Sensitivity Analysis of Simulation Models using Divergence Minimisation

Vaishno Devi Makam

Supervisors: Pietro Millossovich
Andreas Tsanakas

Faculty of Actuarial Science and Insurance
Bayes Business School, City, University of London

This dissertation is submitted for the degree of
Doctor of Philosophy

September 2023

I would like to dedicate this thesis to my supervisors, Pietro Millossovich and Andreas Tsanakas. I am deeply grateful for the time and effort that my supervisors devoted to guiding me through this academic journey.

Acknowledgements

Throughout my PhD journey, I have received invaluable support and guidance from numerous individuals, and I would like to express my sincere gratitude and appreciation to them. Their constant encouragement and assistance have been crucial in completing this research project, and without them, this achievement would not have been possible.

First and foremost, I want to express my heartfelt appreciation and gratitude to my supervisors, Dr. Pietro Millossovich and Prof. Andreas Tsanakas, for their unwavering guidance and support throughout the course of this research project. Even during times when it was difficult and challenging, they patiently supported me and helped me navigate through complex mathematical concepts and proofs required for this study. I am thankful for their invaluable feedback, patience, and encouragement.

In addition to my supervisors, I would like to thank Ronald Richman for his valuable contributions and profound insights on Solvency II. His extensive knowledge has significantly enriched my understanding of this research area. His expertise and feedback have been crucial in the successful implementation of the Solvency II standard model in Chapter 3.

I would also like to thank everyone at the PhD programme and Faculty of Actuarial Science and Insurance at Bayes Business School, who have assisted and supported me in the completion of my research.

Last but not least, I want to express my deepest gratitude to my friends and family. Their support, encouragement, and understanding have been a constant source of inspiration and motivation throughout my academic journey. I am indebted to my family and friends for their unwavering support throughout my academic journey. They have stood by me during challenging times, and given me the strength to overcome obstacles.

Declaration

I declare that this thesis and the work presented in it are my own. I hereby confirm that

- This work was done wholly while in candidature for a research degree at Bayes Business School.
- This thesis contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted the work of others, the source is always given. With the exception of such quotation, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- This work was done in collaboration with my two supervisors, Pietro Millossovich and Andreas Tsanakas and I have provided proper acknowledgments for individuals consulted throughout the process.
- I grant powers of discretion to the University Librarian of Bayes Business School, City, University of London, to allow the thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgements.

Vaishno Devi Makam
September 2023

CITY UNIVERSITY LONDON

STATEMENT OF CO-AUTHORS of JOINT PUBLICATIONS

TO WHOM IT MAY CONCERN

Title of publication: Sensitivity Analysis using χ^2-divergence

Name of candidate: Vaishno Devi Makam

Title of research thesis: Sensitivity Analysis of Simulation Models using Divergence Minimisation

Name of first supervisor: Pietro Millossovich

We, the undersigned, co-authors of the above publication, confirm that the above publication has not been submitted as evidence for which a degree or other qualification has already been awarded.

We, the undersigned, further indicate the candidate's contribution to the publication in our joint statement below.

Signature: ANDREAS TSANAKAS

Name: Andreas Tsanakas
Date: 31/03/2023

Signature: PIETRO MILLOSSOVICH

Name: Pietro Millossovich
Date: 31/03/2023

Signature:

Name:
Date:

Signature:

Name:
Date:

Statement indicating the candidate's contribution to the publication
{Statement in support of candidate's contribution to the publication}
As co-authors of this work, we can confirm that Vaishno Devi Makam had a major contribution to the published paper. The authors' contributions are commensurate to her role as PhD student and our roles as supervisors.

Abstract

Sensitivity analysis is a critical tool used by risk managers for developing and interpreting quantitative models, as it provides invaluable information regarding model behaviour and the importance of model assumptions. Building an effective sensitivity analysis approach leads to more robust decisions ultimately impacting the financial stability of a company. In this thesis, I develop a novel sensitivity approach that expands the current literature on sensitivity analysis and stress testing. A change of measure is derived by minimising the χ^2-divergence subject to a moment constraint on a chosen risk factor. A reverse-forward sensitivity analysis framework is then specified such that the behaviour of model inputs and output can be assessed consistently. Metrics specific to the approach I adopt are also introduced to effectively capture the sensitivities of risk factors in a model to rank them according to their order of importance in a model. A model's sensitivities to different risk factors are contingent on parametric assumptions. It is thus important to consider what the impact of (statistical) uncertainty with respect to such assumptions is, on the measured sensitivities. The reverse and forward sensitivity analysis, along with specific bootstrapping procedures, are employed to assess model and parameter uncertainty in a Solvency II type model for non-life premium and reserve risk. I show that a change in the distributional assumptions can have an impact on the evaluation of sensitivities of risk factors and consequently result in reordering of risk factors' importance. The thesis then progresses towards developing a more flexible approach for conducting sensitivity analysis, giving the modeller a choice to select from a broader class of f-divergences which include the well-known χ^2- and Kullback-Leibler divergences. We characterise the properties of the solutions to the divergence minimisation problems we consider, in terms of the properties of the divergence function used. In particular, I focus on the shape of the solution and on stochastic ranking of a variable under different solutions of the divergence minimisation problems.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Sensitivity analysis in insurance risk management

Insurance and financial institutions use sophisticated quantitative simulation models to effectively manage risk that they face. Specifically, for an insurance company, numerous factors, such as interest and inflation rates, expected premium growth, claims frequency and severity, are taken into consideration for model building. An important task of a risk manager is to identify potential risks and integrate them into the model to enable a more comprehensive understanding of the model portfolio. As such models are often high-dimensional and have non-linear relationships between model inputs and output [51], the high complexity of model building necessitates the application of sensitivity analysis to identify the impact of model components and assumptions. Sensitivity analysis can have different aims depending on the problem posed. Typically, it aids in identification of model inputs that have the greatest influence on the model output and vice versa. By determining the most critical inputs that drive the output, input factors can be prioritised for further investigation or management intervention. Parameter and model uncertainty, which are persistent issues in insurance risk modelling, can be explored through sensitivity analysis by investigating the change in model predictions when a parameter is altered. Furthermore, sensitivity analysis can be used for quality and reasonableness checks as it allows comparing the model's interval dynamics with modeller's expectations.

Often sensitivity analyses can be time-consuming as the modeller is required to vary individual model features multiple times to observe the effect on the model output. As a result, the costs of performing a sensitivity analysis can be high. For this reason, it is vital that sensitivity analysis approaches are developed that are both efficient and when dealing with complex simulation models.

Saltelli [75] defines sensitivity analysis as the study of how uncertainty in the model's output is apportioned to various sources of uncertainty in model inputs. Through sensitivity analysis, a systematic approach can be applied to identify risk factors, small changes of which have a high impact on the output of a simulation model. This is particularly important when dealing with events that have a low probability of occurrence but severe financial implications. In cases where rare events are modelled using heavy-tailed distributions, sensitivity analysis becomes even more crucial; examples include modelling losses of a stock [70], flood risk [6], nuclear waste disposal [48] among others. The tail behaviour of a distribution is also studied in many actuarial and risk management applications. In the context of capital setting applications, for example Solvency II models, the evaluation of capital requirement is based on the 99.5th percentile of the distribution of losses over a one-year period. Furthermore, depending on the objectives of the modeller, sensitivity analysis can be used to uncover various behaviours of parameters in a model. Bermúdez et al. [14] use sensitivity analysis in non-life underwriting risk to evaluate the dependence of copula choice on the importance of correlation between business lines.

Several approaches to sensitivity analysis are available in literature. In recent years, there has been a shift in focus towards global sensitivity analysis in actuarial science applications due to the various shortcomings in local methods. Local sensitivity methods are limited in their ability to deal with complex models, as they focus on assessing model output sensitivity only around a specific point of interest. On the other hand, global sensitivity analysis proposed by Leamer [57] allows analysts to have a better understanding of model behaviour as it takes the entire space of input values into consideration. This approach has been motivated by the need to acknowledge the importance of considering alternative model structures and assessing the impact of each source of uncertainty on model predictions, in turn allowing the assessment of the robustness of model driven decisions. An extensive literature is available for the numerous sensitivity analysis techniques that have been proposed, see Borgonovo and Plischke [21], Saltelli [75]. We specifically focus on global methods, given the fundamental role of uncertainty in insurance applications.

A more recent study by Pesenti et al. [67] focuses on developing a global method, referred to as *reverse sensitivity analysis*, whereby a function of the output is stressed such that it corresponds to an increase in risk measures of the output. This approach attempts to determine the most influential factors in model, as those factors are associated with very adverse output states. Furthermore, the advantages of using such a method are manifold; factor prioritisation is based on the changes in the output distribution rather than a specific output state and requires only a single set of input/output scenarios, making it computationally cost-effective.

In the context of factor prioritisation, a sensitivity measure assigns a score to each input factor based on the effect it has on the model output. The most common approaches in evaluating sensitivity measures in the literature include moment-based and variance-based methods. They have been shown to insufficiently portray information especially when models are complex, involving correlated input factors. Further, it is possible that a change in one model input not only affects the output but indirectly affects other input factors in the model. Pesenti et al. [68] introduces a cascade sensitivity measure to accurately capture such spin-off effects to reflect the dependence structure between input factors. This is done by taking directional partial derivatives of risk measures applied to the output. For other literature closely relating to calculating directional derivatives, see Tsanakas and Millossovich [82].

We focus on developing a sensitivity analysis framework that closely follows the approach taken by Pesenti et al. [67] to evaluate the importance of inputs in relation to the output and conversely, to ascertain the significance of output with respect to changes in the inputs. Such an approach relies on f-divergence minimisation and thus relates to ideas that have been explored by many others such as Breuer and Csiszár [23], Csiszár [30], Cambou and Filipović [24]. f-divergences evaluate the dissimilarity between two probability distributions. The overall approach involves generating alternative scenarios relative to a baseline scenario by applying a stress on the baseline scenario. By doing this, the model is reconsidered under a more severe set of conditions. Alternative scenarios generated through this method are deemed plausible as they reflect the uncertainty while remaining broadly consistent with the baseline scenario. Often Kullback-Leibler divergence is heavily relied upon for many applications in statistics, information theory, finance and economics due to its relation with entropy. However, we note here that there are certain drawbacks for using Kullback-Leibler divergence which we aim to address in this thesis and the use of alternate f-divergences is explored.

The thesis consists of three self-contained chapters with their own introduction, notation and conclusion. Note that there may be some overlap and repetition of information presented throughout. Chapter 2 has been published in Insurance: Mathematics and Economics (IME) journal. A brief summary of the following chapters is given below.

1.2 Thesis overview

1.2.1 Sensitivity analysis with χ^2-divergence

In Chapter 2, we focus on developing a sensitivity analysis framework to investigate the importance ranking between various inputs in a model. While our approach to sensitivity analysis draws from the ideas of Pesenti et al. [67], we have added a novel approach where both reverse and forward sensitivity analyses can be integrated in a simple way to obtain a more comprehensive understanding of the behaviour of model inputs and their rankings. Novel sensitivity measures that assign a score to each input factor have been proposed, which reflect both reverse and forward sensitivity analyses.

To carry out the sensitivity analysis methodology, we use optimisation problems in a discrete space setting, reflecting e.g., a Monte Carlo simulation context. We rely on a change of measure that is derived by minimising the χ^2-divergence subject to a constraint on the expectation of a random variable and derive an explicit solution to the respective optimisation problem. By stressing a chosen variable, distortions to the empirical distributions are produced. The distortions are obtained such that they are close to the baseline model to reflect the level of underlying uncertainties in a model.

Our choice of minimising the χ^2-divergence stems from the fact that Kullback-Leibler divergence leads to somewhat less stable solutions when heavy-tailed distributions are modelled. Though in a simulation setting this disadvantage can be overlooked to a certain extent, we show that using the Kullback-Leibler divergence produces higher sampling errors compared to the χ^2-divergence.

1.2.2 Stress testing the solvency II standard model

Sensitivity analysis, in addition to identifying factors that are important, can be used to explore parameter uncertainty in the model. This is done by altering one parameter of the model individually and evaluating sensitivities of input factors. This allows us to pinpoint those parameters that contribute to model uncertainty significantly. Further, we can assess the impact a parameter has on the sensitivities of input factors by comparing risk factor sensitivities obtained for different parameter alterations.

In Chapter 3, we delve more specifically into understanding parameter uncertainty of a models' risk factors in the context of Solvency II. We specifically look at the premium and reserve risk module under the non-life underwriting risk as an example. We propose a stochastic model for the non-life premium and reserve risk module that broadly aligns with the standard formula, the distributional assumptions of which are partially determined by the regulator.

Two specific methods are used to investigate parameter uncertainty: re-simulation and re-weighting. While the re-simulation technique focuses on obtaining alternative scenarios, the re-weighting technique focuses on deriving alternative probabilities to a single set of simulated scenarios. We borrow and integrate the reverse and forward sensitivity analyses from Chapter 2 in both methodologies. In our analysis, we see that parameter uncertainty can have a significant impact on the model, particularly with respect to the model's correlation matrix.

1.2.3 Characterising minimal divergence stresses for general f-divergences

In Chapter 4, we delve deeper into studying f-divergences and the characteristics of solutions. Our aim here is to generalise the divergence minimisation problem stated in Chapter 2, which involved minimising the χ^2-divergence subject to a moment constraint. Our purpose here is to develop a common framework where different divergences can be implemented effectively to obtain alternate plausible models for stress testing. We provide results that allow for a characterisation of solutions when an f-divergence is minimised while, satisfying a constraint on the expectation of a random variable. The characterisation of the solution depends on the derivative value of the divergence at zero. This allows for the systematisation of f-divergences in terms of their features in the study of sensitivity analysis. Furthermore, we present results that address stochastic dominance and convex ordering of distribution functions when one parameter in the optimisation problem is altered. In particular, we see that the convexity of a function closely related to the chosen divergence is key for comparing the volatilities of stressed random variables. Finally through a simple example, we show the change in the influence of input factors in a model when different divergences are minimised.

Chapter 2

Sensitivity analysis using χ^2-divergence

With minor changes, this chapter has been published as:

Makam, V. D., Millossovich, P., and Tsanakas, A. (2021). Sensitivity analysis with χ^2-divergences. Insurance: Mathematics and Economics, 100:372–383.

Abstract

We introduce an approach to sensitivity analysis of quantitative risk models, for the purpose of identifying the most influential inputs. The proposed approach relies on a change of measure derived by minimising the χ^2-divergence, subject to a constraint ('stress') on the expectation of a chosen random variable. We obtain an explicit solution of this optimisation problem in a finite space, consistent with the use of simulation models in risk management. Subsequently, we introduce metrics that allow for a coherent assessment of reverse (i.e. stressing the output and monitoring inputs) and forward (i.e. stressing the inputs and monitoring the output) sensitivities. The proposed approach is easily applicable in practice, as it only requires a single set of simulated input/output scenarios. This is demonstrated by application on a simple insurance portfolio. Furthermore, via a simulation study, we compare the sampling performance of sensitivity metrics based on the χ^2- and the Kullback-Leibler divergence, indicating that the former can be evaluated with lower sampling error.
Keywords: Sensitivity analysis, χ^2-divergence, Kullback-Leibler divergence, simulation, sensitivity measures, reverse stress testing.
JEL codes: C15, G22, D81.

2.1 Introduction

2.1.1 Problem statement

Insurance and financial firms often employ complex quantitative models to analyse and evaluate the risks pertaining to their organisations; see McNeil et al. [64] for an overview of relevant methods and techniques. In insurance risk management applications, such models are typically implemented via Monte Carlo simulation. Scenarios are generated from modelled sources of uncertainty (*risk factors*) and are mapped via an *aggregation function* to *model outputs* of interest (e.g. the portfolio loss). Thus, aggregating risk factors allows the calculation of the probability distribution of model outputs. As the intricacy of such models increases, it becomes harder to develop insights from them and understand clearly the relationship between inputs and outputs [see, e.g. 82]. The complexity of quantitative risk models arises from the potential high-dimension and stochastic dependence of risk factors [e.g. 32, 4], as well as the non-linearity of the aggregation function [e.g. 50, 82], which may itself be numerically demanding in its evaluation at particular simulated scenarios [72, 38].

Our main focus in this paper is to develop an approach to sensitivity analysis, which enables users to rank model inputs by their importance, while being applicable to the simulation models used by insurers and financial firms. Sensitivity analysis generates insights into models and supports robust decision making; for comprehensive reviews see Christopher Frey and Patil [27], Saltelli [75], Saltelli et al. [77], Borgonovo and Plischke [21], Rabitti and Borgonovo [71].

We propose a sensitivity analysis framework relying on a change of measure, which requires only a single set of Monte-Carlo simulations, thus avoiding multiple model runs. The given set of simulations defines an (empirical) baseline probability measure. The model is stressed by a change of measure that should reflect specified distortions on the distributions of risk factors, with the stressed model remaining close to the baseline model. Specifically, working in a discrete probability space, we derive a change of measure by minimising the χ^2-divergence [29] with respect to the baseline model, subject to a constraint on the expectation of a model component (e.g. risk factor, model output, or a function thereof). The constraint reflects the desired stress on the variable of interest. We derive an explicit analytical solution to the relevant optimisation problem, which allows easy and efficient implementation.

Focusing on risk management applications, we use the terms *reverse* and *forward sensitivity analysis*, when the change of measure is, respectively, derived by stressing a model output or input. While forward sensitivity analysis refers to the well-understood problem of monitoring the impact of input changes on outputs, reverse sensitivity analysis [67] offers a generalisation of the reverse stress testing approach often used in risk regulation [36]. We develop a framework that combines the two analyses by, first, stressing the model output and evaluating the optimal χ^2-divergence and, second, maximising the expectations of input factors one at a time, while constraining the χ^2-divergence to the level obtained from the first step. By requiring that for both the reverse and forward stresses the χ^2-divergence is the same, we ensure that under all stresses applied the level of distortion to the baseline model is comparable and derived from an output stress specification, which is itself interpretable in risk management terms. In our view, this approach ensures the consistency of reverse and forward sensitivity analyses.

The changes in the distributions of inputs and output, under the reverse and forward stresses, are quantified via two novel sensitivity measures that we introduce in this paper. These sensitivity measures are associated with the above reverse/forward framework and enable the ranking of input factors, based on their importance in the model. We note that similar sensitivity measures can be defined if, in the relevant optimisation problems, we replace the χ^2-divergence with a different divergence measure – e.g. Kullback-Leibler divergence [30, 23]. By a numerical study we show that sensitivity measures based on the χ^2-divergence are obtained with lower sampling error, compared to the case when the Kullback-Leibler divergence is used.

2.1.2 Review of literature

In recent years, the literature on sensitivity analysis has largely focused on *global* methods, which reflect the model behaviour over the entire range of the input distribution; for comprehensive reviews see Christopher Frey and Patil [27], Saltelli [75], Saltelli et al. [77], Borgonovo and Plischke [21]. Major advances in sensitivity analysis can be accredited to Sobol [81], Homma and Saltelli [49], Saltelli et al. [77]. The range of sensitivity analysis methods available in the literature is substantial, with variance-based [77, 76] and moment-independent methods [19] being the most common. Recently, local and global sensitivity methods have been applied to evaluate the comparative importance of demographic and financial factors in an annuity portfolio [71]. Variance-based measures implicitly assume that knowledge of the second moment is sufficient to determine the uncertainty of an input factor, which is problematic in the case of heavy tails [59]. Efforts towards overcoming this shortcoming include the use of conditional Kullback-Leibler divergences, in order to quantify the importance of a model input [7, 59].

