3.4081	−0.0357	0.0132	26.23	29.05	9.70
57	0.004474	93280	−3.3131	−0.0348	0.0106	25.34	28.02	9.59
58	0.004913	92863	−3.2186	−0.0340	0.0083	24.45	26.99	9.47

Table 19.7: 2013 Australian male life table.

age	death prob.	surv. up to x	intercept param.	trend param.	trend reduc.	EFLT old	EFLT new	std. dev.
x	q_x	l_x	α_x	β_x	η_x	e_x^*	e_x	s_x
59	0.005426	92407	−3.1244	−0.0333	0.0063	23.57	25.97	9.35
60	0.006015	91905	−3.0307	−0.0327	0.0047	22.70	24.96	9.22
61	0.006678	91352	−2.9373	−0.0323	0.0033	21.84	23.96	9.08
62	0.007411	90742	−2.8443	−0.0319	0.0023	20.99	22.97	8.94
63	0.008206	90070	−2.7516	−0.0316	0.0015	20.14	22.00	8.78
64	0.009055	89331	−2.6594	−0.0315	0.0011	19.31	21.04	8.62
65	0.009952	88522	−2.5675	−0.0314	0.0010	18.49	20.09	8.44
66	0.010934	87641	−2.4760	−0.0313	0.0010	17.67	19.16	8.27
67	0.012053	86683	−2.3849	−0.0312	0.0011	16.87	18.23	8.08
68	0.013332	85638	−2.2942	−0.0310	0.0011	16.07	17.32	7.90
69	0.014798	84496	−2.2039	−0.0306	0.0011	15.29	16.43	7.70
70	0.016481	83246	−2.1140	−0.0302	0.0011	14.52	15.56	7.50
71	0.018408	81874	−2.0249	−0.0297	0.0012	13.76	14.70	7.30
72	0.020608	80367	−1.9373	−0.0291	0.0012	13.02	13.86	7.09
73	0.023126	78710	−1.8512	−0.0285	0.0012	12.30	13.05	6.88
74	0.026012	76890	−1.7665	−0.0277	0.0012	11.59	12.25	6.66
75	0.029328	74890	−1.6832	−0.0268	0.0012	10.90	11.48	6.43
76	0.033144	72694	−1.6014	−0.0259	0.0012	10.23	10.74	6.21
77	0.037544	70284	−1.5210	−0.0249	0.0012	9.58	10.02	5.98
78	0.042629	67646	−1.4421	−0.0237	0.0012	8.95	9.33	5.74
79	0.048516	64762	−1.3646	−0.0225	0.0012	8.35	8.67	5.51
80	0.055346	61620	−1.2886	−0.0212	0.0012	7.77	8.05	5.27
81	0.062957	58210	−1.2139	−0.0200	0.0012	7.23	7.45	5.04
82	0.071038	54545	−1.1408	−0.0189	0.0012	6.72	6.90	4.80
83	0.079511	50670	−1.0691	−0.0179	0.0011	6.23	6.38	4.56
84	0.088278	46641	−0.9988	−0.0171	0.0011	5.77	5.88	4.33
85	0.097223	42524	−0.9300	−0.0165	0.0011	5.33	5.41	4.11
86	0.106559	38390	−0.8608	−0.0160	0.0022	4.90	4.97	3.89
87	0.117261	34299	−0.7896	−0.0156	0.0055	4.48	4.53	3.69
88	0.131289	30277	−0.7163	−0.0155	0.0110	4.08	4.11	3.51
89	0.150717	26302	−0.6410	−0.0154	0.0188	3.70	3.72	3.34
90	0.175783	22338	−0.5636	−0.0156	0.0288	3.35	3.37	3.20
91	0.202091	18411	−0.4905	−0.0158	0.0394	3.07	3.08	3.08
92	0.225491	14690	−0.4282	−0.0160	0.0488	2.84	2.85	2.97
93	0.245558	11378	−0.3767	−0.0161	0.0571	2.67	2.68	2.89
94	0.262068	8584	−0.3359	−0.0163	0.0644	2.54	2.54	2.82
95	0.274834	6334	−0.3058	−0.0165	0.0705	2.44	2.44	2.77
96	0.284875	4593	−0.2824	−0.0166	0.0755	2.37	2.37	2.72
97	0.293376	3285	−0.2615	−0.0168	0.0794	2.31	2.30	2.69
98	0.300386	2321	−0.2430	−0.0169	0.0822	2.27	2.26	2.66
99	0.305911	1624	−0.2271	−0.0170	0.0838	2.24	2.22	2.65
100+	0.309923	1127	−0.2136	−0.0172	0.0844	2.23	2.20	2.63

Table 19.8: 2013 Australian female life table.

age	death prob.	surv. up to y	intercept param.	trend param.	trend reduc.	EFLT old	EFLT new	std. dev.
y	q_y	l_y	α_y	β_y	η_y	e_y^*	e_y	s_y
0	0.002206	100000	−3.4305	−0.0500	0.0116	83.92	89.48	11.42
1	0.000584	99779	−4.8679	−0.0488	0.0142	83.11	88.64	10.64
2	0.000213	99721	−5.9787	−0.0476	0.0164	82.16	87.65	10.45
3	0.000107	99700	−6.7627	−0.0463	0.0184	81.17	86.63	10.41
4	0.000074	99689	−7.2199	−0.0451	0.0201	80.18	85.60	10.41
5	0.000070	99682	−7.3505	−0.0438	0.0215	79.19	84.56	10.42
6	0.000077	99675	−7.3310	−0.0425	0.0227	78.19	83.53	10.43
7	0.000081	99667	−7.3381	−0.0412	0.0235	77.20	82.49	10.44
8	0.000083	99659	−7.3717	−0.0398	0.0241	76.21	81.45	10.46
9	0.000083	99651	−7.4318	−0.0385	0.0244	75.21	80.41	10.47
10	0.000079	99642	−7.5185	−0.0371	0.0244	74.22	79.37	10.48
11	0.000078	99635	−7.5819	−0.0357	0.0242	73.22	78.33	10.50
12	0.000081	99627	−7.5722	−0.0343	0.0237	72.23	77.29	10.51
13	0.000092	99619	−7.4895	−0.0329	0.0229	71.24	76.25	10.53
14	0.000110	99610	−7.3337	−0.0314	0.0218	70.24	75.20	10.55
15	0.000143	99599	−7.1049	−0.0299	0.0204	69.25	74.16	10.56
16	0.000185	99584	−6.8718	−0.0284	0.0188	68.26	73.12	10.56
17	0.000225	99566	−6.7034	−0.0269	0.0169	67.27	72.08	10.55
18	0.000257	99543	−6.5997	−0.0254	0.0147	66.29	71.04	10.54
19	0.000275	99518	−6.5607	−0.0238	0.0122	65.30	70.01	10.52
20	0.000278	99491	−6.5864	−0.0223	0.0095	64.32	68.97	10.49
21	0.000274	99463	−6.6383	−0.0209	0.0071	63.34	67.94	10.47
22	0.000273	99436	−6.6781	−0.0198	0.0057	62.36	66.90	10.45
23	0.000272	99408	−6.7059	−0.0192	0.0053	61.38	65.86	10.43
24	0.000272	99381	−6.7215	−0.0189	0.0059	60.39	64.82	10.42
25	0.000273	99354	−6.7251	−0.0190	0.0076	59.41	63.78	10.40
26	0.000276	99327	−6.7165	−0.0194	0.0102	58.42	62.73	10.39
27	0.000282	99300	−6.6959	−0.0202	0.0138	57.44	61.69	10.38
28	0.000293	99272	−6.6632	−0.0214	0.0185	56.46	60.65	10.36
29	0.000310	99243	−6.6183	−0.0230	0.0241	55.47	59.60	10.35
30	0.000334	99212	−6.5614	−0.0250	0.0308	54.49	58.55	10.34
31	0.000364	99179	−6.4924	−0.0270	0.0375	53.51	57.51	10.33
32	0.000399	99143	−6.4113	−0.0291	0.0434	52.53	56.46	10.31
33	0.000439	99103	−6.3180	−0.0310	0.0483	51.55	55.42	10.29
34	0.000486	99060	−6.2127	−0.0329	0.0524	50.57	54.37	10.27
35	0.000541	99012	−6.0953	−0.0347	0.0556	49.60	53.32	10.24
36	0.000602	98958	−5.9736	−0.0364	0.0579	48.62	52.28	10.22
37	0.000664	98898	−5.8551	−0.0381	0.0593	47.65	51.24	10.18
38	0.000724	98833	−5.7401	−0.0397	0.0598	46.68	50.20	10.14
39	0.000783	98761	−5.6284	−0.0413	0.0595	45.72	49.16	10.10
40	0.000836	98684	−5.5201	−0.0427	0.0582	44.75	48.12	10.06
41	0.000888	98601	−5.4151	−0.0440	0.0565	43.79	47.08	10.01
42	0.000943	98514	−5.3135	−0.0449	0.0546	42.83	46.04	9.97
43	0.001003	98421	−5.2152	−0.0454	0.0526	41.87	45.00	9.92
44	0.001068	98322	−5.1203	−0.0456	0.0504	40.91	43.96	9.87
45	0.001138	98217	−5.0287	−0.0454	0.0482	39.96	42.92	9.82
46	0.001215	98106	−4.9406	−0.0449	0.0457	39.00	41.88	9.77
47	0.001300	97986	−4.8557	−0.0440	0.0432	38.05	40.84	9.71
48	0.001395	97859	−4.7743	−0.0427	0.0405	37.10	39.80	9.66
49	0.001503	97723	−4.6962	−0.0411	0.0377	36.15	38.77	9.60
50	0.001628	97576	−4.6214	−0.0392	0.0347	35.21	37.73	9.54
51	0.001765	97417	−4.5476	−0.0371	0.0316	34.26	36.70	9.48

Table 19.8: 2013 Australian female life table.

age	death prob.	surv. up to y	intercept param.	trend param.	trend reduc.	EFLT old	EFLT new	std. dev.
y	q_y	l_y	α_y	β_y	η_y	e_y^*	e_y	s_y
52	0.001908	97245	−4.4721	−0.0352	0.0284	33.32	35.67	9.41
53	0.002057	97059	−4.3951	−0.0334	0.0250	32.39	34.64	9.34
54	0.002210	96860	−4.3165	−0.0319	0.0215	31.45	33.62	9.26
55	0.002371	96646	−4.2363	−0.0304	0.0179	30.52	32.59	9.19
56	0.002557	96416	−4.1545	−0.0292	0.0145	29.60	31.57	9.11
57	0.002790	96170	−4.0712	−0.0281	0.0117	28.67	30.56	9.02
58	0.003071	95902	−3.9863	−0.0271	0.0095	27.75	29.54	8.94
59	0.003402	95607	−3.8997	−0.0263	0.0079	26.84	28.54	8.84
60	0.003780	95282	−3.8116	−0.0257	0.0069	25.93	27.54	8.74
61	0.004195	94922	−3.7220	−0.0252	0.0063	25.03	26.54	8.63
62	0.004636	94523	−3.6307	−0.0249	0.0056	24.13	25.56	8.51
63	0.005102	94085	−3.5379	−0.0247	0.0051	23.25	24.58	8.39
64	0.005589	93605	−3.4435	−0.0247	0.0045	22.36	23.61	8.25
65	0.006095	93082	−3.3475	−0.0249	0.0040	21.49	22.65	8.12
66	0.006645	92515	−3.2499	−0.0251	0.0036	20.62	21.70	7.97
67	0.007277	91900	−3.1507	−0.0253	0.0032	19.76	20.76	7.83
68	0.008004	91231	−3.0500	−0.0254	0.0029	18.90	19.82	7.68
69	0.008842	90501	−2.9476	−0.0254	0.0026	18.06	18.89	7.52
70	0.009809	89701	−2.8437	−0.0254	0.0023	17.22	17.98	7.36
71	0.010920	88821	−2.7389	−0.0253	0.0021	16.39	17.08	7.19
72	0.012192	87851	−2.6338	−0.0252	0.0020	15.57	16.19	7.02
73	0.013651	86780	−2.5284	−0.0250	0.0018	14.76	15.31	6.84
74	0.015328	85595	−2.4228	−0.0247	0.0018	13.97	14.45	6.66
75	0.017260	84283	−2.3169	−0.0244	0.0018	13.18	13.61	6.47
76	0.019489	82828	−2.2108	−0.0241	0.0018	12.42	12.78	6.27
77	0.022067	81214	−2.1044	−0.0237	0.0019	11.66	11.97	6.08
78	0.025057	79422	−1.9977	−0.0232	0.0020	10.93	11.19	5.87
79	0.028532	77432	−1.8908	−0.0227	0.0022	10.21	10.42	5.66
80	0.032581	75223	−1.7836	−0.0221	0.0024	9.51	9.68	5.45
81	0.037298	72772	−1.6773	−0.0215	0.0029	8.83	8.97	5.24
82	0.042821	70058	−1.5731	−0.0208	0.0040	8.17	8.28	5.03
83	0.049377	67058	−1.4709	−0.0200	0.0057	7.53	7.62	4.81
84	0.057282	63747	−1.3708	−0.0192	0.0080	6.92	6.99	4.60
85	0.066925	60095	−1.2727	−0.0183	0.0108	6.35	6.39	4.39
86	0.078333	56073	−1.1767	−0.0176	0.0143	5.80	5.83	4.17
87	0.091310	51681	−1.0827	−0.0169	0.0183	5.29	5.31	3.97
88	0.105841	46962	−0.9908	−0.0165	0.0229	4.83	4.84	3.76
89	0.121806	41991	−0.9009	−0.0163	0.0281	4.40	4.40	3.56
90	0.139025	36877	−0.8131	−0.0163	0.0338	4.01	4.01	3.37
91	0.157304	31750	−0.7273	−0.0165	0.0402	3.65	3.65	3.18
92	0.176480	26755	−0.6436	−0.0168	0.0471	3.33	3.32	3.01
93	0.196426	22034	−0.5620	−0.0174	0.0546	3.05	3.03	2.84
94	0.217055	17706	−0.4824	−0.0182	0.0627	2.79	2.77	2.69
95	0.238312	13863	−0.4048	−0.0191	0.0713	2.57	2.54	2.55
96	0.259794	10559	−0.3293	−0.0201	0.0794	2.37	2.33	2.42
97	0.281172	7816	−0.2559	−0.0210	0.0856	2.21	2.15	2.31
98	0.302434	5618	−0.1845	−0.0218	0.0901	2.07	1.99	2.21
99	0.323461	3919	−0.1152	−0.0224	0.0928	1.97	1.85	2.13
100	0.344038	2651	−0.0479	−0.023	0.0937	1.91	1.74	2.00