In this paper, we use the χ^2-divergence as a criterion for deriving stressed probability measures, under which the model's behaviour is examined. Hence our approach is more closely related to the literature involving divergence minimisation (under moment constraints) or moment maximisation (under divergence constraints). Specifically, we build on the reverse sensitivity testing approach proposed by Pesenti et al. [67], where the stressed probability measures are derived by minimising the Kullback-Leibler divergence, subject to a constraint on risk measures such as Value-at-Risk and Expected Shortfall. Working with the χ^2-divergence, we explore problems analogous to the ones stated in Breuer and Csiszár [23], who use Kullback-Leibler divergence in the context of model uncertainty. Model uncertainty is also addressed in Glasserman and Xu [44], by bounding the worst-case model error under a divergence constraint. However, in contrast to those papers, our focus is on understanding the sensitivities to risk factors within a given model rather than the study of model uncertainty.

Recently, Borgonovo et al. [20] have defined sensitivity measures that reflect the divergence between the unconditional distribution of model output and the conditional distribution, given an input factor, and have shown that several well-known sensitivity measures fall into this category. While our paper also utilises divergence measures and shares some conceptual parallels with Borgonovo et al. [20], it does not fall neatly under that framework, as the impact of a random variable is assessed by stressing its moment and then minimising divergence, rather than conditioning. Our use of the χ^2-divergence is motivated by the fact that the Radon-Nikodym derivative obtained when using the Kullback-Leibler divergence is typically exponential in form and, hence, when heavy-tailed distributions are used in a model, it might lead to issues with existence or (in a Monte Carlo setting) convergence. Related concerns are found in Glasserman and Xu [44], who use the α-divergence (of which the χ^2-divergence is a special case), when distributions are heavy-tailed. Similarly, Dey and Juneja [33] minimise a related divergence measure under linear constraints in a portfolio selection problem.

2.1.3 Structure of the paper

The rest of the paper is organised as follows. In Section 2.2, we discuss the Kullback-Leibler and χ^2-divergences. In Section 2.3, we provide the main result of the paper, relating to minimising χ^2-divergence, under a moment constraint. Furthermore, extensions and variations of the optimisation problem are considered. Finally, the reverse and forward sensitivity analysis framework is presented and the related sensitivity measures are defined. In Section 2.4, we apply our results to a simple non-linear insurance portfolio model. Furthermore, a simulation study is presented, where we assess the extent of simulation error in the evaluation of our sensitivity measures, when either χ^2- or Kullback-Leibler divergence is used. Brief conclusions are stated in Section 2.5.

2.2 Preliminaries

Let P and Q be two probability measures defined on a common measurable space $(\Omega, \mathscr{A})$. $Q \ll P$ indicates the absolute continuity of Q with respect to P and, in this case, we write the Radon-Nikodym derivative of Q with respect to P as $\frac{dQ}{dP}$. We denote the expectation operator under P and Q by $\mathbb{E}$ and $\mathbb{E}^Q$, respectively.

In the paper, we use special cases of the f-divergence [3, 58, 24], as measures of discrepancy between two probability measures.

Definition 2.2.1. Let $f : (0, \infty) \to \mathbb{R}$ be a convex function and suppose that $Q \ll P$. The f-divergence of Q with respect to P, denoted by $D_f(Q||P)$, is defined as

$$D_f(Q||P) = \int_\Omega f\left(\frac{dQ}{dP}\right) dP = \mathbb{E}\left[f\left(\frac{dQ}{dP}\right)\right].$$

The f-divergence is non-negative, monotone and jointly convex. The Kullback-Leibler (KL-) divergence, first introduced by Kullback and Leibler [55], and the χ^2-divergence [29, 58] are two special cases corresponding to $f(u) = u\log u$ and $f(u) = u^2 - 1$, respectively.

Definition 2.2.2. The KL-divergence of Q with respect to P with $Q \ll P$, is defined as

$$D_{KL}(Q||P) = \int_\Omega \frac{dQ}{dP}\log\left(\frac{dQ}{dP}\right)dP = \mathbb{E}^Q\left[\log\left(\frac{dQ}{dP}\right)\right].$$

The KL-divergence is positive, i.e. $D_{KL}(Q||P) > 0$, except if $Q = P$ when it becomes 0. It is also in general asymmetric i.e., $D_{KL}(Q||P) \neq D_{KL}(P||Q)$.

Definition 2.2.3. The χ^2-divergence of Q with respect to P with $Q \ll P$ is defined as

$$D_{\chi^2}(Q||P) = \int_\Omega \left(\left(\frac{dQ}{dP}\right)^2 - 1\right)dP = \mathbb{E}\left[\left(\frac{dQ}{dP}\right)^2\right] - 1 = \mathrm{var}\left(\frac{dQ}{dP}\right).$$

In Saraswat [78], it is shown that $D_{\chi^2}(Q||P) \geq D_{KL}(Q||P)$ for all P, Q.

Consider a finite probability space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ with $\mathscr{A} = 2^\Omega$. Assume that $Q \ll P$. Let p_i and q_i denote the probability of obtaining the state of the world $\omega_i \in \Omega$, under those two measures, that is, $p_i = P(\omega_i)$ and $q_i = Q(\omega_i)$. We assume that $p_i > 0$ for all i, while there may be states for which $q_i = 0$. Then, the definitions of KL-divergence and χ^2-divergence become:

$$\begin{aligned} D_{KL}(Q||P) &= \sum_i q_i \log\left(\frac{q_i}{p_i}\right) = \sum_i p_i w_i \log w_i \\ D_{\chi^2}(Q||P) &= \sum_i \left(\frac{{q_i}^2}{p_i}\right) - 1 = \sum_i p_i w_i^2 - 1, \end{aligned}$$

where $w_i = \dfrac{q_i}{p_i} = \dfrac{dQ}{dP}(\omega_i)$ for all i.

Finite spaces are typical in a Monte Carlo setting, where we have $p_i = \dfrac{1}{n}$, with each state of world corresponding to a simulated scenario, with equal probability of occurrence.

A risk measure is a functional ρ mapping a random variable X (a loss), to a real number $\rho(X)$ and it may represent e.g. the capital to be allocated in order to make the risk X acceptable. There are several ways of classifying of risk measures (see Artzner et al. [5], Föllmer and Schied [39]). We focus on the percentile-based risk measures Value-at-Risk (VaR) and Expected Shortfall (ES) [64].

Definition 2.2.4. The Value-at-Risk for risk X, at confidence level $\alpha \in (0,1)$, is defined as the left quantile of the distribution of X,

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha),$$

where $F_X^{-1}(\alpha) = \inf\{x \in \mathbb{R} | F_X(x) \geq \alpha\}$.

Definition 2.2.5. The Expected Shortfall for a risk X with $\mathbb{E}[|X|] < \infty$, at confidence level $\alpha \in (0,1)$, is given by

$$\begin{aligned}\mathrm{ES}_\alpha[X] &= \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_q(X)dq \\ &= \frac{1}{1-\alpha}\mathbb{E}\left[((X - \mathrm{VaR}_\alpha(X))_+\right] + \mathrm{VaR}_\alpha(X).\end{aligned}$$

From the above formula, the Expected Shortfall can also be interpreted as an average of Value-at-Risk for confidence levels greater than α, thus taking into account the entire tail of the distribution [74]. Expected Shortfall is a coherent risk measure, whereas the Value-at-Risk in general is not [5].

2.3 Stress testing models

2.3.1 Problem definition

We introduce a basic model within a sensitivity analysis framework, where the model inputs are mapped to a model output by means of an aggregation function. Let the random vector $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_d)$ on a measurable space $(\Omega, \mathscr{A})$ denote the random variables representing the input factors of the model under consideration. The aggregation function $g : \mathbb{R}^d \to \mathbb{R}$, when applied on the inputs, gives a one-dimensional output $Y = g(\mathbf{Z})$. We will throughout assume that high values of Y correspond to adverse outcomes, as is the case, for example, when Y represents the total loss of an insurance portfolio. The main focus of this paper concerns the understanding of model behaviour subject to changes in an input factor or the output. Specifically, we look at the changes in the distributional characteristics of input factors when there is a change in the output and vice versa.

Let $\mathscr{P}$ denote the set of all probability measures on the measurable space $(\Omega, \mathscr{A})$ and, for a given $P \in \mathscr{P}$, define the *baseline model* as $(\mathbf{Z}, g, P)$. A change of measure is introduced via the Radon-Nikodym derivative $W = \dfrac{dQ}{dP}$. We then refer to $(\mathbf{Z}, g, Q)$ as an *alternative* or *stressed model*.

The measure Q is chosen such that the expectation of random variable X becomes

$$\mathbb{E}^Q[X] = \mathbb{E}[WX] = t,$$

for a specified $t \in \mathbb{R}$. The variable X may be chosen to be one of the model inputs ($X = Z_i$), the model output ($X = Y$) or indeed a function of the input vector $\mathbf{Z}$. Depending on the problem context, the expectation of X may be stressed upwards ($t > \mathbb{E}[X]$) or downwards ($t < \mathbb{E}[X]$). The choice of $Q \in \mathscr{P}$ is such that the distortion to the baseline model is minimised. Specifically, we aim to minimise $D_{\chi^2}(Q|P)$, subject to the constraint $\mathbb{E}^Q[X] = t$ being fulfilled. In terms of the Radon-Nikodym derivative W, we arrive at the optimisation problem

$$\begin{cases} \min_W \frac{1}{2}\mathbb{E}[W^2] \quad \text{st} \\ \mathbb{E}[W] = 1, \\ \mathbb{E}[WX] = t, \\ W \geq 0. \end{cases} \tag{I}$$

Such a stress on X can be interpreted in two ways. First, we are concerned about model change. We can consider what would happen to the probability measure – and hence the distribution of *all* random variables of interest – if the expected value of X would move to the stressed value t. The second interpretation is concerned with model mis-specification. If the current model is not correctly specified, and the actual expectation of X is t, the Radon-Nikodym derivative arising as a solution to Problem (I) allows the calculation of a plausible distribution for all variables, under a corrected model. Note that in this paper we are not concerned with statistical arguments pertaining to how the baseline model was selected from data.

Portfolio models used in risk management typically require numerical evaluation of probability distributions of interest, with Monte Carlo simulation often used. For that reason, in the rest of this paper, we restrict our analysis to a finite probability space $\Omega = \{\omega_1, \ldots, \omega_n\}$, with baseline probability $P(\omega_i) = p_i$ for $i = 1, \ldots, n$. We denote by $\mathbf{w} = (w_1, \ldots, w_n)$, with $w_i = W(\omega_i)$, the vector of Radon-Nikodym derivative values, such that $Q(\omega_i) = q_i = p_i w_i$. Furthermore, let $X(\omega_i) = x_i$ and denote $\mathbf{x} = (x_1, \ldots, x_n)$. We assume that $x_1 < \ldots < x_n$; this is relaxed in Remark 2.3.5. In that context, Problem (I) becomes:

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2}\sum_{i=1}^{n} p_i w_i^2 \quad \text{st} \\ \sum_{i=1}^{n} p_i w_i = 1, \\ \sum_{i=1}^{n} p_i w_i x_i = t, \\ w_i \geq 0 \quad \text{for all} \quad i = 1, \ldots, n. \end{cases} \tag{II}$$

A ‘dual’ version of Problem (II) arises from maximising the expectation of a random variable with respect to the measure Q, subject to a constraint on the χ^2-divergence. A similar optimisation problem, with a constraint on the KL-divergence, is discussed in Breuer and Csiszár [23]. Here, we define the problem:

$$\begin{cases} \max \sum_{i=1}^{n} p_i v_i x_i \quad \text{s.t} \\ \sum_{i=1}^{n} p_i v_i = 1, \\ \frac{1}{2}\sum_{i=1}^{n} p_i v_i^2 \leq \theta, \\ v_i \geq 0 \quad \text{for all} \quad i = 1, \ldots, n. \end{cases} \tag{III}$$

Remark 2.3.1. The KL-divergence as a measure of plausibility of an alternate model is by far the most popular choice in the family of f-divergences. Applications in financial risk management include Breuer and Csiszár [23] and Glasserman and Xu [44]. Nonetheless, there are potential problems in the characterization of solutions obtained when the KL-divergence is used, if X follows a heavy-tailed distribution, as is often the case in insurance and finance applications. Specifically, if in Problem (I) we change the χ^2- to the KL-divergence, it is known that the optimal Radon-Nikodym derivative takes the form [30, 23]

$$W = \frac{\exp(\beta X)}{\mathbb{E}[\exp(\beta X)]},$$

for some $\beta \in \mathbb{R}$. The above expression is not well defined if X is heavy tailed, such that exponential moments are not defined (e.g. Log-normal or Student t). To avoid this pitfall Dey and Juneja [33] have replaced the KL-divergence with polynomial divergence in a portfolio selection problem. This also motivates our choice of the χ^2-divergence. In the case of a finite space (Problem II), issues of heavy-tailedness do not arise. However, if the discrete space is generated through the realisations of a Monte Carlo simulation, with the underlying model containing heavy tailed components, then convergence issues may appear – we return to this issue in Section 2.4.4.

2.3.2 Main results

In this section, we present the mathematical results of the paper, specifically the solution to the optimisation Problem (II) and its various corollaries.

Throughout this section, we use the following notations. Let the sum of the first j probabilities associated with each of the corresponding states of the world be denoted by $\pi_j = \sum_{i=1}^{j} p_i$ and similarly, indicate the sum of the probabilities corresponding to that of the latter $n-j$ states of the world by $\pi_{>j} = \sum_{i=j+1}^{n} p_i$. The mean, second moment and variance of $\mathbf{x}$ respectively are defined as:

$$\bar{x} = \sum_{i=1}^{n} p_i x_i, \quad \bar{x}^{(2)} = \sum_{i=1}^{n} p_i x_i^2, \quad s^2 = \bar{x}^{(2)} - \bar{x}^2.$$

For any integer $j \in \{1,\dots,n-2\}$, the mean of the first j values of $\mathbf{x}$ is given by:

$$\bar{x}_j = \frac{\sum_{i=1}^{j} p_i x_i}{\pi_j}.$$

The mean, second moment and variance of the latter $n-j$ values of $\mathbf{x}$ is given by:

$$\bar{x}_{>j} = \frac{\sum_{i=j+1}^{n} p_i x_i}{\pi_{>j}}, \quad \bar{x}_{>j}^{(2)} = \frac{\sum_{i=j+1}^{n} p_i x_i^2}{\pi_{>j}}, \quad s_{>j}^2 = \bar{x}_{>j}^{(2)} - \bar{x}_{>j}^2.$$

Proposition 2.3.1. Let $\bar{x} < t \leq x_n$. Then, the optimisation problem (II) has a unique solution $\mathbf{w}$, given below.

a) If $t < \bar{x} + \dfrac{s^2}{\bar{x} - x_1}$, then $w_i = \lambda_1 + \lambda_2 x_i > 0$, for $i = 1,\dots,n$,

   where $\lambda_2 = \dfrac{t - \bar{x}}{s^2}$ and $\lambda_1 = 1 - \lambda_2 \bar{x}$.

b) If $\bar{x} + \dfrac{s^2}{\bar{x} - x_1} \leq t < x_n$, then

$$w_i = \begin{cases} 0, & i = 1,\dots,k \\ l_1(k) + l_2(k) x_i, & i = k+1,\dots,n, \end{cases}$$

where $1 \leq k \leq n-2$ is the unique integer satisfying $l_1(k) + l_2(k)x_k \leq 0$ and $l_1(k) + l_2(k)x_{k+1} > 0$ and where the functions l_1 and l_2 are defined by

$$l_1(j) = \frac{1 - l_2(j)(\bar{x} - \bar{x}_j \pi_j)}{\pi_{>j}}$$

$$l_2(j) = \frac{t - \bar{x} - \dfrac{\pi_j}{\pi_{>j}}(\bar{x} - \bar{x}_j)}{\pi_{>j} s^2_{>j}} \qquad \text{for} \quad j = 1, \dots, n-2.$$

c) If $t = x_n$, then $w_i = \begin{cases} 0, & i = 1, \dots, n-1 \\ n, & i = n. \end{cases}$

Note that the positivity of the weights w_i in parts a), and in part b) for $i > k$, is guaranteed by the constraint on the parameter t.

Proposition 2.3.2. For a given t with $\bar{x} < t < x_n$, denote the solution of Problem (II) by $\mathbf{w}^*$ and the optimal value of the objective function by $\theta^* = \frac{1}{2} \sum_{i=1}^n p_i w_i^{*2}$. Then, $\mathbf{w}^*$ solves Problem (III) with $\theta = \theta^*$.

Finally, from the proof of Proposition 2.3.2 it can be seen that the χ^2-divergence constraint in Problem (III) is always binding at the optimum.

Remark 2.3.2. The Optimal χ^2-divergence in Problem (II) ranges from 0, corresponding to $t = \bar{x}$, to its maximum value, corresponding to $t = \max(X) = x_n$. Furthermore, the optimal χ^2-divergence is a strictly increasing function of t, thanks to the Sensitivity Theorem [61].

Remark 2.3.3. The increasingness of the optimal Radon-Nikodym derivative in Proposition 2.3.1 implies that the distribution of X under the stressed measure Q first order stochastically dominates the distribution of X under P, see for e.g. Pesenti et al. [67, Prop. A.1]. As a result the expectation of *any* increasing function of X is stressed upwards.

Remark 2.3.4. In Proposition 2.3.1, we state the solution to Problem (II) for an upward stress only, $\bar{x} < t \leq x_n$. Consider Problem (II) with a downward stress, that is $x_1 \leq t < \bar{x}$. Its solution is the same as that of the following problem:

$$\begin{cases} \min_{\mathbf{w}} \sum_{i=1}^n p_i w_i^2 \quad \text{s.t} \\ \sum_{i=1}^n p_i w_i = 1, \\ \sum_{i=1}^n p_i w_i r_i = -t, \\ w_i \geq 0 \quad \text{for all} \quad i = 1, \dots, n, \end{cases}$$

and where $r_1 = -x_n, r_2 = -x_{n-1}, \dots, r_n = -x_1$. This problem can be solved once again using Proposition 2.3.1, since $\bar{r} = -\bar{x} < -t \leq r_n = -x_1$.

Remark 2.3.5. In a Monte Carlo model where the states of the world are assumed to be equiprobable, the optimisation Problem (II) simplifies to

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2n} \sum_{i=1}^{n} w_i^2 \quad \text{s.t} \\ \frac{1}{n} \sum_{i=1}^{n} w_i = 1, \\ \frac{1}{n} \sum_{i=1}^{n} w_i x_i = t \\ w_i \geq 0 \quad \text{for all} \quad i = 1, \dots, n, \end{cases} \tag{IV}$$

The solution of Problem (II), as reported in Proposition 2.3.1, holds for Problem (IV), after substituting $p_i = \frac{1}{n}, \pi_j = \frac{j}{n}$ and $\pi_{>j} = \frac{n-j}{n}$.