Glossary

a (index)	index for age category, page 108
AIC	Akaike information criterion, page 167
AIHW	Australian Institute of Health and Welfare, page 7
a.a.	$\mathbb{P}$-almost-all, page 35
a.s.	$\mathbb{P}$-almost-surely, page 3
$\alpha = (\alpha_{a,g})$	intercept parameter for death probabilities, page 110
$B_{\alpha,\beta}$	cumulative conditional beta distribution function with parameters $\alpha - \beta$ and $\beta + 1$, page 61
$\beta = (\beta_{a,g})$	trend parameter for death probabilities, page 110
$\beta_{\mathcal{G},\delta,X}$	adjustment for indicator function $f_{\mathcal{G},\delta,X}$, page 39
BIC	Bayesian information criterion, page 167
$\mathrm{CEPS}_\delta[X\,\vert\,\mathcal{G}]$	conditional expected proportional shortfall given $\mathcal{G}$, page 43
δ	level of risk aversion, page 21
DIC	deviance information criterion, page 167
$e_{a,g}$, e_x,e_y	expected future life time of a person with age a and gender g, or males from group x, or females from group y, respectively, with trends considered, page 158
e^*_x, e^*_y	expected future life time of males in group x or females in group y, respectively, without trends considered, page 188
$\operatorname{ess\,inf}$	essential infimum, page 20
$\operatorname{ess\,sup}$	essential supremum, page 20
$\eta = (\eta_{a,g})$	trend reduction parameter for death probabilities, page 110
$\mathrm{ES}_\delta[X]$	classical, unconditional expected shortfall, page 8
$\mathrm{ES}_\delta[X\,\vert\,\mathcal{G}]$	conditional expected shortfall given $\mathcal{G}$, page 40
$\mathrm{ES}_\delta[X, L\,\vert\,\mathcal{G}]$	contributions to conditional expected shortfall given $\mathcal{G}$ of subportfolio loss X to loss L, page 64
$\mathrm{ES}_G[X]$	G-weighted expected shortfall, page 77
$\mathrm{ES}_{B_{\alpha,\beta}}[X\,\vert\,\mathcal{G}]$	Beta-weighted conditional expected shortfall given $\mathcal{G}$, page 61
$\mathrm{ES}_G[X\,\vert\,\mathcal{G}]$	G-weighted conditional expected shortfall given $\mathcal{G}$, page 56
$\mathrm{ES}_G[X, L\,\vert\,\mathcal{G}]$	contributions to G-weighted conditional expected shortfall given $\mathcal{G}$ of subportfolio loss X to loss L, page 64
$F^{\leftarrow}$	lower quantile function of increasing function F, page 21

$F^{\rightarrow}$	upper quantile function of increasing function F, page 21
$f_{\mathcal{G},G,X}$	G-weighted adjusted indicator function, page 59
$f_{\mathcal{G},\delta,X}$	adjusted indicator function, page 39
$(\mathcal{F}_t)_{t\in[0,\infty)}$	filtration, page 3
F^{Lap}	Laplace distribution function, page 110
$\mathcal{F}^{Y}_{\mathcal{G},\delta,X}$	optimality set for conditional expected shortfall, page 41
$\mathcal{F}_{\mathcal{G},\delta}$	conditional probability densities given $\mathcal{G}$ bounded from above by $\frac{1}{1-\delta}$, page 42
$\mathcal{F}_{\mathcal{G},\delta,X}$	conditional probability densities given $\mathcal{G}$ bounded from above by $\frac{1}{1-\delta}$ with integrability condition, page 42
g (index)	index for gender, page 108
$\mathcal{G},\mathcal{H}$	sub-σ-algebras, page 3
ICD	International Statistical Classification of Diseases and Related Health Problems, page 7
i.i.d.	independent and identically distributed, page 21
K	number of non-idiosyncratic risk factor, page 99
k (index)	index for risk factor, page 99
L	portfolio loss, page 97
$L^0(\Omega,\mathcal{F},\mathbb{P},\overline{\mathbb{R}}^+)$	set of equivalence classes of all $\mathbb{P}$-a.s. equal $\mathcal{F}$-measurable random variables with values in $\mathbb{R}\cup\{\infty\}$, page 3
$L^0(\Omega,\mathcal{F},\mathbb{P})$	all elements of $L^0(\Omega,\mathcal{F},\mathbb{P},\overline{\mathbb{R}}^+)$ which are real-valued, page 3
Λ_k	k-th non-idiosyncratic risk factor or death cause, page 99
λ_k	risk factor realisation of risk factor Λ_k, page 117
LC	Lee–Carter model, page 150
$\mathcal{L}_{\mathcal{G},\mathrm{quant}}(\mathbb{P})$	domain for conditional lower quantiles as risk measure, page 30
$\mathcal{L}_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$	domain for conditional distortion risk measures with integrability condition, page 32
$\mathcal{L}^{-}_{\mathcal{G},g,\mathrm{cdrm}}(\mathbb{P})$	domain for conditional distortion risk measures with semi-integrability condition, page 32
$\mathcal{L}_{\mathcal{G},\mathrm{env}}(\mathbb{P})$	domain for upper envelopes as risk measure, page 19
$\mathcal{L}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$	domain for G-weighted conditional expected shortfall with integrability condition, page 55
$\mathcal{L}^{-}_{\mathcal{G},G,\mathrm{wces}}(\mathbb{P})$	domain for G-weighted conditional expected shortfall with semi-integrability condition, page 55
$\mathcal{L}^{-}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$	domain for G-weighted conditional expected shortfall contributions with semi-integrability condition, page 63
$\mathcal{L}_{\mathcal{G},G,L,\mathrm{contr}}(\mathbb{P})$	domain for G-weighted conditional expected shortfall contributions with integrability condition, page 63
$\ell(n\,\vert\,\theta_q,\theta_w,\sigma)$	likelihood function of parameters given death data, page 122
$\ell(n\,\vert\,\theta_q,\theta_w,\lambda,\sigma)$	likelihood function of parameters given death data, page 117
l_x, l_y	number of male survivors up to group x, or female survivors up to group y, respectively, page 188

MAP	maximum a posteriori, labelling for corresponding estimators (upper case) and estimates, page 118
MAPappr	approximations in the maximum a posteriori approach, page 121
MCMC	Markov chain Monte Carlo, page 9
MLE	maximum likelihood estimation, labelling for corresponding estimators (upper case) and estimates, page 123
MM	matching of moments, labelling for corresponding estimators (upper case) and estimates, page 113
$m_{a,g}$	population with age a and gender g, page 108
$M_{a,g}$	set of representative people with age a and gender g such that $\lvert M_{a,g}\rvert = m_{a,g}$, page 109
na	not applicable, page 179
$N_{a,g,k}$, $n_{a,g,k}$	death indicator and realisation for age a, gender g and risk factor k, page 109
$N'_{a,g,k}$, $n'_{a,g,k}$	transformed death indicator and realisation for age a, gender g and risk factor k, page 113
$N^*_{a,g,k}$, $n^*_{a,g,k}$	normalised transformed death indicator and realisation for age a, gender g and risk factor k, page 164
N_i	death indicator for policyholder i, page 95
$N_{i,k}$	death indicator for policyholder i and risk factor k, page 99
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space, page 2
$\leq_{\mathrm{st}(\mathcal{G})}$	conditional first stochastic order, page 87
$\leq_{\mathrm{cx}(\mathcal{G})}$	conditional convex order, page 87
$\leq_{\mathrm{icx}(\mathcal{G})}$	conditional increasing convex order, page 87
$\phi = (\phi_k)$	trend reduction parameter for weights, page 110
$\Phi_{\mathcal{G}}(X)$	set of $\mathcal{G}$-measurable $Z\colon \Omega \to \mathbb{R} \cup \{\infty\}$ with $X \leq Z$ a.s., page 18
Poisson(λ)	Poisson distribution with intensity λ, page 78
$\pi(\theta_q, \theta_w, \lambda, \sigma \,\vert\, n)$	posterior distribution of parameters given death data, page 117
$\pi(\theta_q, \theta_w, \sigma)$	prior distribution of corresponding parameters, page 117
$\pi(\lambda \,\vert\, \theta_q, \theta_w, \sigma)$	prior distribution of risk factor realisations given parameters, page 117
$\propto$	equal up to a constant multiplicative factor, page 117
$\psi = (\psi_k)$	trend reduction parameter for weights, page 110
$q_{a,g}$	death probability (intensity) for age a and gender g, page 108
q^*_i	death probability for policyholder i, page 96
q_i	death probability (intensity) for policyholder i, page 99
$q_\delta(X)$	lower quantile, page 21
$q_{\mathcal{G},\delta}(X)$	conditional lower quantile given $\mathcal{G}$, page 20
$q^\delta(X)$	upper quantile, page 21
$q^{\mathcal{G},\delta}(X)$	conditional upper quantile given $\mathcal{G}$, page 20
$R_{a,g,a',g',k,k'}$	sample correlation coefficient, page 165
$\rho_g[X \,\vert\, \mathcal{G}]$	conditional g-distortion risk measure given $\mathcal{G}$, page 32
$\rho^{\mathrm{ent}}_\gamma[X \,\vert\, \mathcal{G}]$	conditional entropic risk measure given $\mathcal{G}$, page 86

S	gain due to deaths in portfolio, page 97
$s_{a,g}$, s_x, s_y	standard deviation of the future life time of a person with age a and gender g, or males from group x, or females from group y, respectively, page 158
$\sigma = (\sigma_k)$	standard deviation for risk factors, page 99
$\widehat{\Sigma}^2_{a,g,k}$	estimator for risk factor standard deviations with matching of moments, page 115
T	time horizon, $T > 0$, page 108
t	time variable, page 108
θ_q	short for $(\alpha, \beta, \zeta, \eta)$, page 117
θ_w	short for (u, v, ϕ, ψ), page 117
$\mathcal{T}_{\zeta,\eta}$	trend reduction with parameters ζ and η, page 110
$u = (u_{a,g,k})$	intercept parameter for weights, page 110
$v = (v_{a,g,k})$	trend parameter for weights, page 110
$w_{a,g,k}$	weight for age a, gender g and risk factor k, page 108
$W^*_{a,g,k}(t)$, $\overline{W}^*_{a,g,k}$	estimators for weights with matching of moments, page 114
$\zeta = (\zeta_{a,g})$	trend reduction parameter for death probabilities, page 110

Bibliography

[1] B. Acciaio, H. Föllmer, and I. Penner, *Risk assessment for uncertain cash flows: Model ambiguity, discounting ambiguity and the role of bubbles*, Finance & Stochastics **16** (2012), no. 4, 669–709. 2, 4

[2] B. Acciaio and V. Goldammer, *Optimal portfolio selection via conditional convex risk measures on L^p*, Decisions in Economics and Finance. A Journal of Applied Mathematics **36** (2013), no. 1, 1–21. 2, 4, 17, 25, 39, 85

[3] B. Acciaio and I. Penner, *Dynamic risk measures*, in Di Nunno and Øksendal [38], pp. 1–34. 5, 41, 51, 52, 84, 85

[4] C. Acerbi, *Spectral measures of risk: A coherent representation of subjective risk aversion*, Journal of Banking & Finance **26** (2002), no. 7, 1505–1518. 4

[5] H. Akaike, *A new look at the statistical model identification*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control **AC-19** (1974), 716–723, System identification and time-series analysis. 167

[6] P. K. Andersen and M. Vaeth, *Simple parametric and nonparametric models for excess and relative mortality*, Biometrics (1989), 523–535. 149

[7] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, *Thinking coherently*, Risk **10** (1997), 68–71. 2

[8] ———, *Coherent measures of risk*, Mathematical Finance **9** (1999), no. 3, 203–228. 2, 52

[9] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford Studies in Probability, vol. 2, Oxford University Press, 1992. 9

[10] F. Bellini and V. Bignozzi, *On elicitable risk measures*, Quantitative Finance **15** (2015), no. 5, 725–733. 5