Furthermore, assume that in addition to scenarios being equiprobable, we are in a situation where there are ties in $\mathbf{x}$. For example, if Y is a portfolio loss, we may be interested in stressing the random variable $X = (Y - \beta)_+$; in that case we may have $X(\omega_i) = 0$ for more than one state ω_i. In particular, assume that there is a unique tie consisting of $m+1$ values, $x_1 < x_2 < \cdots < x_{j-1} < x_j = x_{j+1} = \cdots = x_{j+m} < x_{j+m+1} < \cdots < x_n$. Then, we can replace Problem (IV) with

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2n} \sum_{i=1}^{\widetilde{n}} \widetilde{w}_i^2 \quad \text{s.t} \\ \frac{1}{\widetilde{n}} \sum_{i=1}^{\widetilde{n}} \widetilde{w}_i \widetilde{p}_i = 1, \\ \frac{1}{\widetilde{n}} \sum_{i=1}^{\widetilde{n}} \widetilde{w}_i \widetilde{p}_i \widetilde{x}_i = t, \\ \widetilde{w}_i \geq 0, \end{cases}$$

and $\widetilde{n} = n - m$,

$$\widetilde{x}_i = \begin{cases} x_i \quad \text{if} \quad i < j, \\ x_i \quad \text{if} \quad i = j, \\ x_{i+m} \quad \text{if} \quad i = j+1, \cdots, \widetilde{n}. \end{cases}$$

$$\widetilde{p}_i = \begin{cases} p_i \quad \text{if} \quad i < j, \\ p_j + p_{j+1} + \cdots + p_{j+m} \quad \text{if} \quad i = j, \\ p_{i+m} \quad \text{if} \quad i = j+1, \cdots, \widetilde{n}. \end{cases}$$

Remark 2.3.6. We have solved Problem (II) with a non-negativity constraint on the weights. Thus, information pertaining to some states of nature is lost when they are assigned a zero weight, i.e. if for the i^{th} scenario, we have $w_i = 0$. To avoid such a drastic intervention to the probability measure P, we slightly generalise Problem (II) by introducing a strictly positive lower bound δ for the weights. Specifically, for a given $t \in \mathbb{R}$ and $0 < \delta < 1$, consider the optimisation problem

$$\begin{cases} \min_{\mathbf{w}} \sum_{i=1}^n p_i w_i^2 \quad \text{s.t} \\ \sum_{i=1}^n p_i w_i = 1, \\ \sum_{i=1}^n p_i w_i x_i = t, \\ w_i \geq \delta > 0 \quad \text{for all} \quad i = 1, \dots, n. \end{cases} \tag{V}$$

The solution to Problem (V) follows from Problem (II), by the following argument. Let $\mathbf{v}^* = (v_1, \dots, v_n)$ be the solution of the auxiliary problem

$$\begin{cases} \min \sum_{i=1}^n p_i v_i^2 \quad \text{s.t} \\ \sum_{i=1}^n p_i v_i = 1, \\ \sum_{i=1}^n p_i v_i x_i = \dfrac{1}{1-\delta}(t - \delta \bar{x}), \\ v_i \geq 0 \quad \text{for all} \quad i = 1, \dots, n. \end{cases}$$

Then, $\mathbf{w}^* = \delta + (1-\delta)\mathbf{v}^*$ is the solution to the Problem (V). This can be verified by substituting $\mathbf{w}^*$ in the constraints of Problem (V). The objective function of Problem (V) becomes:

$$\begin{aligned} \sum_{i=1}^n p_i w_i^2 &= \sum_{i=1}^n p_i (\delta + (1-\delta) v_i)^2 \\ &= \delta^2 + 2\delta(1-\delta) + (1-\delta)^2 \sum_{i=1}^n p_i v_i^2. \end{aligned}$$

Hence, minimising the left hand side is equivalent to minimising $\sum_{i=1}^n p_i v_i^2$.

2.3.3 Reverse and forward sensitivity analyses

Here we return to the problem definition of Section 2.3.1, considering a model with output Y and risk factors $\mathbf{Z}$, linked through an aggregation function, $Y = g(\mathbf{Z})$. Depending on the purpose of the sensitivity analysis, we may set X in Problem (I) as either $X = Y$, leading to a *reverse* sensitivity analysis (see also Pesenti et al. [67]), or $X = Z_i$, a *forward* sensitivity analysis. Reverse sensitivity analysis aims at evaluating the behaviour of risk factors under a stress on the model output (portfolio loss), while forward sensitivity is concerned with the impact on the output distribution of stressing individual risk factors.

Specifically, for reverse sensitivity analysis, we solve Problem (II) with the constraint $\mathbb{E}^Q(Y) = t$, where $t > \mathbb{E}[Y]$ represents a stress on the expected value of the output Y, and let $\frac{dQ_Y}{dP}$ the optimal Radon-Nikodym derivative. As high values of Y are interpreted as adverse outcomes, the value of t can be selected to represent a critical threshold for a decision maker. Subsequently, the input factors' importance is assessed according to the impact on their distribution, caused by the change of measure $\frac{dQ_Y}{dP}$. A substantial change observed in the distribution of an input factor can be interpreted as a high sensitivity of that factor.

Conversely, for forward sensitivity testing, a change of measure is obtained by specifying a stress on one input factor at a time. In order that the stresses on different input factors are consistent with each other, we obtain the relevant changes of measure by solving Problem (III) with $X = Z_i$ and under the same constraint on the χ^2-divergence. Denote the resulting Radon-Nikodym derivatives by $\frac{dQ_{Z_i}}{dP}$, $i = 1, \dots, d$. Then, these changes of measure are used to evaluate stressed distributions of the output Y; we attribute a higher sensitivity to input factors that lead to a more substantial change in the distribution of Y.

Furthermore, we can link reverse and forward sensitivity, to ensure consistency between the stresses applied under each of the two approaches and detect any dissonance that may arise between the importance rankings they produce. Here, we propose the following process. We start with reverse sensitivity analysis, as a stress on the output may be calibrated with reference to an unacceptable level of adverse movement in portfolio risk (Problem (II)). Subsequently, the optimal χ^2-divergence is calculated from that analysis. Then, this divergence value is used as a constraint in Problem (III) to find the maximal stress possible on an input factor for the forward sensitivity analysis.

Two sensitivity measures specific to our framework are defined below.

Definition 2.3.1. Let $\mathbb{E}[Y] < t < \max Y$, Q_Y be the probability measure arising from the solution of Problem (II) with $X = Y$, and denote by θ^* the corresponding optimal value of the objective function. Let Q_{Z_i} be the probability measure arising from the solution of Problem (III) with $X = Z_i$ and $\theta = \theta^*$. Then, the *reverse sensitivity* of an input Z_i is defined by

$$R_i := \frac{\mathbb{E}^{Q_Y}[Z_i] - \mathbb{E}[Z_i]}{\mathbb{E}^{Q_{Z_i}}[Z_i] - \mathbb{E}[Z_i]},$$

while the *forward sensitivity* of Z_i is defined as

$$F_i := \frac{\mathbb{E}^{Q_{Z_i}}[Y] - \mathbb{E}[Y]}{\mathbb{E}^{Q_Y}[Y] - \mathbb{E}[Y]}.$$

The sensitivity measure R_i (resp. F_i) represents the change in the expectation of an input (resp. output), when the output (resp. input) is stressed. The denominators act a normalising constants, as is seen from Proposition 2.3.3 below. We remark that the sensitivity analysis framework we present, including Definition 2.3.1, can be altered as necessary to include functions of input factors to enable the assessment of different distributional characteristics.

Proposition 2.3.3. The sensitivity measures of Definition 2.3.1 satisfy the following properties:

1. R_i, F_i are well defined.
2. $R_i, F_i \leq 1$.
3. $R_i = F_i = 0$ if Z_i, Y are independent.
4. $R_i, F_i \geq 0$ if (Z_i, Y) are positive quadrant dependent.

Remark 2.3.7. In the reverse/forward stress testing framework proposed in this section, we have assumed throughout that the starting point is a stress that increases the mean of the output Y to a level t. We believe that this has an appealing risk management interpretation, as it allows us to consider the way that a specified adverse movement in the distribution of Y (e.g. a portfolio loss) is reflected in corresponding movements of input factors. Of course, an alternative analysis can be carried out, with $t < \mathbb{E}[Y]$, thus considering improvements in Y.

Furthermore, the formulation of the metrics R_i and F_i encodes an analyst's prior expectations on a positive relationship between the risk factors and model input. This is implicit in the process of stressing *upwards* a risk factor Z_i in order to observe a (presumably) adverse effect on Y. Put differently, unless such a prior expectation exists, an analyst may choose to *minimise* rather than maximise $\mathbb{E}^Q[Z_i]$ in Problem (III), subject to the same constraint on the χ^2-divergence. However, as sensitivity analysis is a process of discovery, the analyst's expectations may be confounded, e.g. if $F_i < 0$ is observed, signifying that an upward stress in Z_i results in a reduction in $\mathbb{E}^{Q_{Z_i}}[Y]$. Though such a negative value is interpretable in its own right, the analyst may subsequently choose to carry out a different forward stress, involving minimisation of $\mathbb{E}^Q[Z_i]$, and recalculate the forward sensitivity measure accordingly. To formalise this argument, consider F_i as in Definition 2.3.1, and let $\tilde{F}_i$ be the same quantity, with the only difference that in the derivation of Q_{Z_i}, minimisation replaces maximisation in Problem (III). Then one could consider the couple $(F_i, \tilde{F}_i)$, or indeed some quantity like $\max\{|F_i|, |\tilde{F}_i|\}$, as an importance measure that does not rely on prior expectations – though the latter would not reveal the direction of association between Z_i and Y.

Remark 2.3.8. A further observation relates to the extent that our approach can be extended in order to allow the stressing of second moments. For example, in Problem (I) one can clearly set $X = Z_i^2$, to capture volatility effects, or $X = Z_i Z_j$, to capture interaction effects. However, such a stress specification would lead to results that are hard to interpret, since under the stressed model with such a second order constraint, the first moments would also move. A more appropriate method of stressing second moments would involve introducing more constraints to Problem (I), e.g. fixing the first moments to their values under the baseline measure P and stressing the second moments to higher values. However, for such a procedure our main analytical result, Proposition 2.3.1, can no longer be used and one would need to revert to numerical optimisation. While this is tractable using quadratic programming methods [e.g. 45], we believe that it lies outside the scope of the current paper.

2.4 Case study of an insurance portfolio

Here we apply the framework of Section 2.3 to the example of a simplified insurance portfolio. In Section 2.4.1, we introduce the model, while in Sections 2.4.2 and 2.4.3 we, respectively, perform reverse and forward sensitivity analyses. Finally, in Section 2.4.4, we evaluate the sensitivity measures of Definition 2.3.1; furthermore, we examine their sampling performance, comparing them to similar measures that are constructed by replacing the χ^2-with KL-divergence.

2.4.1 Baseline model

Consider a model of an insurance portfolio, with inputs factors $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)$ and output Y, representing the portfolio loss. Z_1 and Z_2 represent claims from two lines of business. Claims are subject to a common multiplicative (e.g. inflation) factor, Z_3, such that the portfolio loss, before reinsurance, is given by

$$L = (Z_1 + Z_2) Z_3.$$

The insurance company buys reinsurance on L with limit l and deductible d. Z_4 represents the percentage of reinsurance recovery lost in circumstances when the re-insurer fails to make a payment. The total portfolio loss thus is:

$$Y = L - (1 - Z_4) \min\{(L - d)_+, l\}.$$

Z_1 follows a truncated Log-normal distribution with mean 150 and standard deviation 35, where the truncation point is at the 99.9% quantile; Z_2 follows a Gamma distribution with mean 200 and standard deviation 20; Z_3 follows a Log-normal distribution with mean 1.05 and standard deviation 0.05; Z_4 follows a Beta distribution with mean 0.1 and standard deviation 0.2. We assume that Z_1, Z_2, Z_3 are independent. Furthermore, Z_4 is dependent on L through a Gaussian Copula with a correlation of 0.6 and, conditional on L, Z_4 is independent of (Z_1, Z_2, Z_3). For the reinsurance parameters, we set $l = 30$ and $d = 380$. We simulate $(\mathbf{Z}, Y)$ using a Monte Carlo sample of $n = 10^5$ scenarios.

2.4.2 Reverse sensitivity analysis of the insurance model

Using the above model, we follow the sensitivity analysis process outlined in Section 2.3.3. We denote by Q_Y the measure for which $\frac{dQ_Y}{dP}$ is the solution of Problem (II) after setting $X = Y$. We stress the expectation of Y upwards by 10%, such that $\mathbb{E}^{Q_Y}(Y) = 1.1,\ \mathbb{E}(Y) = t$.

Figure 2.1 (left) displays the Radon-Nikodym derivative of the stressed probability measure Q_Y, as a piecewise linearly increasing function of Y. On the right of Figure 2.1, the empirical distributions of Y under the baseline (dashed) and stressed (solid) measure are shown. The stressed output distribution first-order stochastically dominates the output distribution under the baseline model, as remarked after Proposition 2.3.1.

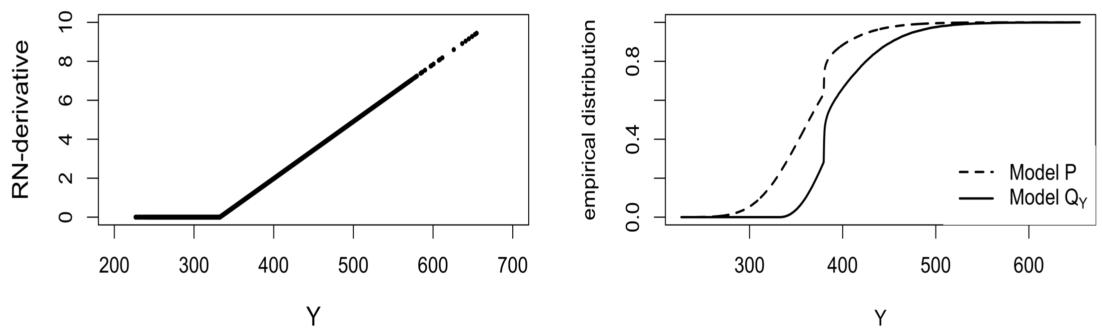


Fig. 2.1 Left: Radon-Nikodym derivative of Q_Y against Y. Right: Stressed probability distributions of Y under models P, Q_Y.

Figure 2.2 displays the distribution of the input factors under the stressed model Q_Y. The stressed probability distributions appear to stochastically dominate the baseline distributions. We can see that Z_1 and Z_4 undergo a larger change, compared to Z_2 and Z_3. We attribute this behaviour to the heavier tail of Z_1 and the role of Z_4 in the aggregation function, since the loss of reinsurance recoveries is important in those scenarios where losses L before reinsurance are high.

These observations are confirmed in Table 2.1, which reports the percentage increases in the mean, standard deviation, and VaR/ES risk measures, at the 95% level, of the four input factors. If for example we focus on $\mathrm{ES}_{0.95}$, we observe an approximate increase of 15% and 18% for Z_1 and Z_4 respectively, with the corresponding values for Z_2 and Z_3 being much lower.

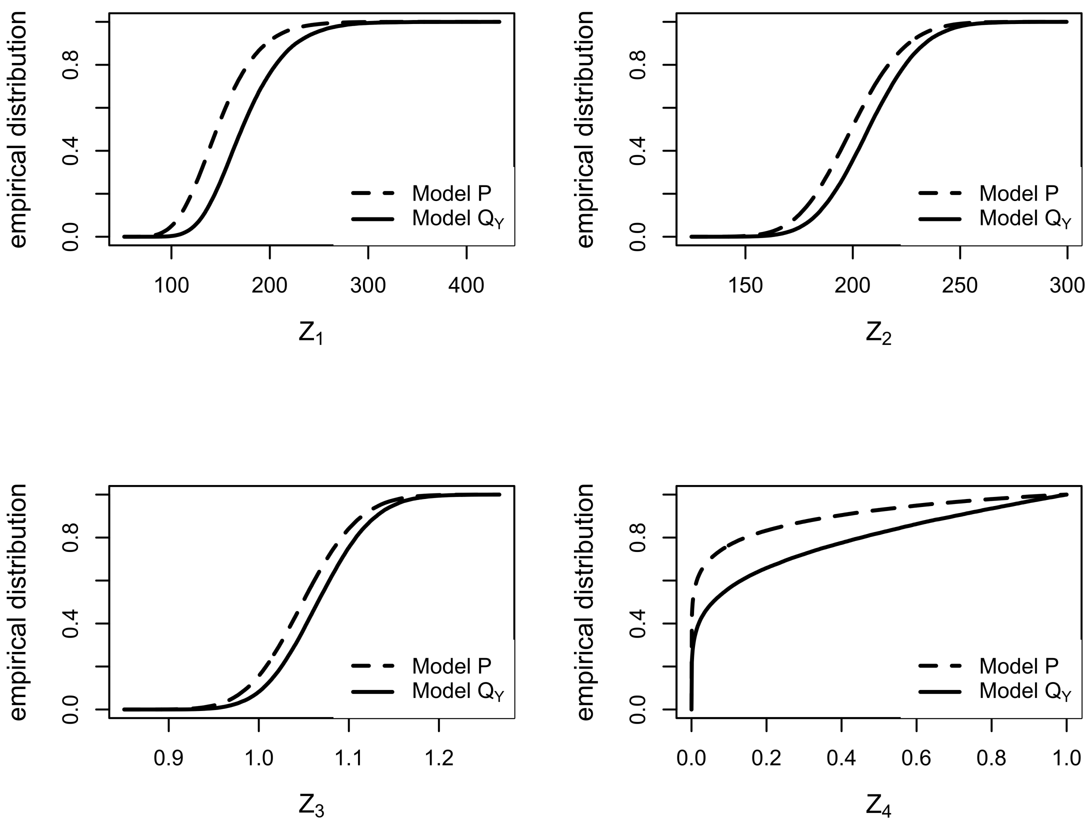


Fig. 2.2 Empirical distributions of the input factors under the baseline and stressed models P, Q_Y.

Table 2.1 Percentage increase in statistics of output Y and input factors Z_i under the stressed model Q_Y, with respect to the baseline model P.

Factors	Mean	St. Dev.	$\text{VaR}_{0.95}$	$\text{ES}_{0.95}$
Y	10.0	−1.31	11.17	10.90
Z_1	17.44	8.67	15.38	14.79
Z_2	3.99	−0.80	3.27	3.14
Z_3	1.60	−1.48	1.39	1.35
Z_4	108.52	39.63	39.18	18.10

2.4.3 Forward sensitivity analysis of the insurance model

Now we carry out forward sensitivity analysis, as discussed in Section 2.3.3. We denote by Q_{Z_i} the measure for which $\frac{dQ_{Zi}}{dP}$ is the solution of Problem (III) after setting $X = Z_i$ and θ equal to the optimal χ^2-divergence of the reverse sensitivity problem in Section 2.4.2. Figure 2.3 displays the Radon-Nikodym derivative of the stressed probability measures Q_{Z_i}, $i = 1, \dots, 4$. It is seen that each Radon-Nikodym derivative is an increasing function of the factor being stressed. Note that while the different Radon-Nikodym derivatives have the same standard deviation (due to the χ^2-divergence constraint) their distributions are generally not the same.

In Figure 2.4, the empirical distributions of Y under the baseline (P, dashed black) and all stressed (Q_{Z_i}, grey; Q_Y, black) models are displayed. As each input factor is subject to a stress with the same optimal χ^2-divergence, arising from the reverse analysis, the stressed measures under the forward analysis cannot produce greater distortions to the distribution of Y compared to that obtained in Section 2.4.2. This is evident from Figure 2.4, where we can see that the red lines are always between the black and dashed grey ones. This is precisely the effect that the Definition 2.3.1 of sensitivity measures aims to reflect.

We observe that greater distortions to the distribution of Y arise under stressed models Q_{Z_1}, Q_{Z_4}, compared to Q_{Z_2} and Q_{Z_3}, implying a higher sensitivity to Z_1 and Z_4. This is broadly consistent with the observations of Section 2.4.2. In Table 2.2 we report percentage changes in distributional characteristics of Y, under the stresses on all input factors. We note that, for example, the largest changes in the 95%-ES measure are observed for Z_1 and Z_4, 9.2% and 8.4% respectively.

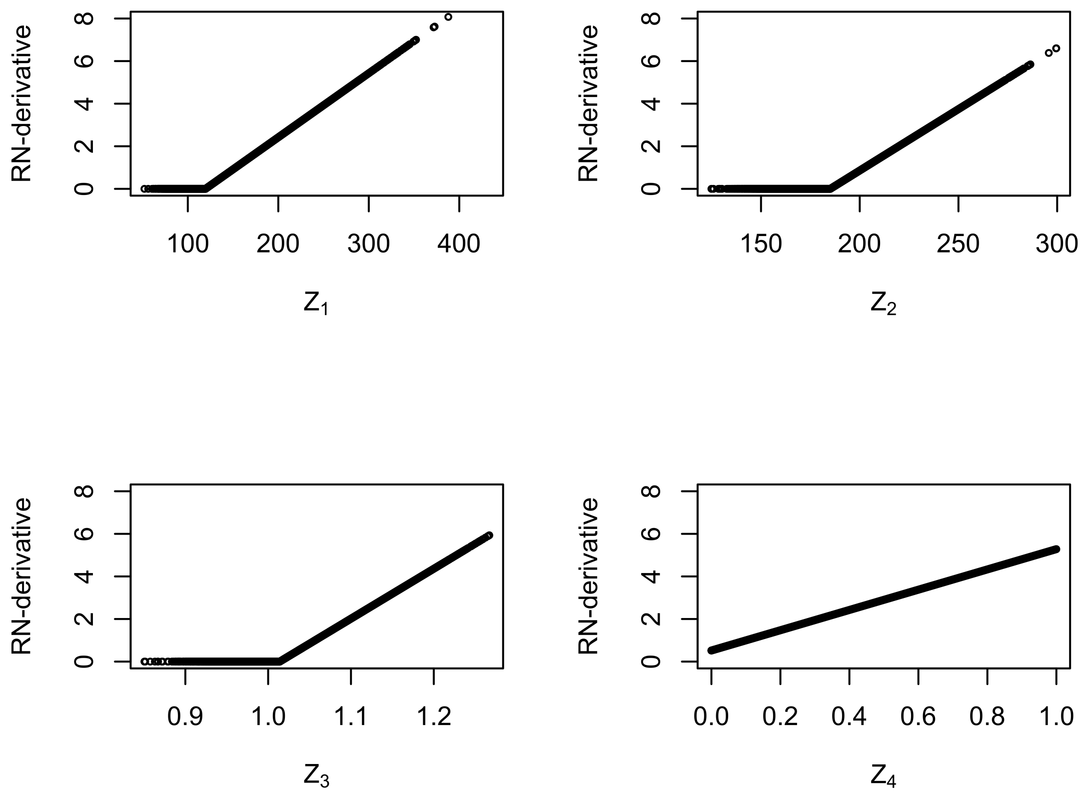


Fig. 2.3 Radon-Nikodym derivatives of stressed models Q_{Z_i} for $i = 1, \dots, 4$
.

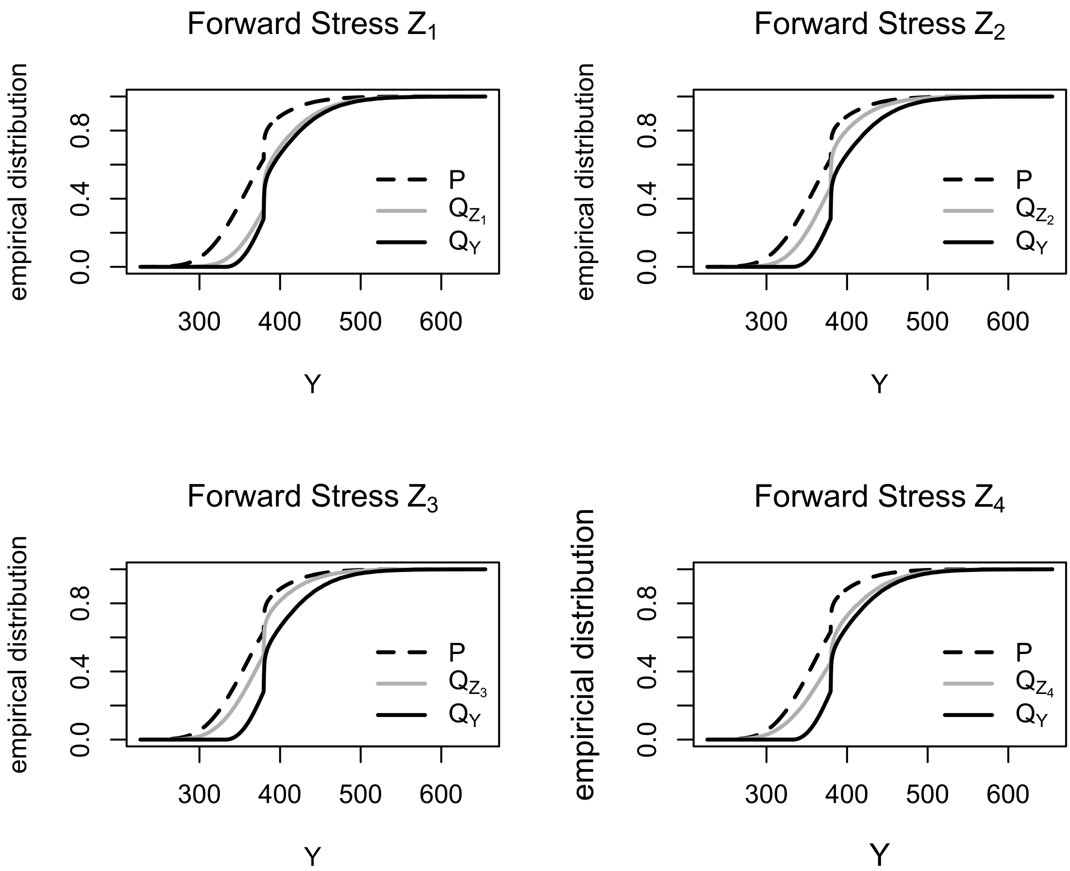


Fig. 2.4 Stressed probability distributions of Y under the baseline model P and the stressed models Q_Y and Q_{Z_i} for $i = 1, \ldots, 4$.

Table 2.2 Percentage increase in statistics of Y under the stressed models Q_{Z_i} for $i = 1, \ldots, 4$, with respect to the baseline model P.

Variables stressed	Mean	St. Dev.	$\text{VaR}_{0.95}$	$\text{ES}_{0.95}$
Z_1	8.00	2.46	9.41	9.21
Z_2	4.52	−4.44	4.65	4.34
Z_3	3.72	−0.32	4.35	4.15
Z_4	5.51	17.94	8.80	8.41

2.4.4 Evaluation of sensitivity measures

The aim of this section is to evaluate the sensitivity measures defined in Section 2.3.3 for our insurance portfolio model and assess the extent of simulation error in their calculation.

The reverse and forward sensitivities of different input factors are reported, respectively, in the second and fourth column of Table 2.3, such that e.g. $R_1 = 0.794$ and $F_1 = 0.800$. It can be seen that, according to both the reverse and forward sensitivity measures, the ranking of risk factors, from the most to the least sensitive, is Z_1, Z_4, Z_2, Z_3. This is broadly consistent with the discussion of Sections 2.4.2 and 2.4.3.

Furthermore, for comparison purposes, in the third and fifth column of Table 2.3, we report sensitivity measures calculated with respect to the KL- rather than the χ^2-divergence. These sensitivity measures are still calculated according to Definition 2.3.1, with the difference that the measures Q_Y, Q_{Z_i} are the solutions of modified versions of Problems (II) and (III), with the χ^2-divergence replaced with the KL-divergence. The solution to these problems is given by e.g. [23] and the numerical implementation is carried out via the **R** package SWIM by [66]. We observe that a change in the divergence measure does not impact the relative importance of input factors.

Table 2.3 Reverse and forward sensitivities of input factors Z_1, Z_2, Z_3, Z_4 under χ^2-divergence and KL-divergence (calculated as the average over 1000 sets of $n = 10^5$ simulated scenarios).

	Reverse SM		Forward SM	
Input	χ^2-divergence	KL-divergence	χ^2-divergence	KL-divergence
Z_1	0.794	0.809	0.800	0.806
Z_2	0.433	0.389	0.451	0.417
Z_3	0.370	0.356	0.374	0.346
Z_4	0.568	0.570	0.551	0.580

To quantify simulation error, we simulate m sets of n simulated scenarios from our model. The sensitivity measures are evaluated on each of the m sets of simulations, resulting in empirical distributions representing sampling error. Specifically, for $k = 1, \dots, m$, we follow the algorithm:

1. Multivariate scenarios $\mathbf{z}^{(k)}$ are sampled from $\mathbf{Z}$ under P, where $\mathbf{z}^{(k)} = \left(z_{j,i}^{(k)}\right)_{\substack{j=1,\dots,n \\ i=1,\dots,d}}$.
Subsequently, evaluate $\mathbf{y}^{(k)} = \left(y_j^{(k)}\right)_{j=1,2,\dots,n}$, where $y_j^{(k)} = g\left(\mathbf{z}_{j\bullet}^{(k)}\right)$ and $\mathbf{z}_{j\bullet}^{(k)} = (z_{ji}^{(k)})_{i=1,\dots,d}$.
2. Set $t^{(k)} = 1.1\frac{1}{n}\sum_{j=1}^{n} y_j^{(k)}$ for the reverse sensitivity test.
3. Working first with the χ^2-divergence, we obtain the corresponding Radon-Nikodym densities $(\mathbf{w}_j^{(k)})_{j=1,\dots n}$ by solving Problem (II) with $\mathbf{x} = \mathbf{y}^{(k)}$ and $t = t^{(k)}$.

4. Evaluate the optimal divergence, $\theta^{(k)} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{w}_j^{(k)})^2$.

5. For the forward sensitivity test, set $\theta = \theta^{(k)}$ and solve Problem (III) with $\mathbf{x} = \mathbf{z}_{\bullet i}^{(k)}$, where $\mathbf{z}_{\bullet i}^{(k)} = (z_{ji}^{(k)})_{j=1,\dots,n}$, to obtain the Radon-Nikodym densities $\mathbf{w}_i^{(k)} = (w_{ji}^{(k)})_{j=1,\dots,n}$.

6. Using $\mathbf{w}^{(k)}$ and $\mathbf{w}_i^{(k)}$, we measure the reverse and forward sensitivity measures R_i, F_i as given in Definition 2.3.1.

In addition, we carry out the same algorithm, but using the KL-divergence for the calculation of sensitivity measures, as discussed above. We aim to compare the simulation error of sensitivity measures under each of the two divergence measures. This is motivated by Remark 2.3.1, where we argued that, due to the form of the solution of the KL-divergence minimisation problem, high numerical errors may arise.

Figure 2.5 displays box plots of input factors' sensitivity measures. The top left and right box plots are associated with reverse sensitivity with χ^2- and KL-divergences respectively, while the bottom two plots represent forward sensitivities for the two divergences. We observe greater volatility in the estimates of both reverse and forward sensitivities, when the KL-divergence is used. This is particularly visible in the case of the reverse sensitivity, where the KL-divergence produces a high number of outliers. This confirms our concerns raised in Remark 2.3.1 about the use of the KL-divergence and demonstrates the better numerical properties of sensitivity measure estimates, when the χ^2-divergence is used.

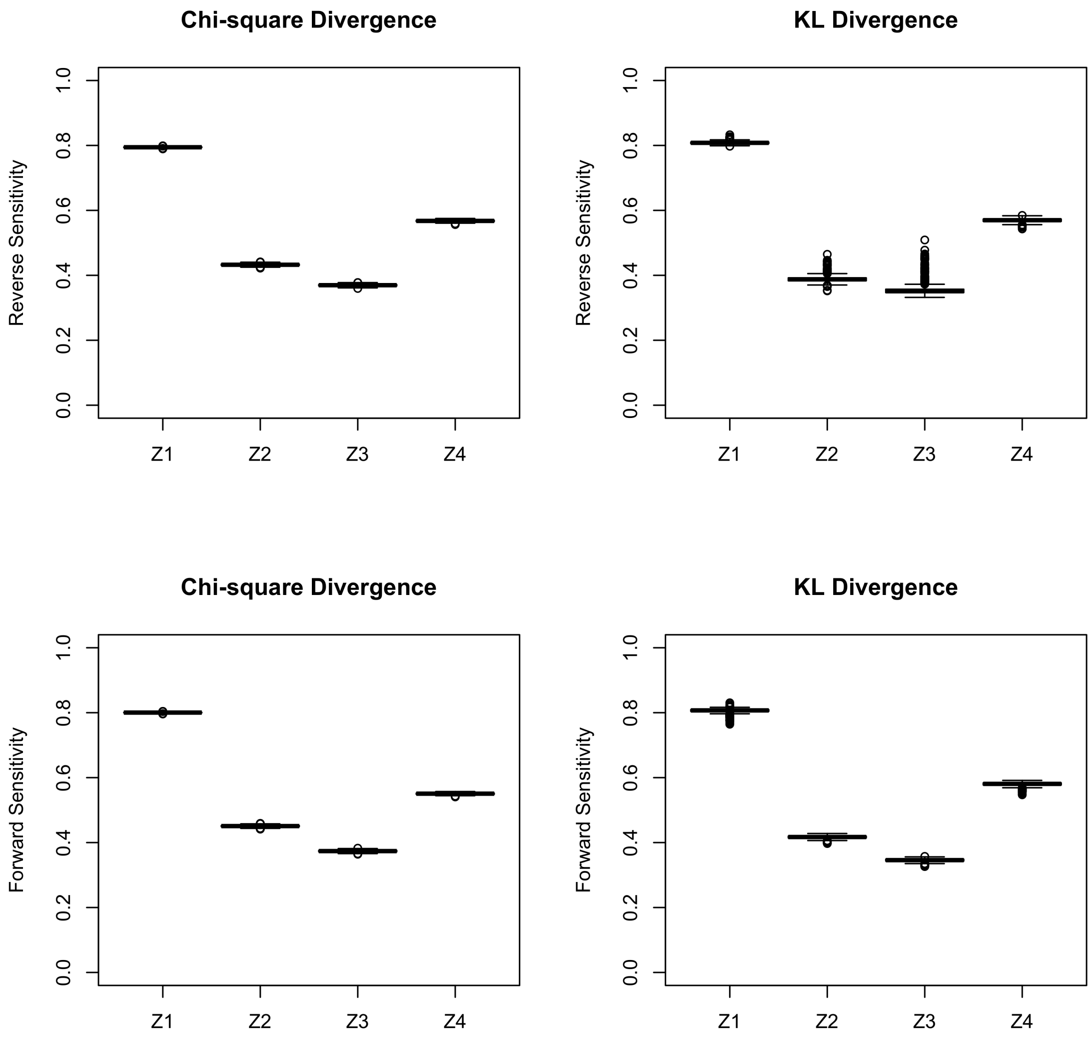


Fig. 2.5 Box plots of reverse and forward sensitivities of input factors under χ^2- and Kullback-Leibler divergences, for $m = 1000$ sets of $n = 10^5$ simulated scenarios.

In Table 2.4, we show the standard errors of reverse and forward sensitivities of input factors Z_1, Z_2, Z_3, Z_4 under the χ^2- and KL-divergences, for $m = 1000$ sets of $n \in \{10^3,\ 10^4,\ 10^5\}$ simulated scenarios. Once more, we observe the higher error of sensitivity measures based on the KL-divergence, particularly for lower sample sizes n.

Table 2.4 Standard errors of reverse and forward sensitivities of input factors Z_1, Z_2, Z_3, Z_4 under χ^2-divergence and KL-divergence for $m = 1000$ sets of $n = 10^3, 10^4, 10^5$ simulations.

	Reverse Sensitivity		Forward Sensitivity	
$n = 10^3$	χ^2-divergence	KL-divergence	χ^2-divergence	KL-divergence
Z_1	0.013	0.023	0.012	0.034
Z_2	0.028	0.057	0.027	0.034
Z_3	0.029	0.088	0.029	0.036
Z_4	0.029	0.046	0.026	0.040
$n = 10^4$	χ^2-divergence	KL-divergence	χ^2-divergence	KL-divergence
Z_1	0.004	0.010	0.004	0.016
Z_2	0.009	0.026	0.008	0.012
Z_3	0.009	0.045	0.009	0.012
Z_4	0.009	0.015	0.008	0.014
$n = 10^5$	χ^2-divergence	KL-divergence	χ^2-divergence	KL-divergence
Z_1	0.001	0.004	0.001	0.006
Z_2	0.003	0.010	0.003	0.005
Z_3	0.003	0.019	0.003	0.004
Z_4	0.003	0.005	0.003	0.006

2.5 Conclusions

We have proposed a sensitivity analysis framework based on the χ^2-divergence, to investigate in a coherent fashion the relationship between a model's inputs and output. Two approaches to sensitivity analysis are considered; for the reverse approach, the expectation of the output was stressed to ascertain the output to input relationship whereas, for the forward approach, the input factors were stressed subject to the same optimal divergence. The analytical solution obtained for the divergence minimisation problem allows an easy implementation of the sensitivity analyses using Monte-Carlo simulation. We introduced sensitivity measures specific to our framework, to investigate the changes in the distributions of inputs and output. Finally, a numerical study is presented, comparing the simulation error of sensitivity measures based on the KL- and χ^2-divergences. The lower errors observed in the case of the χ^2-divergence and its applicability in the context of heavy-tailed distributions, make it a competitive alternative to the more commonly used KL- divergence.

2.6 Appendix: Proofs

Proposition (2.3.1). If $t \leq x_n$, Problem (II) is a quadratic programming problem which admits a unique solution. The Karush-Kuhn-Tucker (KKT) conditions will then be both necessary and sufficient for optimality of a candidate solution **w** [61].

The KKT conditions are:

$$
\begin{aligned}
p_i w_i = p_i\lambda_1 + p_i\lambda_2 x_i + \mu_i, &\qquad \sum_{i=1}^{n} p_i w_i = 1,\\
w_i\mu_i = 0, &\qquad \sum_{i=1}^{n} p_i w_i x_i = t,\\
\mu_i \geq 0, &\qquad w_i \geq 0,
\end{aligned}
$$

for $i = 1, \ldots, n$.
To find the general form of λ_1 and λ_2, we substitute the equation $p_i w_i$ in the equality constraints of Problem II. We get

$$\lambda_1 = 1 - \lambda_2\bar{x} - \sum_{i=1}^{n} \mu_i, \tag{i}$$

$$\lambda_2 = \frac{t - \bar{x} + \sum_{i=1}^{n} \mu_i(\bar{x} - x_i)}{s^2}. \tag{ii}$$

We note that $w_i > 0$ implies that $\mu_i = 0$ and $w_i = \lambda_1 + \lambda_2 x_i$.