[11] F. Bellini, B. Klar, A. Müller, and E. Rosazza Gianin, *Generalized quantiles as risk measures*, available at SSRN: `http://ssrn.com/abstract=2225751`, 2013. 21

[12] F. Belzunce, J. F. Pinar, J. M. Ruiz, and M. A. Sordo, *Comparison of risks based on the expected proportional shortfall*, Insurance: Mathematics & Economics **51** (2012), no. 2, 292–302. 43

[13] P. Billingsley, *Probability and Measure*, second ed., Wiley Series in Probability and Statistics, John Wiley & Sons Inc., New York, 1986. 84

[14] ———, *Convergence of Probability Measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, 1999, Wiley-Interscience [John Wiley & Sons]. 31

[15] S. Bochner, *Partial ordering in the theory of martingales*, Annals of Mathematics. Second Series **62** (1955), 162–169. 18

[16] Credit Suisse First Boston, *Creditrisk$^+$: A credit risk management framework*, Tech. report, CSFB, 1997. 8, 105

[17] N. Brouhns, M. Denuit, and J. K. Vermunt, *A Poisson log-bilinear regression approach to the construction of projected lifetables*, Insurance: Mathematics and Economics **31** (2002), no. 3, 373–393. 9, 111, 149, 150, 154

[18] A. J. G. Cairns, D. Blake, and K. Dowd, *A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration*, Journal of Risk and Insurance **73** (2006), no. 4, 687–718. 149

[19] A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich, *A quantitative comparison of stochastic mortality models using data from England and Wales and the United States*, North American Actuarial Journal **13** (2009), no. 1, 1–35. 112, 149

[20] M. Carter and B. van Brunt, *The Lebesgue–Stieltjes Integral*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2000, A practical introduction. 36, 38, 60

[21] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio, *Risk measures: Rationality and diversification*, Mathematical Finance **21** (2011), no. 4, 743–774. 2

[22] A. Chaudhry and S. M. Zubair, *On a Class of Incomplete Gamma Functions with Applications*, Taylor & Francis, 2001. 119, 120

[23] P. Cheridito, F. Delbaen, and M. Kupper, *Dynamic monetary risk measures for bounded discrete-time processes*, Electronic Journal of Probability **11** (2006), no. 3, 57–106. 2, 85

[24] P. Cheridito and M. Stadje, *Time-inconsistency of VaR and time-consistent alternatives*, Finance Research Letters **6** (2009), 40–46. 4, 17, 19, 25, 51

[25] A. S. Cherny and D. B. Madan, *Coherent measurement of factor risks*, available at SSRN: `http://ssrn.com/abstract=904543` or `http://dx.doi.org/10.2139/ssrn.904543`, 2006. 5, 55, 61, 62

[26] K. C. Cheung and S. Vanduffel, *Bounds for sums of random variables when the marginal distributions and the variance of the sum are given*, Scandinavian Actuarial Journal **2013** (2013), no. 2, 103–118. 178

[27] G. Choquet, *Theory of capacities*, Université de Grenoble. Annales de l'Institut Fourier **5** (1953), 131–295. 32

[28] DAV-Unterarbeitsgruppe Todesfallrisiko, *Herleitung der Sterbetafel DAV 2008 T für Lebensversicherungen mit Todesfallcharakter*, Blätter der DGVFM **30** (2009), no. 1, 189–224 (German). 134, 135

[29] F. Delbaen, *Coherent Risk Measures*, Cattedra Galileiana. [Galileo Chair], Scuola Normale Superiore, Classe di Scienze, Pisa, 2000. 2

[30] ———, *Coherent risk measures on general probability spaces*, Advances in finance and stochastics (K. Sandmann and P. J. Schönbucher, eds.), Springer-Verlag, Berlin, 2002, Essays in honour of Dieter Sondermann, pp. 1–37. 52

[31] ———, *The structure of m-stable sets and in particular of the set of risk neutral measures*, in Émery and Yor [44], pp. 215–258. 4, 19, 25, 51, 75, 77, 83, 84, 85

[32] M. A. H. Dempster (ed.), *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, 2002, Papers from the workshop held in Cambridge, October 2–3, 1998. 199

[33] D. Denneberg, *Non-Additive Measure and Integral*, Theory and Decision Library. Series B: Mathematical and Statistical Methods, vol. 27, Kluwer Academic Publishers Group, Dordrecht, 1994. 32, 37

[34] D. Dentcheva, S. Penev, and A. Ruszczyński, *Kusuoka representation of higher order dual risk measures*, Annals of Operations Research **181** (2010), 325–335. 5

[35] K. Detlefsen and G. Scandolo, *Conditional and dynamic convex risk measures*, Finance and Stochastics **9** (2005), no. 4, 539–561. 2, 51, 52, 85

[36] J. Dhaene, S. Vanduffel, M. J. Goovaerts, R. Kaas, Q. Tang, and D. Vyncke, *Risk measures and comonotonicity: A review*, Stochastic Models **22** (2006), no. 4, 573–606. 32

[37] J. M. L. Dhaene, A. Kukush, D. Linders, and Q. Tang, *Some remarks on quantiles and distortion risk measures*, European Actuarial Journal **2** (2012), no. 2, 319–328. 4, 31, 32

[38] G. Di Nunno and B. Øksendal (eds.), *Advanced Mathematical Methods for Finance*, Springer-Verlag, Heidelberg, 2011. 197

[39] P. Embrechts and M. Hofert, *A note on generalized inverses*, Mathematical Methods of Operations Research **77** (2013), no. 3, 423–432. 21, 22, 49, 61

[40] P. Embrechts, A. J. McNeil, and D. Straumann, *Correlation and dependence in risk management: Properties and pitfalls*, in Dempster [32], Papers from the workshop held in Cambridge, October 2–3, 1998, pp. 176–223. 2, 30

[41] P. Embrechts and G. Puccetti, *Bounds for functions of dependent risks*, Finance and Stochastics **10** (2006), no. 3, 341–352. 98

[42] P. Embrechts, G. Puccetti, and L. Rüschendorf, *Model uncertainty and VaR aggregation*, Journal of Banking and Finance **37** (2013), no. 8, 2750–2764. 98

[43] P. Embrechts, G. Puccetti, L. Rüschendorf, R. Wang, and A. Beleraj, *An academic response to Basel 3.5*, Risks **2** (2014), no. 1, 25–48. 2, 43

[44] M. Émery and M. Yor (eds.), *In Memoriam Paul-André Meyer: Séminaire de Probabilités XXXIX*, Lecture Notes in Mathematics, vol. 1874, Springer-Verlag, Berlin, 2006. 199

[45] N. Falkner and G. Teschl, *On the substitution rule for Lebesgue–Stieltjes integrals*, Expositiones Mathematicae **30** (2012), no. 4, 412–418. 59, 60, 62