We now show that $\lambda_2 > 0$. Let's suppose by contradiction that $\lambda_2 \leq 0$ and consider the case where $x_h < x_j$ for some indices $1 \leq h < j \leq n$ such that $w_j > 0$. It follows that $w_j = \lambda_1 + \lambda_2 x_j$ and

$$
\begin{aligned}
p_h w_h &= p_h\lambda_1 + p_h\lambda_2 x_h + \mu_h \\
&\geq p_h\lambda_1 + p_h\lambda_2 x_h && (\text{since} \quad \mu_h \geq 0)\\
&\geq p_h(\lambda_1 + \lambda_2 x_j) && (\text{since} \quad x_h < x_j)\\
&= p_h w_j.
\end{aligned}
$$

We conclude that w_i is non-increasing in i and that there is a counter-monotonic relationship between X and W. In the case where $w_j = 0$, the conclusion still holds. Therefore, by

Chebyshev's Sum Inequality,

$$t = \sum_{i=1}^{n} p_i x_i w_i \leq \sum_{i=1}^{n} p_i w_i \sum_{i=1}^{n} p_i x_i = \bar{x}$$

which contradicts $t > \bar{x}$. Therefore, $\lambda_2 > 0$.

Let now $x_h < x_j$ for $1 \leq h < j \leq n$ such that $w_h > 0$. Then $\mu_h = 0$ and we have

$$\begin{aligned} p_j w_j &= p_j \lambda_1 + p_j \lambda_2 x_j + \mu_j \\ &\geq p_j \lambda_1 + p_j \lambda_2 x_j \\ &> p_j(\lambda_1 + \lambda_2 x_h) \\ &= p_j w_h. \end{aligned}$$

Hence w_i is non-decreasing in i and the solution will be of the form

$$w_i = \begin{cases} 0 & i < k^* \\ \lambda_1 + \lambda_2 x_i & i \geq k^* \end{cases} \tag{iii}$$

for some $k^* \in \{1, \ldots, n\}$, where k^* is the smallest index such that $w_{k^*} > 0$.

Note that the implications in the statement of the proposition can be inverted as the three cases are mutually exclusive and exhaustive. If $\mathbf{w}$ is the unique solution of Problem II, we will proceed by proving the following:

a) If $w_i > 0$ for $i = 1, \ldots, n$, then $t < \bar{x} + \dfrac{s^2}{\bar{x} - x_1}$.

b) If $w_i = 0$ for some i and $w_j > 0$ for at least two indices j, then $\bar{x} + \dfrac{s^2}{\bar{x} - x_1} \leq t < x_n$.

c) If $w_i = 0$ for all i but one, then $t = x_n$.

We proceed with the proof by considering three different cases for k^* and establish the condition on t for each case.

Case $k^* = 1$:

Let $k^* = 1$, which implies that $w_i > 0$ for all $i = 1, 2, \ldots, n$. Therefore, from (iii), the solution is $w_i = \lambda_1 + \lambda_2 x_i > 0$ for any $i = 1, 2, \ldots, n$.

The general formulas derived for λ_1 and λ_2 in equations (i) and (ii) simplify as follows:

$$\lambda_1 = 1 - \bar{x}\lambda_2, \quad \lambda_2 = \frac{t-\bar{x}}{s^2}.$$

In order to obtain a condition on t we substitute λ_1 and λ_2 in w_i, to get

$$w_i = 1 - \frac{t-\bar{x}}{s^2}(\bar{x} - x_i).$$

As all w_i are positive, $t < \bar{x} + \dfrac{s^2}{\bar{x} - x_i}$ for each i. Since the x_i's are increasing, this is equivalent to $t < \bar{x} + \dfrac{s^2}{\bar{x} - x_1}$.

<u>Case $1 < k^* < n$</u>:

We let $k^* = k+1$ for some $1 \leq k \leq n-2$.
Thus

$$w_i = \lambda_1 + \lambda_2 x_i + \frac{\mu_i}{p_i} = 0 \quad \text{for} \quad i \leq k, \tag{iv}$$

$$w_i = \lambda_1 + \lambda_2 x_i > 0 \quad \text{for} \quad i > k. \tag{v}$$

Rearranging the terms in (iv), we get $\mu_i = -(\lambda_1 + \lambda_2 x_i)p_i$ for $i \leq k$ and subsequently, substituting for μ_i in equations (i) and (ii), we solve for λ_1 and λ_2.
Solving for λ_1:

$$\lambda_1 = 1 - \lambda_2\bar{x} - \sum_{i=1}^{k} \mu_i = \frac{1 - \lambda_2(\bar{x} - \bar{x}_k\pi_k)}{\pi_{>k}}. \tag{vi}$$

Solving for λ_2 gives:

$$\lambda_2 = \frac{t - \bar{x} + \sum_{i=1}^{k} \mu_i(\bar{x} - x_i)}{s^2},$$

which leads to

$$\lambda_2 s^2 = t - \bar{x} - \sum_{i=1}^{k} (\lambda_1 p_i + \lambda_2 p_i x_i)(\bar{x} - x_i).$$

Hence

$$\lambda_2 = \frac{t - \bar{x} - \pi_k(\bar{x}_{>k} - \bar{x}_k)}{\left(s^2 - \pi_k\left(\bar{x}_{>k}(\bar{x} - \bar{x}_k) - \bar{x}\bar{x}_k + \bar{x}_k^{(2)}\right)\right)}.$$

After some algebra, the denominator becomes

$$s^2 - \pi_k\left(\bar{x}_{>k}(\bar{x} - \bar{x}_k) - \bar{x}\bar{x}_k + \bar{x}_k^{(2)}\right) = \pi_{>k}s^2_{>k}.$$

Therefore,

$$\lambda_2 = \frac{(t - \bar{x}) - \dfrac{\pi_k}{\pi_{>k}}(\bar{x} - \bar{x}_k)}{\pi_{>k}s^2_{>k}}. \tag{vii}$$

We know that from the KKT conditions, $0 = w_k = \lambda_1 + \lambda_2 x_k + \dfrac{\mu_k}{p_k}$. Since, $\dfrac{\mu_k}{p_k} \geq 0$, we have $\lambda_1 + \lambda_2 x_k \leq 0$. Substituting the values of λ_1 and λ_2 in the above, we get:

$$1 - \frac{(t - \bar{x}) - \dfrac{\pi_k}{\pi_{>k}}(\bar{x} - \bar{x}_k)}{\pi_{>k}s^2_{>k}}(\bar{x} - \bar{x}_k\pi_k - x_k\pi_{>k}) \leq 0.$$

Therefore, the last inequality can be written as

$$t - \bar{x} \geq \frac{A(k)}{B(k)} \tag{viii}$$

where, $A(i) = \pi_{>i}s^2_{>i} + \pi_i(\bar{x} - \bar{x}_i)(\bar{x}_{>i} - x_i)$ and $B(i) = \pi_{>i}(\bar{x}_{>i} - x_i)$.
To see that $\dfrac{A(i)}{B(i)}$ is increasing in i note that, after some algebra, we have

$$A(i)B(i+1) - A(i+1)B(i) \leq 0 \implies \pi_{>i}s^2_{>i}(x_i - x_{i+1}) \leq 0.$$

Setting $i = 1$, we get $\dfrac{A(1)}{B(1)} = \dfrac{s^2_{>1} + \pi_1(\bar{x}_{>1} - \bar{x}_1)(\bar{x}_{>1} - x_1)}{\bar{x}_{>1} - x_1} = \dfrac{s^2}{\bar{x} - x_1}$.
Since $t = \sum_{i=1}^n p_i w_i x_i = \sum_{i=k^*}^n p_i w_i x_i$ and $k^* < n$, it follows that $t < x_n$.

To find the value of k, we use $0 = w_k = \lambda_1 + \lambda_2 x_k + \dfrac{\mu_k}{p_k} \geq \lambda_1 + \lambda_2 x_k$ and $w_{k+1} = \lambda_1 + \lambda_2 x_{k+1} > 0$.
Hence, k will be the unique value such that $\lambda_1 + \lambda_2 x_k \leq 0 < \lambda_1 + \lambda_2 x_{k+1}$. By noting the

dependence of λ_1 and λ_2 on k, through equations (vi) and (vii), the expression for calculating k, that is given in the proposition's statement, follows.

Case $k^* = n$:

We get $w_n = n$ and $w_i = 0$ for $i = 1, \dots, n-1$. In such a case, it is clear from the second constraint of Problem II that $t = x_n$. □

Proposition 2.3.2. It can be confirmed that $\mathbf{w}^*$ is a solution to Problem (III) by verifying that it satisfies the KKT conditions, by choosing $\eta_2^* = \frac{-1}{\lambda_2}, \eta_1^* = \frac{\lambda_1}{\lambda_2}$ and $\varepsilon_i^* = \frac{\mu_i}{\lambda_2}$, where λ_1, λ_2 and μ_i are the Lagrangian multipliers in Problem (II).

The KKT conditions for for Problem (III), $i = 1, \dots, n$, are:

$$\begin{aligned}
\eta_2 p_i v_i &= -p_i x_i - \eta_1 p_i - \varepsilon_i, & v_i \varepsilon_i &= 0,\\
\eta_2 \left(\frac{1}{2} \sum_{i=1}^{n} p_i v_i^2 - \theta \right) &= 0 & \varepsilon_i &\geq 0\\
\sum_{i=1}^{n} p_i v_i &= 1, & \eta_2 &\leq 0,\\
\frac{1}{2} \sum_{i=1}^{n} p_i v_i^2 &= \theta, & v_i &\geq 0.
\end{aligned}$$

As (III) is a convex problem, satisfying the KKT conditions is necessary and sufficient for $\mathbf{w}^*$ to be a solution. □

Proposition (2.3.3).

1. The denominator of F_i is strictly positive, by assumption. The denominator of R_i is strictly positive by Proposition 2.3.2.

2. For $R_i \leq 1$, $\mathbb{E}^{Q_Y}(Z_i) \leq \mathbb{E}^{Q_{Z_i}}(Z_i)$ must hold. This follows from $D_{\chi^2}(Q_Y || P) = D_{\chi^2}(Q_{Z_i} || P)$ and Q_{Z_i} being the maximiser in Problem (III).
The claim $F_i \leq 1$ follows similarly, by considering Problem (III) and Proposition 2.3.2, for $X = Y$.

3. If (Z_i, Y) are independent, $\mathbb{E}^{Q_Y}(Z_i) = \mathbb{E}(Z_i)$ and $\mathbb{E}^{Q_{Z_i}}(Y) = \mathbb{E}(Y)$, implying directly that $R_i = F_i = 0$.

4. Let $\eta(Y) = \frac{dQ_Y}{dP}$. From Proposition 2.3.1, we know that η is a non-decreasing function. By the PQD assumption it follows that $\mathbb{E}^{Q_Y}(Z_i) = \mathbb{E}(\eta(Y) Z_i) \geq \mathbb{E}(Z_i)$, which shows that $R_i \geq 0$. The case $F_i \geq 0$ is similar.

□

Chapter 3

Stressing the solvency II standard model

Abstract

Solvency II provides a risk management framework for insurance companies by setting out strict regulatory requirements to reflect the level of risk undertaken. We apply the reverse-forward sensitivity analysis framework from Makam et al. [62] to a stochastic model for non-life premium and reserve risk, which partially adheres to the standard formula Solvency II. We investigate model and parameter uncertainty, by employing bootstrapping approaches, namely *re-simulation* and *re-weighting*, to assess the impact of a parameter change on the sensitivities of risk factors. While re-simulation involves generating multiple sets of scenarios under different distributional assumptions, the re-weighting technique involves constructing alternate probabilities for one given set of simulated scenarios. We highlight the importance of considering parameter uncertainty in sensitivity studies by evaluating its impact on risk factors' sensitivity ranking and examining how this impact is itself affected by variations in distributional assumptions.
Keywords: Solvency II, parameter uncertainty, sensitivity analysis, sensitivity measures.

3.1 Introduction

In Chapter 2, we introduced a sensitivity analysis framework that involved both reverse and forward sensitivity approaches and relied on prior knowledge about the underlying statistical model. In this chapter, we focus on a Solvency II-inspired model and specifically examine the non-life premium and reserve risk sub-module. Furthermore, we assess the way that the sensitivity of the portfolio risk to different lines of business is impacted by parameter uncertainty.

The Solvency II directive for insurance firms in Europe, implemented in 2016, aims to establish strict capital requirements to reflect the level of risk undertaken by firms and to safeguard interests of policyholders, by ensuring that insurers have sufficient resources to meet their obligations. Of the three pillars constituting Solvency II, the first pertains to the quantitative requirements for the level of capital that a company must maintain, referred to as the solvency capital requirement (SCR). The SCR is calculated to be the capital necessary for a company in order to meet its obligations with a minimum of 99.5% likelihood within one year [25]. This definition is applied individually to all sub-modules for calibration [35] and typically risk measures such as Value-at-risk are used for computing SCR. For a detailed overview of Solvency II, see Eling et al. [37].

In this chapter, we examine the non-life premium and reserve risk sub-module of the Solvency II standard formula. Under Solvency II, insurance companies have the option to use a standard formula, an internal model or a partial internal model to assess their risk-based capital requirements. Although companies often rely on the standard formula for its convenience, using a stochastic model instead of a deterministic one would allow analysts to have a more comprehensive understanding of their risks. A stochastic model can enhance the firms' capabilities to apply stress testing and sensitivity analysis. In Section 3.3, we construct a stochastic model for the non-life premium and reserve risk sub-module incorporating various distributional assumptions, while partially adhering to the standard formula. This enables risk analysts to gain further insights into different business lines – e.g. by assessing their relative importance – with such insights obtained at a lower cost compared to developing an internal model from scratch.

Nonetheless, model and parameter uncertainty will persist around the assumptions behind any stochastic model. Model uncertainty refers to the uncertainty surrounding the choice of the model used (e.g. the family of (joint) distribution), while parameter uncertainty involves the uncertainty around the selected parameters, given the model [16, 8]. The question then arises as to how such uncertainties impact the results of sensitivity analyses and particularly the relative importance ranking of risk factors. Investigating this question forms the main contribution of this chapter, focusing on the issue of parameter uncertainty.

Parameter uncertainty arises from limited samples being used to estimate parameters, resulting in estimates that can deviate from the true parameters substantially [17]. The risk associated with parameter uncertainty can have a significant impact on the company's probability of insolvency. Gerrard and Tsanakas [42] show that for a class of loss distributions derived by increasing transforms of location-scale, ignoring parameter uncertainty while calculating risk capital can substantially increase the probability of insolvency beyond the 0.5% threshold. Both underestimating and overestimating parameter risk can have drawbacks. For example, underestimation of parameter risk would lead to an increased risk of insolvency and overestimating the risk can lead to incorrect management decisions [40].

In this chapter, we assume that the distributions of risks are known, and the parameters are estimated. To tackle parameter uncertainty, several approaches in the existing literature point to bootstrapping and Bayesian procedures. The bootstrap procedure assumes that the true parameter of a risk distribution is unknown but fixed, and the uncertainty around the estimate is modelled as a random variable, from which new observations can be drawn in order to quantify the parameter uncertainty [63, 69, 40]. On the other hand, the Bayesian approach treats the unknown parameter as a realisation of a random variable with a prior distribution. For further details on the Bayesian approach in the context of insurance risk, see Borowicz and Norman [22], where parameter uncertainty is studied in the context of extreme event frequency-severity model using data on Danish fire losses. Diers et al. [34] also study the importance of parameter uncertainty in the context of premium risk in a multi-year horizon using the Bayesian approach. Other methods in the existing literature to measure parameter uncertainty include raising the confidence level to adjust the risk capital [42], the asymptotic normality approach [34] or the capital add-on method [40].

We explore parameter uncertainty in the Solvency II-inspired model proposed for calculating non-life premium and reserve risk capital requirements. To this effect, we propose two different bootstrapping procedures are implemented: re-simulation and re-weighting. The re-simulation approach, often used by practitioners, involves generating different sets of simulated scenarios for risk factors, under alternative parameter estimate realisations obtaining a set of simulated scenarios using current information. In addition, we consider a re-weighting scheme, whereby a single set of risk factor simulations is used, which are assigned different probabilities, under alternative parameter assumptions.

Our objective is to investigate model and parameter uncertainty through a sensitivity analysis framework. We adopt the reverse and forward sensitivity analysis techniques presented in Makam et al. [62], as described in Chapter 2. This methodology allows us to evaluate how the ranking of model inputs changes when we implement alternative model parameterisations. In our analysis, we modify the model assumptions independently and observe the changes in the sensitivities and rankings of model input factors. We use the reverse sensitivity metric to evaluate the sensitivities of risk factors. We note that when a parameter is changed, there is a significant impact on the rankings of risk factors. Although the most/least important risk factors retain their spots across different scenarios, their actual sensitivity values may be much higher/lower for some comparisons. Furthermore, we find that the sensitivities' error bounds often overlap, indicating a lack of robustness in sensitivity evaluation, as the ranking of risk factors may depend on the particular realisation of parameter estimators.

The rest of the chapter is structured as follows: In Section 3.2, we present the reverse and forward sensitivity analysis frameworks from Makam et al. [62], along with the re-simulation and re-weighting approaches to parameter uncertainty. In Section 3.3, we present the stochastic model for non-life premium and reserve risk sub-modules, as well as the statistical assumptions and data used. Section 3.4 covers the implementation of bootstrapping procedures on the proposed statistical model and the investigation of the impact of parameter uncertainty. Finally in Section 3.5, we summarise our conclusions.

3.2 Stress testing by change of measure

3.2.1 Reverse and forward sensitivities

We consider a standard sensitivity analysis framework. A model has d inputs $\mathbf{Z} = (Z_1, \ldots, Z_d)$. These inputs or *risk factors* are mapped to a model output Y via an *aggregation function*, $g : \mathbb{R}^d \to \mathbb{R}$, such that $Y = g(\mathbf{Z})$. The model output Y is understood as a portfolio loss.

The statistical behaviour of the risk factors $\mathbf{Z}$ is governed by an underlying parametric distribution $\mathbf{Z} \sim F(\cdot|\theta^0)$, which reflects an analyst's current best estimates. For the sake of practicality, a set of n multivariate scenarios is simulated from $F(\cdot|\theta^0)$ and we denote these n realisations of $\mathbf{Z}$ by $\mathbf{z}_1, \ldots, \mathbf{z}_n$.

Restricted to those scenarios, the analyst is essentially working on a discrete probability space, where each realisation $\mathbf{z}_i = (z_{1,i}, \ldots, z_{d,i}),\ i = 1, \ldots, n$ is a state itself, taking the same probability $p_i = 1/n$. We call this the *reference model*. Different simulation exercises could hence give rise to different reference models, in this terminology. The corresponding realisations of the portfolio loss Y are denoted by $y_i = g(\mathbf{z}_i),\ i = 1, \ldots, n$ and we write $\mathbf{y} = (y_1, \ldots, y_n)$. Without loss of generality, we assume that $y_1 < \ldots < y_n$.

For the purpose of sensitivity analysis, the analyst also considers *stressed versions of a reference model*. Under a stressed model, different probability weights are assigned to different states of the reference model. Specifically, under the stressed model, the probability of the state $\mathbf{z}_i,\ i = 1, \ldots, n$ is given by w_i/n, where the vector $\mathbf{w} = (w_1, \ldots, w_n)$ with $w_i \geq 0$ and $\frac{1}{n}\sum_{i=1}^{n} w_i = 1$ plays the role of a Radon-Nikodym derivative.

In this paper, we conduct reverse and forward sensitivity analyses to evaluate the relative importance of risk factors in the presence of model uncertainty. Reverse sensitivity analysis [67] focuses on the distributional changes in the risk factors if the output distribution is stressed in a certain way, while forward sensitivity analysis considers the effect on the model output when an input is stressed. For further details on the two sensitivity analysis approaches, see Makam et al. [62]. The two analyses are implemented in tandem as they provide a fuller understanding of factor importance. The optimisation problems formulated in Makam et al. [62] for the reverse and forward sensitivity analysis are as follows. First, we consider the problem:

$$\begin{cases} \min_{\mathbf{w}\in\mathbb{R}^n} \frac{1}{2}\sum_{i=1}^n p_i w_i^2 \quad \text{s.t.} \\ \sum_{i=1}^n p_i w_i = 1, \\ \sum_{i=1}^n p_i w_i \zeta(y_i) \geq t, \\ w_i \geq 0 \quad \text{for all} \quad i = 1,\dots,n. \end{cases} \tag{I}$$

In this reverse analysis, the χ^2-divergence is minimised with a constraint on the mean of Y, or of any function ζ of Y under the stressed model. Problem (I) allows us to find a stressed version of the reference model subject to a desired shock on the output.

The forward analysis is an extension of the reverse sensitivity analysis. For the forward analysis, the χ^2-divergence is first evaluated as $\lambda^* = \sum_{i=1}^n p_i w_i^{*2}$, where $\mathbf{w}^* = (w_1^*,\dots,w_n^*)$ is the solution of Problem (I). The mean of a risk factor is then maximised with a constraint on the divergence. For each fixed Z_j, $j = 1,\dots,d$, the problem is as follows:

$$\begin{cases} \max_{\mathbf{v}\in\mathbb{R}^n} \sum_{i=1}^n p_i v_i \zeta_j(z_{j,i}) \quad \text{s.t} \\ \sum_{i=1}^n p_i v_i = 1, \\ \frac{1}{2}\sum_{i=1}^n p_i v_i^2 \leq \lambda^*, \\ v_i \geq 0 \quad \text{for all} \quad i = 1,\dots,n. \end{cases} \tag{II}$$

where ζ_j is some function of the j-th risk factor.

The importance of the risk factors Z_j can be quantified through a measure that evaluates the difference between the probability distributions before and after stressing the mean of Y. Here, for our analysis, we use the *reverse sensitivity measure* of Makam et al. [62] to rank each Z_j according to the change in its expectation. The reverse sensitivity measure for factor Z_j is

$$R_j = \frac{\sum_{i=1}^n p_i w_i^* z_i - \sum_{i=1}^n p_i z_i}{\sum_{i=1}^n p_i v_{j,i}^* z_i - \sum_{i=1}^n p_i z_i}$$

where, $\mathbf{v}_j^* = (v_{j,1}^*,\dots,v_{j,n}^*)$ is the solution of Problem (II). The greater the change in a risk factor's distributional characteristics, the greater its sensitivity score will be.

3.2.2 Reflecting parameter uncertainty

Our aim is to examine how model uncertainty impacts the assessment of input factors' sensitivity using the reverse sensitivity framework described above. Specifically, we generate models that are plausible alternatives to the reference model. Each of those alternative models is stressed and the reverse sensitivity measure is subsequently evaluated for all risk factors.

The distribution of sensitivities across models reveals the way that model uncertainty impacts factor importance. The generation of alternative models can be carried out in two distinct ways, which are discussed below.

Parameter uncertainty by re-simulation

Alternative models are obtained via bootstrapping. We start from the underlying parametric distribution $F(\cdot|\theta^0)$, which governs the behaviour of risk factors. In each of b bootstrap iterations, a data set of size m is simulated from $F(\cdot|\theta^0)$. Each of those data sets is used to re-estimate the parameters of F. Subsequently, for each such different set of parameters, we simulate n risk factor values and, on these, evaluate the sensitivity measures.

Through this process we obtain b sets of risk factor realisations $\mathbf{z}_i^{(h)} = (z_{1,i}^{(h)}, \ldots, z_{d,i}^{(h)}),\ i = 1,\ldots,n,\ h = 1,\ldots,b$. For each h, the realisations $\mathbf{z}_i^{(h)},\ i = 1,\ldots,n$ with corresponding probabilities $1/n$ give an *alternative model*. The *stressed version of the h-th alternative model*, has again states $\mathbf{z}_i^{(h)}$, but now with re-weighted probabilities $w_i^{(h)}/n$ attached to them.

Specifically, for $h = 1,\ldots,b$, we proceed as follows:

1. Generate a multivariate sample of size m from the underlying parametric model, $(\tilde{\mathbf{z}}_1^{(h)}, \ldots, \tilde{\mathbf{z}}_m^{(h)}) \sim F(\cdot|\theta^0)$, where $\tilde{\mathbf{z}}_k^{(h)} = (\tilde{z}_{1,k}^{(h)}, \ldots, \tilde{z}_{d,k}^{(h)})$, $k = 1,\ldots,m$.
2. Estimate the distribution parameters $\hat{\theta}^{(h)} \equiv \hat{\theta}(\tilde{\mathbf{z}}_1^{(h)}, \ldots, \tilde{\mathbf{z}}_m^{(h)})$.
3. Simulate n scenarios $(\mathbf{z}_1^{(h)}, \ldots, \mathbf{z}_n^{(h)}) \sim F(\cdot|\hat{\theta}^{(h)})$, where $\mathbf{z}_i^{(h)} = (z_{1,i}^{(h)}, \ldots, z_{d,i}^{(h)}), i = 1,\ldots,n$.
4. Evaluate $y_i^{(h)} = g(\mathbf{z}_i^{(h)}), i = 1,\ldots,n$.
5. Set $t^{(h)} = (1+\beta)\frac{1}{n}\sum_{i=1}^n \zeta(y_i^{(h)}),\ \beta > 0$.
6. Solve Problem (I) for the Radon-Nikodym density $w_i^{*(h)}$ by setting $y_i = y_i^{(h)}, p_i = \frac{1}{n}$ and $t = t^{(h)}$ and evaluate the χ^2-divergence $\lambda^{*(h)} = \frac{1}{n}\sum_{i=1}^n (w_i^{*(h)})^2$.
7. For fixed $j = 1,\ldots,d$, set $\lambda = \lambda^{*(h)}$ in Problem (II) and solve the Radon-Nikodym-density $v_{j,i}^{*(h)}$ by setting $z_{j,i} = z_{j,i}^{(h)}, i = 1,\ldots,n$ and $p_i = \frac{1}{n}$.
8. Evaluate $R_j^{(h)} = \dfrac{\sum_{i=1}^n w_i^{*(h)} z_{j,i}^{(h)} - \sum_{i=1}^n z_{j,i}^{(h)}}{\sum_{i=1}^n v_{j,i}^{*(h)} z_{j,i}^{(h)} - \sum_{i=1}^n z_{j,i}^{(h)}}$ for $j = 1,\ldots,d$.

Parameter uncertainty by re-weighting approach

Here we introduce a second approach to generating alternative models reflecting parameter uncertainty. Rather than re-simulating, we consider a scenario re-weighting procedure. While in the approach of Section 3.2.2 each alternative model is represented by a different set of equiprobable scenarios, here we fix the set of simulated scenarios and vary the probabilities attached to them.

We start again with an underlying parametric distribution $F(\cdot|\theta^0)$ for the risk factors and simulate n risk factor values $\mathbf{z}_i = (z_{1,i},\dots,z_{d,i}),\ i = 1,\dots,n$ from it. On the discrete space with states corresponding to $\mathbf{z}_i$, we consider reference probabilities $p_i^0 = \frac{1}{n}$. As before, we denote $y_i = g(\mathbf{z}_i)$.

Following that, we re-estimate model parameters via the same bootstrap procedure as before in Section 3.2.2. Subsequently, b alternative models are generated by varying the probabilities attached to each realisation $\mathbf{z}_i,\ i = 1,\dots,n$. These probabilities are calculated via a likelihood ratio, induced by the parametric model with re-estimated parameters, $F(\cdot|\theta^{(h)}),\ h = 1,\dots,b$. The corresponding probabilities for each alternative model are denoted by $p_i^{(h)}$, $i = 1,\dots,n$. Then, the *stressed version of the h-th alternative model* has states $\mathbf{z}_i, i = 1,\dots,n$, with re-weighted probabilities $p_i^{(h)}$ attached to them.

Specifically, for $h = 1,\dots,b$, we proceed as follows:

1. Generate a multivariate sample of size m from the underlying parametric model, $(\tilde{\mathbf{z}}_1^{(h)},\dots,\tilde{\mathbf{z}}_m^{(h)}) \sim F(\cdot|\theta^0)$, where $\tilde{\mathbf{z}}_k^{(h)} = (\tilde{z}_{1,k}^{(h)},\dots,\tilde{z}_{d,k}^{(h)})$, $k = 1,\dots,m$.
2. Estimate the parameters of the risk distributions $\hat{\theta}^{(h)} \equiv \hat{\theta}(\tilde{\mathbf{z}}_1^{(h)},\dots,\tilde{\mathbf{z}}_m^{(h)})$.
3. Using the estimated parameters, define $p_i^{(h)} = \dfrac{1}{n}\dfrac{f(\mathbf{z}_i|\hat{\theta}^{(h)})}{f(\mathbf{z}_i|\theta^0)}$ for $i = 1,\dots,n$.
4. Set $t^{(h)} = (1+\beta)\sum_{i=1}^n p_i^{(h)}\zeta(y_i),\ \beta > 0$.
5. Solve Problem (I) for the Radon-Nikodym density $w_i^{*(h)}$ by setting $y_i = y_i, p_i = p_i^{(h)}$ and $t = t^{(h)}$ and evaluate the χ^2-divergence $\lambda^{*(h)} = \frac{1}{n}\sum_{i=1}^n p_i^{(h)}(w_i^{*(h)})^2$.
6. Set $\lambda = \lambda^{*(h)}$ in Problem (II) and solve for the Radon-Nikodym density $v_{j,i}^{*(h)}$ by setting $z_{j,i} = z_{j,i}$ for $j = 1,\dots,d, i = 1,\dots,n$ and $p_i = p_i^{(h)}$.
7. Evaluate $R_j^{(h)} = \dfrac{\sum_{i=1}^n p_i^{(h)} w_i^{*(h)} z_{j,i} - \sum_{i=1}^n p_i^{(h)} z_{j,i}}{\sum_{i=1}^n p_i^{(h)} v_{j,i}^{*(h)} z_{j,i} - \sum_{i=1}^n p_i^{(h)} z_{j,i}}$ for $j = 1,\dots,d$.

In step 3, $f(\cdot|\theta)$ denotes the joint density corresponding to $F(\cdot|\theta)$.

3.3 A statistical version of the Solvency II standard formula

The Solvency II Standard Formula does not in itself constitute a well-specified statistical model. In this section we formulate a model that translates the principles of the Standard Formula into statistical language.

3.3.1 Standard formula

The standard formula for the non-life underwriting risk is a tool for the aggregation of premium and reserve, lapse and catastrophe risks. In this paper we are only concerned with the premium and reserve risk and the definition of the standard formula adopted reflects this. Firstly, the premium risk is the risk that the future premiums earned will be inadequate to cover the volume of losses and expenses. On the other hand, the reserve risk takes into account the risk that the claim provisions might be mis-estimated as well as the risk that stems from the claim payments being stochastic in nature. From an actuarial perspective, it is the risk that the estimated reserves will be insufficient to cover claims payments in a full run-off of the liabilities [65].

Under the standard formula, the solvency capital requirement for the non-life premium and reserve risk is defined as

$$\mathrm{SCR}_{\mathrm{SF}} = 3\sigma V,$$

where σ and V are the portfolio's standard deviation and volume measure respectively.

The portfolio volume measure after aggregating across different business lines is given by:

$$V = \sum_i V_i = \sum_i (V_{i,p} + V_{i,r})(0.75 + 0.25D_i).$$

The standard deviation for a line of business i is calculated as

$$\sigma_i = \frac{\sqrt{V_{i,p}^2\sigma_{i,p}^2 + V_{i,p}V_{i,r}\sigma_{i,p}\sigma_{i,r} + V_{i,r}^2\sigma_{i,r}^2}}{V_{i,p} + V_{i,r}} \left[(V_{i,p} + V_{i,r})(0.75 + 0.25D_i)\right],$$

where $\sigma_{i,p}, \sigma_{i,r}$ are the standard deviations for premium and reserve risk and $V_{i,p}, V_{i,r}$ are the corresponding business volumes.

Finally, the portfolio standard deviation σ is then defined as

$$\sigma = \frac{1}{V}\sqrt{\sum_{i,j} \rho_{i,j}\sigma_i\sigma_j V_i V_j}$$

where $\rho_{i,j}$ is the correlation between business line i and j.

Solvency II calculations take into consideration the diversification between geographical regions. The diversification factor, D_i is interpreted as the Herfindahl index which measures the dispersion of the risk. In the sequel we assume the diversification factor $D_i = 1$.

3.3.2 Model structure

In this section we frame the Solvency II standard formula in terms of an underlying statistical model.

Let $\mathbf{Z} = (Z_1, \dots, Z_d)$ be the losses from d lines of business of a non-life insurance portfolio and $Y = g(\mathbf{Z}) = \sum_{i=1}^{d} Z_i$ the total portfolio loss. The Solvency Capital Requirement (SCR) is given by the Value-at-Risk

$$\mathrm{SCR} = F_Y^{-1}(\alpha),$$

for some $\alpha \in (0,1)$, typically $\alpha = 0.995$, where $F_Y^{-1}(\alpha) = \inf\{y \in \mathbb{R} | F_Y(y) \geq \alpha\}$.

The loss in business line i can be written as

$$Z_i = V_{i,p}X_{i,p} + V_{i,r}X_{i,r},$$

where $X_{i,p}, X_{i,r}$ are the losses for premium and reserve risk and $V_{i,p}, V_{i,r}$ are the corresponding business volumes. The random variables $X_{i,p}, X_{i,r}$ have standard deviations $\sigma[X_{i,p}] = \sigma_{i,p}$ and $\sigma[X_{i,r}] = \sigma_{i,r}$ and means $\mathbb{E}[X_{i,p}] = \mu_{i,p}$ and $\mathbb{E}[X_{i,r}] = \mu_{i,r}$. If we assume that $\mu_{i,r} = \mu_{i,p} = 1$, then $\sigma_{i,p}, \sigma_{i,r}$ can be interpreted as coefficients of variation.

In the standard formula approach, a $d \times d$ correlation matrix $\mathbf{R}$ of the business lines $\mathbf{Z}$ is considered. Furthermore, for any business line, the correlation matrix of premium and reserve risk, $(X_{i,p}, X_{i,r})$, is implicitly assumed to be

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

In order to have a well specified model, we need to construct the correlation matrix Σ of the $2d$-dimensional vector of risk factors,

$$\mathbf{X} = (X_{1,p}, X_{1,r}, X_{2,p}, X_{2,r}, \dots, X_{d,p}, X_{d,r}).$$

While one can design a simulation scheme for $\mathbf{X}$ that reproduces the specified matrix $\mathbf{R}$ as the correlation matrix of $\mathbf{Z}$, such a recipe does not admit a mathematically intuitive model specification. Hence, we propose an alternative way of parameterising the dependence structure of **X**, based on the Kronecker product, which diverges somewhat from the approach in the standard formula.

Specifically, we assume that the correlation matrix Σ of **X** is given by the Kronecker product of **A** and **R**, that is,

$$\Sigma := \mathbf{A} \otimes \mathbf{R} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & r_{12} & \cdots & r_{1d} \\ r_{21} & 1 & \cdots & r_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ r_{d1} & r_{d2} & \cdots & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.5 & r_{12} & 0.5r_{12} & \cdots & r_{1d} & 0.5r_{1d} \\ 0.5 & 1 & 0.5r_{12} & r_{12} & \cdots & 0.5r_{1d} & r_{1d} \\ r_{21} & 0.5r_{21} & 1 & 0.5 & \cdots & r_{2d} & 0.5r_{2d} \\ 0.5r_{21} & r_{21} & 0.5 & 1 & \cdots & 0.5r_{2d} & r_{2d} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{d1} & 0.5r_{d1} & r_{d2} & 0.5r_{d2} & \cdots & 1 & 0.5 \\ 0.5r_{d1} & r_{d1} & 0.5r_{d2} & r_{d2} & \cdots & 0.5 & 1 \end{bmatrix}$$

Thus,

$$Corr(X_{i,p}, X_{j,p}) = r_{ij}, \qquad Corr(X_{i,r}, X_{j,r}) = r_{ij},$$
$$Corr(X_{i,p}, X_{i,r}) = 0.5, \qquad Corr(X_{i,p}, X_{j,r}) = 0.5r_{ij}.$$

However, such an approach does not guarantee that the correlation matrix of business line losses **Z** is exactly equal to the specified correlation matrix **R**.

Building up the matrix Σ this way ensures that, by the properties of the Kronecker product, Σ will be positive definite [83]. Furthermore, it results in a multiplicative structure for cross-correlations which have been used in a Gaussian setting to complete correlation matrices, under a conditional independence assumption – for more on the problem of correlation matrix completion, see e.g. Georgescu et al. [41].

3.3.3 Data and assumptions

In this section, we list out our statistical assumptions and the input data used for our model. We assume that the business lines for each premium and reserve risk $X_{i,p}, X_{i,r}$ are log-Normally distributed with means $\mu_{i,p} = \mu_{i,r} = 1$ and pre-defined standard deviations $\sigma_{i,p}, \sigma_{i,r}$, as given by the regulator. To model the dependence between the random variables in **X**, we use a Gaussian copula with correlation parameters given by the matrix Σ as constructed above.

We use $d = 12$ lines of business. The correlation matrix $\mathbf{R}$ is pre-defined by the regulator [e.g., 60] as

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.50 & 0.50 & 0.25 & 0.50 & 0.25 & 0.50 & 0.25 & 0.50 & 0.25 & 0.25 & 0.25 \\ 0.50 & 1.00 & 0.25 & 0.25 & 0.25 & 0.25 & 0.50 & 0.50 & 0.50 & 0.25 & 0.25 & 0.25 \\ 0.50 & 0.25 & 1.00 & 0.25 & 0.25 & 0.25 & 0.25 & 0.50 & 0.50 & 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1.00 & 0.25 & 0.25 & 0.25 & 0.50 & 0.50 & 0.25 & 0.50 & 0.50 \\ 0.50 & 0.25 & 0.25 & 0.25 & 1.00 & 0.50 & 0.50 & 0.25 & 0.50 & 0.50 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.50 & 1.00 & 0.50 & 0.25 & 0.50 & 0.50 & 0.25 & 0.25 \\ 0.50 & 0.50 & 0.25 & 0.25 & 0.50 & 0.50 & 1.00 & 0.25 & 0.50 & 0.50 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0.50 & 0.50 & 0.25 & 0.25 & 0.25 & 1.00 & 0.50 & 0.25 & 0.25 & 0.50 \\ 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 1.00 & 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.50 & 0.50 & 0.50 & 0.25 & 0.25 & 1.00 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.50 & 0.50 & 0.25 & 0.25 & 0.25 & 0.25 & 0.50 & 0.25 & 1.00 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.50 & 0.25 & 0.25 & 0.25 & 0.50 & 0.25 & 0.25 & 0.25 & 1.00 \end{bmatrix}$$

The premium volumes, meant to be indicative of a large diversified portfolio, were taken from the 2019 Solvency and Financial condition report for AIG's Europe operations [2].