[46] V. Fasen and A. Svejda, *Time consistency of multi-period distortion measures*, Statistics & Risk Modeling with Applications in Finance and Insurance **29** (2012), no. 2, 133–153. 32

[47] A. Feuerverger, *A consistent test for bivariate dependence*, International Statistical Review/Revue Internationale de Statistique (1993), 419–433. 165, 166

[48] H. Föllmer, *Spatial risk measures and their local specification: The locally law-invariant case*, Statistics & Risk Modeling **31** (2014), no. 1, 79–101. 2

[49] H. Föllmer and A. Schied, *Stochastic Finance*, third ed., de Gruyter Studies in Mathematics, vol. 27, Walter de Gruyter & Co., Berlin, 2004, An introduction in discrete time. 4, 18, 20, 22, 85

[50] ———, *Convex and coherent risk measures*, working paper, 2008. 17, 85

[51] D. Gamerman and H. F. Lopes, *Markov chain Monte Carlo*, second ed., Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2006, Stochastic simulation for Bayesian inference. 124

[52] H. U. Gerber, *Life Insurance Mathematics*, third ed., Springer-Verlag; Association of Swiss Actuaries, Zürich, 1997, With exercises contributed by Samuel H. Cox, With a foreword by Hans Bühlmann, Translated from the German by Walther Neuhaus. 7

[53] S. Gerhold, U. Schmock, and R. Warnung, *A generalization of Panjer's recursion and numerically stable risk aggregation*, Finance and Stochastics **14** (2010), no. 1, 81–128. 8, 102

[54] G. Giese, *Enhancing CreditRisk*$^+$, Risk **16** (2003), no. 4, 73–77. 8, 102, 105

[55] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Interdisciplinary Statistics, Taylor & Francis, 1995. 124

[56] L. G. Godfrey, *Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables*, Econometrica **46** (1978), no. 6, 1293–1301. 165

[57] V. Goldammer and U. Schmock, *Generalization of the Dybvig–Ingersoll–Ross theorem and asymptotic minimality*, Mathematical Finance **22** (2012), no. 1, 185–213. 17

[58] B. Gompertz, *On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies*, Philosophical transactions of the Royal Society of London (1825), 513–583. 149

[59] W. Grzenda and W. Zięba, *Conditional central limit theorem*, International Mathematical Forum **3** (2008), no. 29–32, 1521–1528. 164

[60] J. B. Guerard, Jr. (ed.), *Handbook of portfolio construction*, Springer-Verlag, New York, 2010, Contemporary Applications of Markowitz Techniques. 204

[61] M. Gundlach and F. Lehrbass (eds.), *CreditRisk$^+$ in the Banking Industry*, Springer Finance, Springer-Verlag, Berlin, 2004. 201

[62] H. Gzyl and S. Mayoral, *On a relationship between distorted and spectral risk measures*, MPRA Paper 916, University Library of Munich, Germany, 2006. 4

[63] H. Haaf, O. Reiß, and J. Schoenmakers, *Numerically stable computation of CreditRisk$^+$*, in Gundlach and Lehrbass [61], pp. 69–77. 8, 102, 105

[64] D. A. Harville, *Maximum likelihood approaches to variance component estimation and to related problems*, Journal of the American Statistical Association **72** (1977), no. 358, 320–338. 124

[65] S. W. He, J. G. Wang, and J. A. Yan, *Semimartingale Theory and Stochastic Calculus*, Kexue Chubanshe (Science Press), Beijing, 1992. 4, 23, 24, 28, 29, 44, 45, 46, 49, 53, 60, 69, 70, 83, 84, 88, 90, 91

[66] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control, John Wiley & Sons, 2004. 163

[67] K. Jakob, M. Fischer, and S. Kolb, *crp.csfp: CreditRisk$^+$ portfolio model*, 2013, R package version 2.0. 105

[68] E. Jouini and C. Napp, *Conditional comonotonicity*, Decisions in Economics and Finance. A Journal of Applied Mathematics **27** (2004), no. 2, 153–166. 19, 61, 86

[69] R. Kainhofer, M. Predota, and U. Schmock, *The new Austrian annuity valuation table AVÖ 2005R*, Mitteilungen der Aktuarvereinigung Österreichs **13** (2006), 55–135. 9, 110, 111, 133, 136, 150, 154, 158, 159, 160

[70] M. Kalkbrener, *An axiomatic approach to capital allocation*, Mathematical Finance **15** (2005), no. 3, 425–437. 5, 63, 68

[71] O. Kallenberg, *Foundations of Modern Probability*, second ed., Probability and its Applications (New York), Springer-Verlag, 2002. 19, 36, 58, 70, 84, 124

[72] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. 19, 23, 31, 84

[73] A. Klenke, *Probability Theory*, Universitext, Springer-Verlag, London, 2008, A comprehensive course, Translated from the 2006 German original. 53

[74] R. Koenker, *Quantile Regression*, Econometric Society Monographs, vol. 38, Cambridge University Press, Cambridge, 2005. 4, 19

[75] E. J. Kontoghiorghes (ed.), *Handbook of Parallel Computing and Statistics*, Statistics: Textbooks and Monographs, vol. 184, Chapman & Hall/CRC, Boca Raton, FL, 2006. 205

[76] P. A. Krokhmal, *Higher moment coherent risk measures*, Quantitative Finance **7** (2007), no. 4, 373–387. 5

[77] M. Kupper and W. Schachermayer, *Representation results for law invariant time consistent functions*, Mathematics and Financial Economics **2** (2009), no. 3, 189–210. 13, 85

[78] R. D. Lee and L. R. Carter, *Modeling and forecasting U.S. mortality*, Journal of the American Statistical Association **87** (1992), no. 419, 659–671. 9, 111, 149, 150, 154

[79] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, third ed., Springer Texts in Statistics, Springer-Verlag, 2005. 115, 124, 165, 166

[80] S. Leorato, F. Peracchi, and A. V. Tanase, *Asymptotically efficient estimation of the conditional expected shortfall*, Computational Statistics & Data Analysis **56** (2012), no. 4, 768–784. 2

[81] D. B. Madan and A. Cherny, *Markets as a counterparty: An introduction to conic finance*, International Journal of Theoretical and Applied Finance **13** (2010), no. 8, 1149–1177. 4

[82] P. Magnus and K. Sadkowsky, *Mortality over the twentieth century in Australia: Trends and patterns in major causes of death*, no. 4, Australian Institute of Health and Welfare (AIHW): Canberra, 2006. 139

[83] W. M. Makeham, *On the law of mortality and the construction of annuity tables*, The Assurance Magazine, and Journal of the Institute of Actuaries (1860), 301–310. 149