Table 3.1 Standard deviations as specified by the regulator and the premium volumes for the premium and reserve risk for each line of business. [sources: 60, 2]

No	Line of Business	Premium risk		Reserve risk	
		$\sigma_{i,p}$	$V_{i,p}$	$\sigma_{i,r}$	$V_{i,r}$
1	Motor vehicle liability	0.080	113.448	0.090	187.276
2	Other motor	0.080	6.326	0.080	7.554
3	Marine, aviation and transport (MAT)	0.150	347.254	0.110	569.077
4	Fire and other damage to property	0.064	772.964	0.100	844.445
5	General liability	0.112	633.813	0.110	2544.256
6	Credit and suretyship	0.190	56.532	0.172	136.130
7	Legal expenses	0.083	0.000	0.055	0.000
8	Assistance	0.064	2.484	0.220	1.590
9	Miscellaneous financial loss	0.130	43.962	0.200	48.695
10	Non- proportional casualty reinsurance	0.170	0.000	0.200	11.543
11	Non-proportional MAT reinsurance	0.170	0.000	0.200	10.724
12	Non-proportional property reinsurance	0.170	0.000	0.200	62.649

3.4 Results

In this section, we use the model from Section 3.3 to investigate the impact of parameter uncertainty on the sensitivity ranking of risk factors, that is, of the aggregate reserve and premium risk for different lines of business. We first define our baseline model consistently with the assumptions of Section 3.3.3. Integrating the bootstrapping techniques from Section 3.2, the reverse sensitivities of risk factors are calculated, for alternative plausible model parameterisations. We call this the *baseline run*.

Subsequently we present additional results under modelling scenarios whereby different distributional assumptions of baseline run are altered. This allows for the identification of model assumption that have the most impact on sensitivities of input factors.

3.4.1 Baseline run

For the baseline run, we consider $d-1=11$ lines of business instead of 12 as the volumes for business line “legal expenses" are 0 for both the premium and reserve risk. We let each of the premium and reserve risk components, $X_{i,p}, X_{i,r},\ i=1,\dots,11$, follow a log-normal distribution with mean 1 and standard deviations as in Section 3.3.3. The dependence between these random variables is modelled by a Gaussian copula, parameterised by the correlation matrices of Section 3.3.3. We calculate sensitivities to risk factors representing loss by line of business, i.e. to the random variables Z_i as specified in Section 3.3.2.

The settings around parameter uncertainty involve the choice of m, which represents sample size on which parameters are estimated, the methodology used to reflect the parameter uncertainty, and the conditions set on the correlation parameters. For the baseline run, we set $m=20$ to generate a preliminary sample and use re-simulation method of Section 3.2.2 to generate alternative models. The bootstrap procedure is used to derive alternative estimates for the parameters of the marginal log-Normal distributions. Furthermore, we assume that the correlation factors are fixed as given by the regulator and not re-estimated by bootstrapping. The impact of modifying these settings is later explored.

We aim to stress the 95%-Value-at-Risk of Y by 10%. Given that quantiles cannot be written as expectations, this leads to the following formulation of Problem (I): we set $\zeta(y_i)=\mathbb{1}_{\{y_i\leq 1.1\mathrm{VaR}_\alpha(Y)\}}$, where $\alpha=95\%$ and $\mathrm{VaR}_\alpha(Y)$ is the empirical quantile from the simulated sample, and fix $t=\alpha$. Using the premium volumes in Section 3.3.3, we simulate $(\mathbf{Z},Y)$ using a Monte Carlo sample of $n=5\times 10^5$.

We note here that it is important to differentiate between the use of VaR for capital requirement calculation and the analysis of sensitivities through stressing the VaR. In the context of capital requirement calculation, an insurance company is required to hold capital reserves equal to the value represented by the 99.5% VaR in order to cover potential losses. We opted to use instead the 95%VaR for stressing the model, such that our analysis is not exclusively focused on the extreme tails, but also reflects impacts at lower confidence levels that are of interest for management decisions.

To make the following analysis easier to follow, we summarise in Tables 3.2 and 3.3 the notation and baseline parameter values which are used throughout the section.

Table 3.2 Symbols and their meanings in the baseline model.

Symbol	Description
d	Number of business lines
m	Sample size for parameter estimation
b	Number of bootstrap iterations
n	Sample size for Monte-Carlo simulations
X_{ip}	Random variable representing premium risk component for business line i
X_{ir}	Random variable representing reserve risk component for business line i
Z_i	Random variable representing loss for business line i
Y	Random variable representing the output
α	Significance level for Value-at-Risk calculations
t	Threshold value used in Value-at-Risk calculations
$\mathbf{R}$	Correlation matrix of business lines

Table 3.3 Baseline parameter values

Parameter	Value
d	11
m	20
b	50
n	5×10^5
α	0.95
t	α

Figure 3.1 shows box-plots of the sensitivities of $Z_i, i = 1\ldots,11$, in the baseline run, representing the average level and variability of the reverse sensitivity defined in Section 3.2.1, with respect to bootstrapped parameter values. From the plot, it is evident that Z_5 and Z_8 are the most important risk factors, due to their highest sensitivity measure. Furthermore, Z_5 and Z_8 are also most impacted by parameter uncertainty, as they show the most volatility in their respective sensitivities. A greater spread indicates a higher uncertainty with respect to the estimation of the parameter itself. From Figure 3.1, it is seen that Z_2 and Z_{11} exhibit lowest sensitivity levels. However as the error bounds for the risk factors overlap substantially, the exact ranking for the full set of risk factors is difficult to establish. As a result, the

importance rankings between risk factors may often change when the model is run several times. This also indicates that importance rankings for any particular model should be interpreted cautiously, as they may be driven by parameter uncertainty rather than true underlying effects.

We compare most of our results with the baseline run in the following sections. In cases where this is not feasible, we employ a slightly modified baseline run to make comparisons more meaningful.

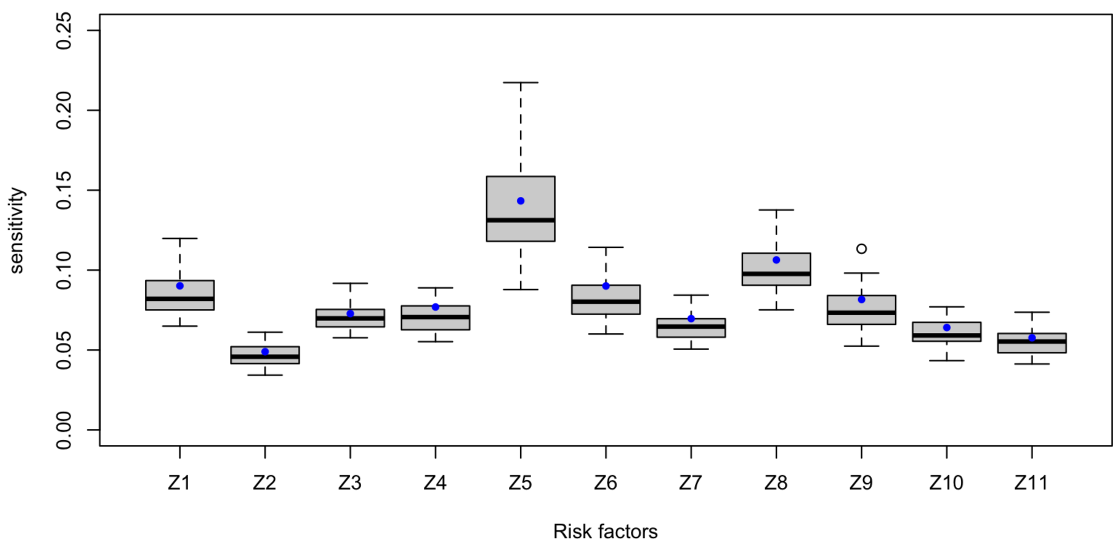


Fig. 3.1 Box plot of sensitivities for risk factors under a χ^2-divergence for $b = 50$ sets of $n = 5 \times 10^5$ simulated scenarios under the baseline run when $\mathrm{VaR}_{95\%}(Y)$ is stressed by 10%; average sensitivities of risk factors are represented by blue dots.

In addition to calculating the reverse sensitivities of risk factors, we also determine the forward sensitivities, refer to Chapter 2 Section 2.3.3. Working with average sensitivities, we show the impact of input factors on the output variables using a radar chart in Figure 3.2. The radar chart is composed of a circular grid with input factors evenly positioned around it. Lines are drawn to connect the data points for each input factor, representing the sensitivities. The length of the lines indicates the magnitude of the sensitivity; longer lines indicate higher importance. In the chart, reverse sensitivities are depicted in blue, while forward sensitivities are depicted in red. We can see that the average forward sensitivities are slightly larger than reverse ones, nonetheless the relative ranking among the input factors remains consistent. For numerical values of the average reverse and forward sensitivities, we refer to Table 3.4, which was used to generate Figure 3.2.

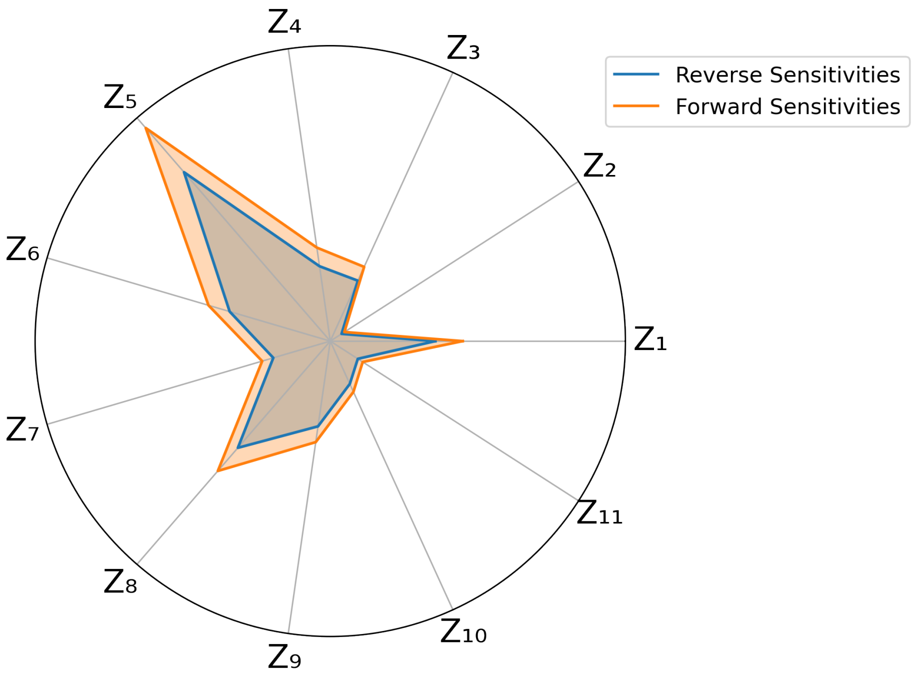


Fig. 3.2 Radar chart of average reverse and forward sensitivities under baseline model for $b = 50$ sets of $n = 5 \times 10^5$ simulated scenarios.

Table 3.4 Table of average reverse and forward sensitivities under the baseline model.

Output	Reverse Sensitivity	Forward Sensitivity
Z_1	0.0901	0.1020
Z_2	0.0489	0.0504
Z_3	0.0728	0.0796
Z_4	0.0769	0.0853
Z_5	0.1433	0.1696
Z_6	0.0900	0.1000
Z_7	0.0696	0.0748
Z_8	0.1063	0.1201
Z_9	0.0816	0.0887
Z_{10}	0.0641	0.0680
Z_{11}	0.0576	0.0602

3.4.2 Variations in type of stress

In this section, we present the results of sensitivities of risk factors when a different type of stress on Y is applied and compare it with the results obtained for the baseline run. In the baseline run, we stressed the 95% VaR of Y by 10%. We now choose to stress the mean of Y

by 10% keeping all other assumptions of the baseline run fixed. This allows us to examine the impact of different types of stresses.

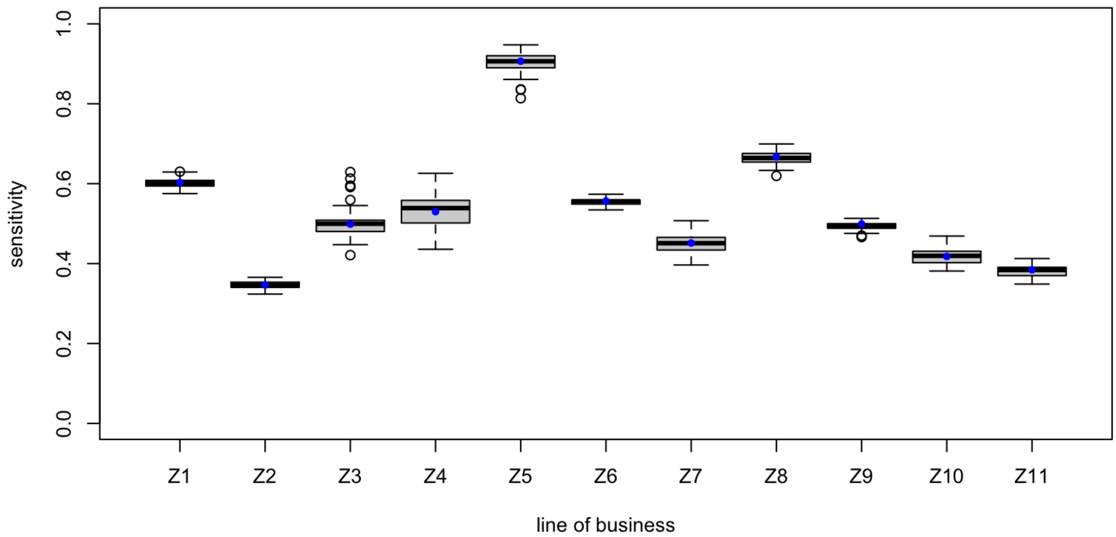


Fig. 3.3 Box plot of sensitivities for risk factors under a χ^2-divergence for $b = 50$ sets of $n = 5 \times 10^5$ simulated scenarios when $\mathbb{E}[Y]$ is stressed by 10%; average sensitivities of risk factors are represented by blue dots.

In Figure 3.3, we plot the sensitivities of $Z_i, i = 1, \dots, 11$ for $b = 50$ bootstrap iterations when the mean of Y is stressed at 10%. The sensitivities of risk factors in this scenario are much higher in absolute terms in Figure 3.3 than what was observed in Figure 3.1. This implies that the sensitivities are greatly impacted by the type of stress that is applied.

However, if we consider the importance rankings, we find results consistent with those obtained for the baseline run; Z_5 and Z_8 come across as being the most sensitive risk factors in the model and Z_2 and Z_{11} being the least sensitive. However, the spread of the sensitivities in Figure 3.3 is much smaller as compared to Figure 3.1. As the error bounds overlap less in Figure 3.3, we note that the order of ranking of the risk factors is much more reliable. This indicates that stress tests based on the mean are more robust to parameter uncertainty compared to stress-tests based on the Value-at-Risk. Hence, there is an important trade-off: while stressing the distribution tail is of interest to risk management purposes, it can also lead to less reliable conclusions.

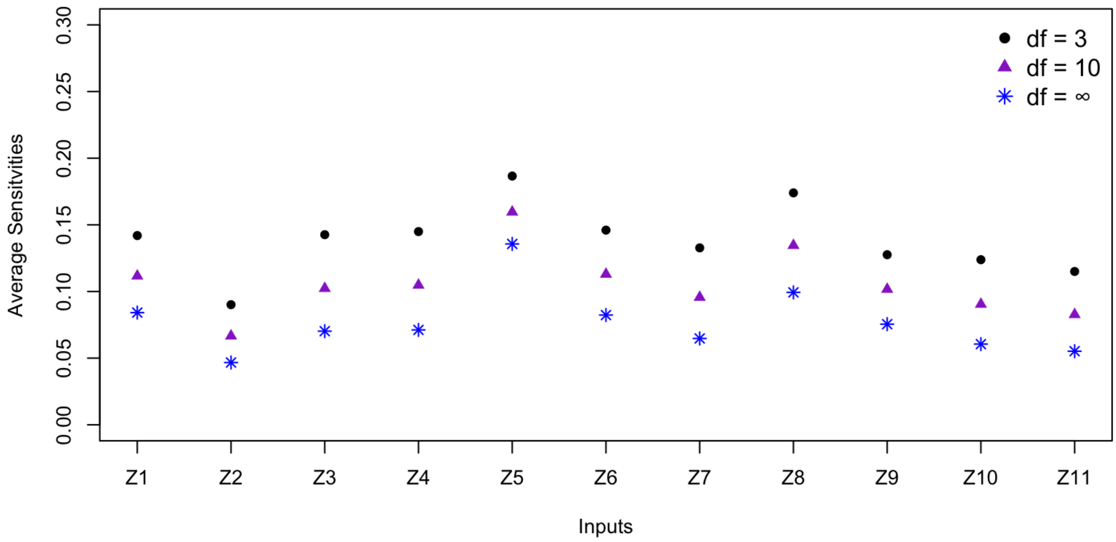
We further substantiate these findings in Table 3.5, where we report the average sensitivities under the two models and accordingly, the order of ranking for the risk factors; 1 being the most important or sensitive risk factor and 11 being the least. We also report the number of times, across bootstrap iterations, that a particular risk factor obtains exactly the average ranking value. This frequency allows us to identify those lines of business which are robust in their rankings. For example, Z_2, Z_5, Z_8, Z_{11} retain their ranking across the two models and also have a high frequency of attaining their average rank.

Table 3.5 Table of average sensitivity, ranking and the occurrence of that particular ranking in the bootstrap procedure for each risk factor.