[84] A. J. McNeil and R. Frey, *Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach*, Journal of Empirical Finance **7** (2000), 271–300. 2, 4, 5, 16, 17, 19, 75, 76

[85] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management*, Princeton Series in Finance, Princeton University Press, Princeton, NJ, 2005, Concepts, techniques and tools. 1, 2, 8, 21, 28, 43, 77, 86, 104, 107, 177

[86] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, second ed., Stochastic Modelling and Applied Probability, vol. 36, Springer-Verlag, Berlin, 2005. 84

[87] P. T. Ng and M. Maechler, *cobs: COBS – constrained B-splines (sparse matrix based)*, 2011, R package version 1.2-2. 159

[88] C. P. Niculescu and L.-E. Persson, *Convex Functions and their Applications*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 23, Springer-Verlag, New York, 2006, A contemporary approach. 36

[89] Australian Institute of Health and Welfare (AIHW), *Multiple causes of death in Australia: An analysis of all natural and selected chronic disease causes of death 1997–2007*, AIHW bulletin **13** (2012), no. 105. 9

[90] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967. 19

[91] U. Pasdika and J. Wolff, *Coping with longvity: The new German annuity valuation table DAV 2004 R*, 2005, Presented at the living to 100 and beyond symposium, sponsored by the Ssociety of Actuaries. 7, 111, 133, 134, 135

[92] F. Peracchi and A. V. Tanase, *On estimating the conditional expected shortfall*, Applied Stochastic Models in Business and Industry **24** (2008), no. 5, 471–493. 2

[93] W. Perks, *On some experiments in the graduation of mortality statistics*, Journal of the Institute of Actuaries (1932), 12–57. 149

[94] G. C. Pflug, *Some remarks on the value-at-risk and the conditional value-at-risk*, in Uryasev [123], Methodology and applications, pp. 272–281. 30

[95] P. E. Protter, *Stochastic integration and differential equations*, second ed., Applications of Mathematics, vol. 21, Springer-Verlag, Berlin, 2004, Stochastic Modelling and Applied Probability. 32

[96] G. Puccetti, *Sharp bounds on the expected shortfall for a sum of dependent random variables*, Statistics & Probability Letters **83** (2013), no. 4, 1227–1232. 98

[97] G. Puccetti and L. Rüschendorf, *Sharp bounds for sums of dependent risks*, Journal of Applied Probability **50** (2013), no. 1, 42–53. 98

[98] F. Qi, Cui R.-Q., Chen C.-P., and Guo B.-N., *Some completely monotonic functions involving polygamma functions and an application*, Journal of mathematical analysis and applications **310** (2005), no. 1, 303–308. 119, 120

[99] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. 119, 123, 128, 166, 177

[100] A. Resti, *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, Risk Books, 2008. 204

[101] J. H. Rob, H. with contributions from Booth, L. Tickle, and J. Maindonald, *demography: Forecasting mortality, fertility, migration and population data*, 2014, R package version 1.17. 151, 154

[102] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed., Springer-Verlag, 2007. 167

[103] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, second ed., Springer Texts in Statistics, Springer-Verlag, New York, 2004. 125

[104] G. O. Roberts, A. Gelman, and W. R. Gilks, *Weak convergence and optimal scaling of random walk Metropolis algorithms*, The Annals of Applied Probability **7** (1997), no. 1, 110–120. 127, 128

[105] B. Roorda and J. M. Schumacher, *Time consistency conditions for acceptability measures, with an application to tail value at risk*, Insurance: Mathematics & Economics **40** (2007), no. 2, 209–230. 51

[106] E. Rosazza Gianin, *Risk measures via g-expectations*, Insurance: Mathematics & Economics **39** (2006), no. 1, 19–34. 85

[107] J. S. Rosenthal, *Parallel computing and Monte Carlo algorithms*, Far East Journal of Theoretical Statistics **4** (2000), no. 2, 207–236. 127

[108] C. Rudolph, *A generalization of Panjer's recursion for dependent claim numbers and an approximation of Poisson mixture models*, Ph.D. thesis, Vienna University of Technology, 2014. 102

[109] L. Rüschendorf, *On conditional stochastic ordering of distributions*, Advances in Applied Probability **23** (1991), no. 1, 46–63. 87

[110] U. Schmock, *Estimating the value of the WinCAT coupons of the Winterthur insurance convertible bond: A study of the model risk*, ASTIN Bulletin **29** (1999), no. 1, 101–163. 134

[111] ———, *Modelling Dependent Credit Risks with Extensions of Credit Risk+ and Application to Operational Risk*, `http://www.fam.tuwien.ac.at/~schmock/notes/ExtensionsCreditRiskPlus.pdf`, Lecture Notes, Version April 29, 2015. 5, 7, 8, 10, 14, 21, 38, 39, 40, 43, 47, 57, 63, 64, 68, 75, 78, 95, 97, 98, 99, 100, 102, 103, 104, 115, 157, 171, 173, 174, 177, 178, 179

[112] E. N. Sereda, E. M. Bronshtein, S. T. Rachev, F. J. Fabozzi, W. Sun, and S. V. Stoyanov, *Distortion risk measures in portfolio optimization*, in Guerard [60], Contemporary Applications of Markowitz Techniques, pp. 649–673 (English). 4, 32

[113] M. Shaked and J. G. Shanthikumar, *Stochastic Orders*, Springer Series in Statistics, Springer-Verlag, New York, 2007. 51, 87

[114] P. V. Shevchenko, *Modelling Operational Risk using Bayesian Inference*, Springer-Verlag, 2011. 106, 116, 117, 118, 119, 124, 125, 126, 127, 129, 141, 177, 180

[115] M. A. Sordo and H. M. Ramos, *Characterization of stochastic orders by L-functionals*, Statistical Papers **48** (2007), no. 2, 249–263. 36

[116] J. M. Stoyanov, *Counterexamples in Probability*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Chichester, 1987. 16, 19

[117] A. Stuart, *Gamma-distributed products of independent random variables*, Biometrika **49** (1962), 564–565. 135

[118] B. Sundt, *On multivariate Panjer recursions*, ASTIN Bulletin **29** (1999), no. 1, 29–45. 102

[119] D. Tasche, *Conditional expectation as quantile derivative*, working paper, Technische Universität München, 2000. 63, 69

[120] ———, *Capital allocation to business units and sub-portfolios: the euler principle*, pp. 423–453, in Resti [100], 2008. 68

[121] L. Tierney, *Markov chains for exploring posterior distributions*, The Annals of Statistics **22** (1994), no. 4, 1701–1762, With discussion and a rejoinder by the author. 125

[122] S. Urbanek and Y. Rubner, *emdist: Earth mover's distance*, 2012, R package version 0.3-1. 177

[123] S. P. Uryasev (ed.), *Probabilistic Constrained Optimization*, Nonconvex Optimization and its Applications, vol. 49, Kluwer Academic Publishers, Dordrecht, 2000, Methodology and applications. 203

[124] P. Vellaisamy and B. Chaudhuri, *Poisson and compound Poisson approximations for random sums of random variables*, Journal of Applied Probability **33** (1996), no. 1, 127–137. 9, 96, 104

[125] G. Vitali, *Sulla definizione di integrale delle funzioni di una variabile*, Annali di Matematica Pura ed Applicata **2** (1925), no. 1, 111–121. 32

[126] S. S. Wang, V. R. Young, and H. H. Panjer, *Axiomatic characterization of insurance prices*, Insurance: Mathematics & Economics **21** (1997), no. 2, 173–183. 4

[127] W. H. Wetterstrand, *Parametric models for life insurance mortality data: Gompertz's law over time*, Transactions of the Society of Actuaries **33** (1981), 159–175. 149

[128] D. J. Wilkinson, *Parallel Bayesian computation*, in Kontoghiorghes [75], pp. 477–508. 127

[129] D. Williams, *Probability with Martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. 84

[130] J. L. Wirch and Hardy M. R., *Distortion risk measures. Coherence and stochastic dominance*, `http://pascal.iseg.utl.pt/~cemapre/ime2002/main_page/papers/JuliaWirch.pdf`, 2006. 37

[131] G. L. Wise, *Counterexamples in Probability and Real Analysis*, Oxford University Press, 1993. 22

[132] M. E. Yaari, *The dual theory of choice under risk*, Econometrica **55** (1987), no. 1, 95–115. 4

[133] A. Zeileis and T. Hothorn, *Diagnostic checking in regression relationships*, R News **2** (2002), no. 3, 7–10. 165

Jonas Hirz

Curriculum Vitae

Education

2011–present **Ph.D. study in Financial and Actuarial Mathematics**, *TU Wien,* Austria*, and CSIRO North Ryde,* Australia.
Research topics: credit risk modelling, modelling of annuity portfolios, conditional risk measures, optimisation of structured products and correlation inequalities.
Along with that: completion of all courses to become fully qualified actuary.

2009–2011 **Master of Science in Mathematics**, *University of Salzburg*, Austria.
Finished within minimum duration of four semesters.

2006–2009 **Bachelor of Science in Mathematics**, *University of Salzburg*, Austria.
Finished within minimum duration of six semesters.

2002–2006 **Secondary School**, *BORG Radstadt*, Austria.

Masters Thesis

title *Design of Optimal Cost-Efficient Payoffs and Corresponding Investment Strategies.*

supervisor Univ.-Prof. Dr. Uwe Schmock, *TU Wien.*

Academic Career

2011–present **Research Assistant**, *TU Wien*, Research Unit of Financial and Actuarial Mathematics (FAM), Austria.

2014 **Researcher**, *CSIRO North Ryde, Research Unit of Computational Informatics*, Australia.
Research with Prof. Dr. Pavel V. Shevchenko via the 2014 Endeavour Research Fellowship

2012–2014 **Two-Year Portfolio Management Program**, *Institute for Strategic Research in Capital Markets (ISK) Vienna*, Austria.
www.iskwien.at

Awards and Fellowships

2014 **Endeavour Research Fellowship**, *CSIRO North Ryde*, Australia.
Scholarship by the Australian Government, Department of Education and Training.

2012 **Hans Stegbuchner Award**, *University of Salzburg*, Austria.

2012 **AVÖ-Award for Theses**, *Actuarial Association of Austria.*

2011 **Award of the Ministry of Science and Research**, *Vienna*, Austria.
Award for best graduates in Austria 2011.

Wiedner Hauptstraße 8–10, 1040 Vienna, Austria
✉ *hirz@fam.tuwien.ac.at* • *fam.tuwien.ac.at*
Vienna, May 19, 2015

2011 **PRisMa Lab Scholarship for Theses**, *TU Wien*, Research Unit of Financial and Actuarial Mathematics (FAM), Austria.

2009/2011 **Performance Scholarship for BSc and MSc**, *University of Salzburg*, Austria.

Conferences and Seminars

2015 **Lecture Series in Financial and Actuarial Mathematics**, *TU Wien*, Austria.
Invited Talk: *Ein Modell zur Risikoaggregation in Pensions- und Lebensversicherungsportfolios*

2014 **2nd Conference on Stochastics of Environmental and Financial Economics**, *Oslo*, Norway.
Talk: *Modelling Annuity Portfolios and Longevity Risk with Extended CreditRisk*$^+$.

2014 **2nd European Actuarial Journal Conference**, *TU Wien*, Austria.
Talk: *Modelling Annuity Portfolios and Longevity Risk with Extended CreditRisk*$^+$.

2014 **4th IMS-FPS Conference**, *University of Technology, Sydney*, Australia.
Invited talk: *Modelling Annuity Portfolios and Longevity Risk with Extended CreditRisk*$^+$.

2013 **FAM Research Seminar about High Frequency Trading**, *TU Wien*, Austria.
Talk: *Introduction to Cointegration with Applications to Finance.*

2013 **One-Day Workshop on Portfolio Risk Management (PRisMa Day 2013)**, *TU Wien*, Austria.
Invited talk: *Risk Measures: From the Unconditional to the Conditional Case.*

2013 **Stochastic Analysis and Applications Conference**, *University of Oxford*, England.

2013 **6th AMaMeF and Banach Center Conference**, *Warsaw*, Poland.
Talk: *Conditional Quantiles, Conditional Weighted Expected Shortfall and Application to Risk Capital Allocation.*

2012 **Young Researcher Workshop**, *Berlin*, Germany.

2012 **Quantitative Methods in Finance Conference**, *Cairns*, Australia.

2012 **Bachelier Finance Society 7th World Congress**, *Sydney*, Australia.
Talk: *Design of Optimal Cost-Efficient Payoffs and Corresponding Investment Strategies.*

2011 **One-Day Workshop on Portfolio Risk Management 2011**, *TU Wien*, Austria.
Talk: *Design of Optimal Cost-Efficient Payoffs and Corresponding Investment Contracts.*

2010 **DAA-Workshop for Young Mathematicians**, *Ulm*, Germany.

Teaching Experience

2013/2014 **Exercise Pension Insurance Mathematics**, *TU Wien*, Austria.

2012/2013 **Exercise Life Insurance Mathematics**, *TU Wien*, Austria.

2012/2013 **Assistant for the lecture Financial Mathematics II**, *TU Wien*, Austria.

Wiedner Hauptstraße 8–10, 1040 Vienna, Austria
✉ *hirz@fam.tuwien.ac.at* • *fam.tuwien.ac.at*
Vienna, May 19, 2015