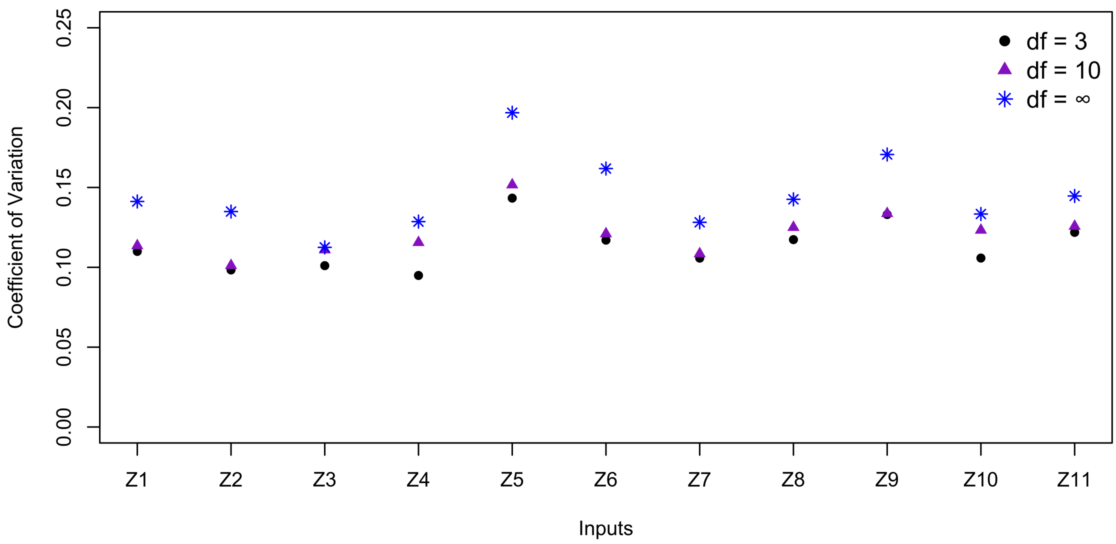
	Stress VaR			Stress Mean		
Input	Avg sensitivity	ranking	frequency	Avg sensitivity	ranking	frequency
Z_1	0.0842	3	0.64	0.6017	3	0.92
Z_2	0.0467	11	1.00	0.3466	11	1.00
Z_3	0.0702	7	0.46	0.5034	6	0.44
Z_4	0.0712	6	0.36	0.5314	5	0.38
Z_5	0.1357	1	1.00	0.9014	1	1.00
Z_6	0.0823	4	0.58	0.5539	4	0.62
Z_7	0.0647	8	0.74	0.4493	8	0.94
Z_8	0.0993	2	1.00	0.6653	2	1.00
Z_9	0.0755	5	0.62	0.4943	7	0.48
Z_{10}	0.0605	9	0.82	0.4192	9	1.00
Z_{11}	0.0552	10	0.94	0.3820	10	1.00

3.4.3 Variations in the copula family

Here, our aim is to understand the effect that the choice of a copula family has on the sensitivities of risk factors. In the baseline run, we implemented a Gaussian copula to model the dependence between risk factors. Here, we use a t-copula and compare our results to the baseline run when all other assumptions are kept fixed. The Gaussian and the t-copula are widely used in risk management frameworks. While the Gaussian copula is mostly considered as the benchmark copula in modelling, the t-copula is popular for its ability to capture the tail dependence [64].



(a)



(b)

Fig. 3.4 Plot of the average sensitivities (b) and coefficient of variation (b) for risk factors' sensitivities when a t-copula is implemented with $3, 10, \infty$ degrees of freedom for $b = 50$ sets of $n = 5 \times 10^5$ simulations under the χ^2-divergence.

Figure 3.4a shows a plot of the average sensitivities of risk factors, Z_i when a t-copula is implemented with $3, 10$ and ∞ degrees of freedom – the last case being the one of a Gaussian copula. The average sensitivities of Z_i are highest when a t-copula with lowest degrees of freedom is implemented and the average sensitivities decrease in absolute terms while retaining their order of ranking between the risk factors. The sensitivities are the lowest when a Gaussian copula is implemented. Hence, in this model, the presence of tail dependence increases the absolute level of sensitivity, while not changing the portfolio's relativities.

Figure 3.4b shows a plot of the coefficient of variation for the risk factor's reverse sensitivity, across bootstrap samples when a t-copula is implemented with $3, 10, \infty$ degrees of freedom. As the coefficient of variation is a scale-invariant quantity, it is useful for comparing the variability among data sets with different means. In Figure 3.4b, we see that the coefficient of variation is highest when a Gaussian copula is used implying a greater variability in the sensitivity values around their means. The coefficient of variation decreases as the number of degrees of freedom is reduced for the t-copula. A possible interpretation of this finding follows by noting that, in this context, we do not fully reflect the uncertainty around correlation values; hence all parameter uncertainty comes from the specification of the marginal distribution parameters. However, at the same time, we know that in t-copula models, the impact of dependence on tail risk is more pronounced compared to Gaussian models.

3.4.4 Variations in the marginal distribution family

Here, we look at the impact on risk factors' sensitivities when a different distribution is used for the marginals. We now assume $Z_i, i = 1, \ldots, 11$ follow an Inverse Gamma distribution with mean and standard deviations consistent with baseline run assumptions. We choose the Inverse Gamma distribution because it is a plausible alternative to the log-Normal distribution with similar properties but a heavier (regularly varying) tail [64, Sec. A.2]. Table 3.6 shows the mean and coefficient of variation of the sensitivities for Z_i, under each marginal distribution assumption. As the Inverse Gamma distribution has a comparatively heavier tail than Log-normal distribution, we see slightly higher average sensitivities for Z_i under Inverse Gamma marginals. We note the same for coefficient of variations.

3.4.5 Uncertainty in the correlation matrix

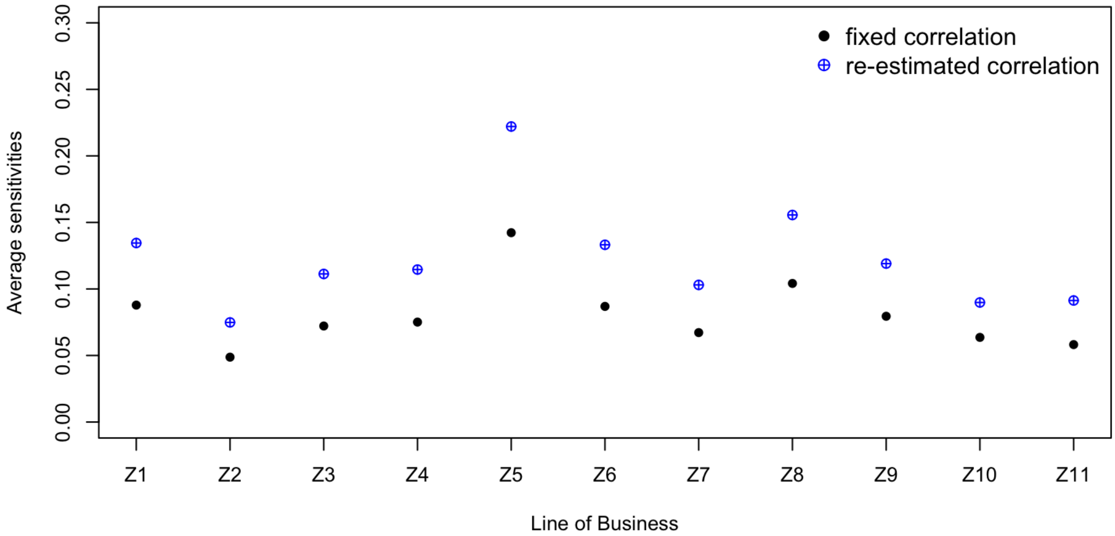
In this section, we aim to understand the effect of correlation matrix, $\mathbf{R}$ on the sensitivities of risk factors. In the baseline run, $\mathbf{R}$ was assumed to be fixed and given by the regulator. Here, we change our assumption on $\mathbf{R}$ and let the correlation parameters be re-estimated in each iteration of the bootstrap procedure. To avoid errors in estimation of correlation factors, we set $m = 100$ in the bootstrap procedure.

Table 3.6 Table of average sensitivity and coefficient of variation of sensitivities for risk factors $Z_i, i = 1, \ldots, d$, when Z_i follow Log-normal versus Inverse-Gamma marginals.

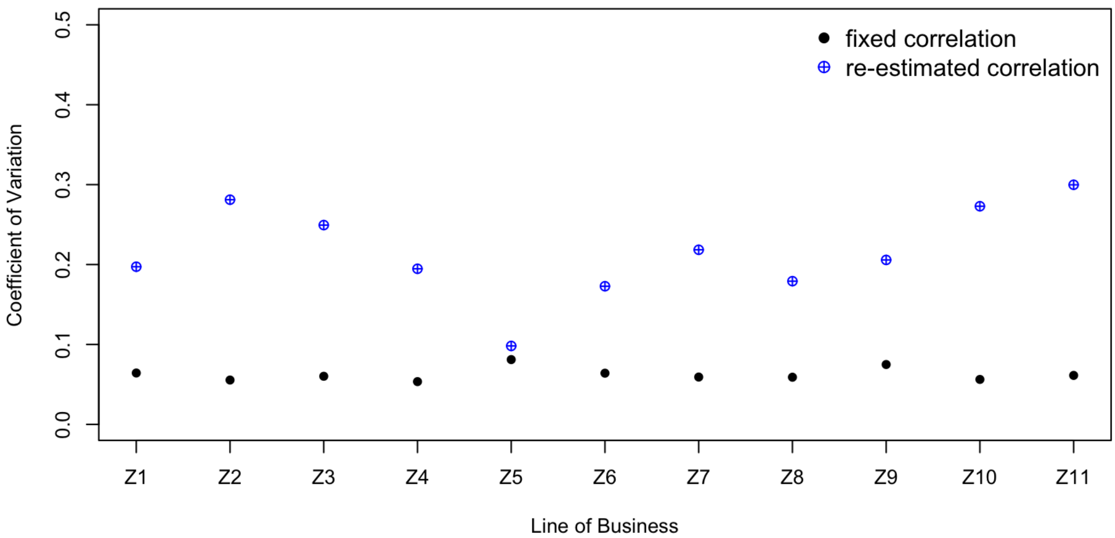
	Log-normal marginals		Inverse-Gamma Marginals	
Input	Avg sensitivity	CoV	Avg sensitivity	CoV
Z_1	0.0842	0.1412	0.0906	0.1467
Z_2	0.0467	0.1349	0.0499	0.1386
Z_3	0.0702	0.1125	0.0754	0.1237
Z_4	0.0712	0.1286	0.0760	0.1455
Z_5	0.1357	0.1970	0.1505	0.2093
Z_6	0.0823	0.1618	0.0904	0.1718
Z_7	0.0648	0.1282	0.0703	0.1417
Z_8	0.0993	0.1426	0.1096	0.1490
Z_9	0.0755	0.1706	0.0833	0.1830
Z_{10}	0.0605	0.1333	0.0655	0.1460
Z_{11}	0.0552	0.1446	0.0602	0.1607

Figure 3.5a and 3.5b show a plot of the average sensitivities of Z_i and the coefficient of variation respectively when the correlation matrix, **R** is kept fixed and when re-estimated. Both the average sensitivities and the coefficient of variation increases substantially when the correlation matrix R is re-estimated, indicating both a higher sensitivity and a lower robustness of the sensitivity rankings observed. This demonstrates the profound impact of uncertainty in the correlation matrix specification on the risk factors' relative importance, as quantified by sensitivity analysis. Interestingly, the higher increase in coefficient of variation is observed for those risk factors (e.g., Z_2, Z_{11}) that had the lowest mean sensitivity in the fixed **R** setting.

Figure 3.6a shows a box-plot of sensitivities of risk factors under the baseline run when $m = 100$ and Figure 3.6b shows a box-plot of sensitivities under the baseline run assumptions but when $m = 100$ and the correlation matrix **R** is re-estimated at every iteration of the bootstrap procedure. From Figure 3.6b, we see that the risk factors Z_5 and Z_8 still take precedence in ranking and have the highest sensitivities. Further, there are stark differences in the spread and variability of the sensitivities between the two figures. The spread of sensitivities is much greater in Figure 3.6b compared to Figure 3.6a. The overlap of the error bounds is also significantly greater in Figure 3.6b. Overall, changes in assumptions on the correlations between the risk factors appears to have a considerable impact on the sensitivities and rankings of the risk factors.

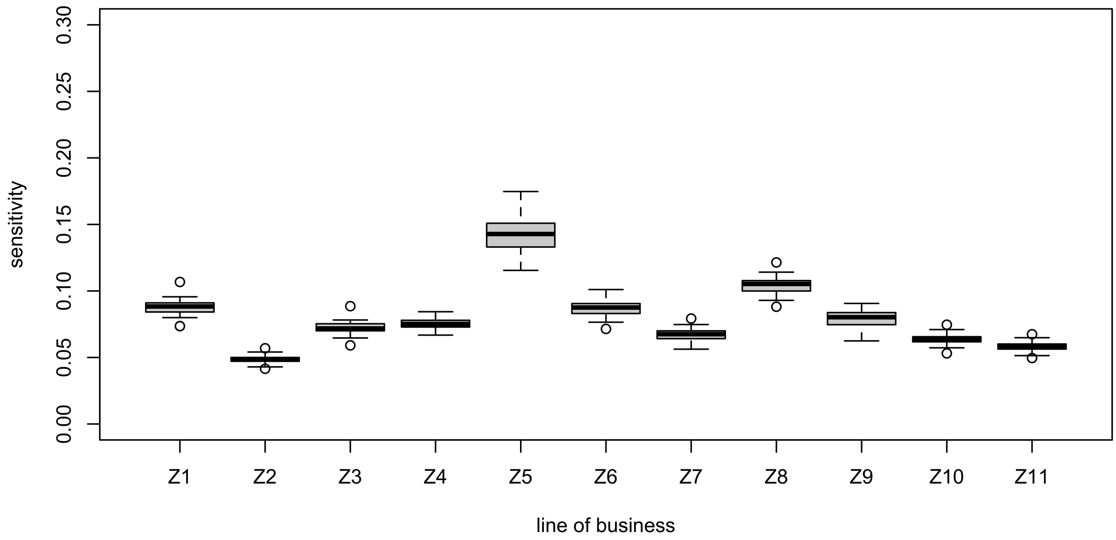


(a)

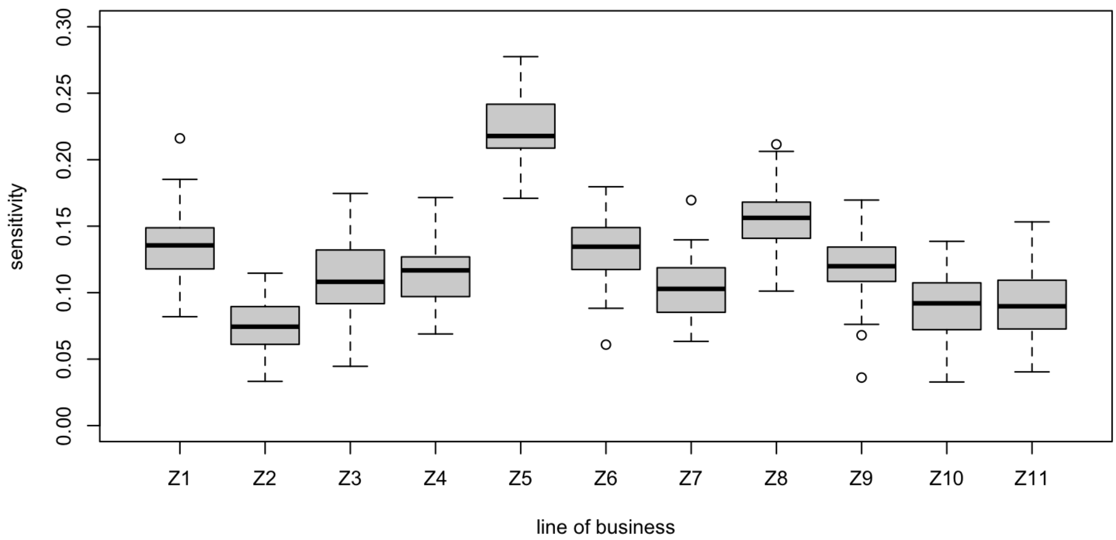


(b)

Fig. 3.5 Plot of (a) average sensitivities and (b) coefficient of variation of risk factors sensitivities for $b = 50$ sets of $n = 5 \times 10^5$ simulated scenarios, when $\mathbf{R}$ is fixed and when $\mathbf{R}$ is re-estimated.



(a)



(b)

Fig. 3.6 Box plot of sensitivities for risk factors under χ^2-divergence for $b = 50$ sets of $n = 5 * 10^5$ simulated scenarios with (a) fixed correlation and (b) re-estimated correlation when $m = 100$.

3.4.6 Variations in the sample size m

In the baseline run, we set $m = 20$, where m is the number of observations used to estimate model parameters. Here, we investigate the impact on parameter uncertainty when the value of m is varied.

As the value of m increases, we expect that estimates become more stable and the ranking of risk factors becomes more robust to parameter uncertainty. To quantify this effect we plot the average pairwise correlation of the risk factor ranks, over bootstrap iterations. Specifically, let $\mathbf{U} = \{u_{i,k}\}_{i=1,\ldots,11,\ k=1,\ldots,b}$ be the matrix of reverse sensitivities of input factors $Z_i, i = 1,\ldots,11$ over $b = 50$ sets of bootstrap iterations. Let $r_{j,k}$ be the pairwise Spearman rank correlation of columns j and k of the matrix $\mathbf{U}$. We then evaluate

$$\bar{r} = \frac{2}{b(b-1)} \sum_{j=1}^{b-1} \sum_{k=j+1}^{b} r_{j,k}.$$

A high value of the average rank correlation $\bar{r}$ indicates that vectors of ranks for the different risk factors are consistent across different bootstrap iterations and hence robust to parameter uncertainty.

In Figure 3.7 we plot $\bar{r}$ against different values of m when (a) the correlation matrix R is kept fixed as in the baseline run and (b) when R is re-estimated in each iteration of the bootstrap procedure. By doing this, we not only investigate how the rank correlations are affected by changes in m, but also examine the impact of uncertainty in R on the robustness of risk factors' importance ranking. For R fixed, we see that the average rank correlation is close to 1 for all values of m. However, when R is re-estimated, the average rank correlation increases very gradually from a level of approximately $\bar{r} = 0.6$ for $m = 50$ and remains consistently below the case of a fixed R. This indicates that parameter uncertainty in the correlation matrix has a profound impact on the robustness of risk factors' importance ranking.

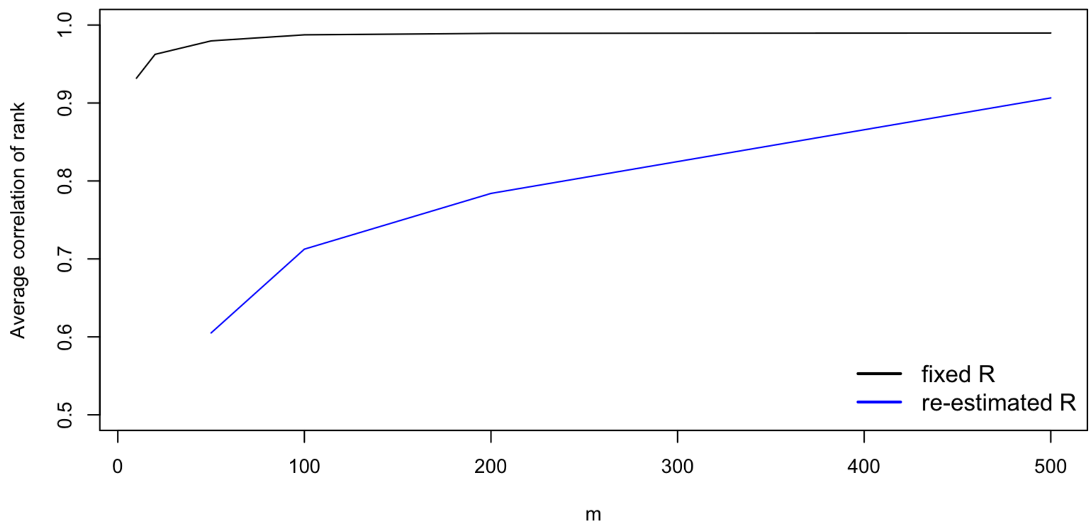


Fig. 3.7 Average correlation of rank (using Spearman's pairwise rank correlation) of risk factors when m is varied for fixed correlation parameters (black) and when correlation parameters are re-estimated (blue).

To illustrate this further, in Figure 3.8a we show a plot of sensitivities of risk factors, Z_i for $b = 50$ sets of bootstrap iterations when $m = 10$ under the baseline run assumptions. Here each line represents the sensitivities of a particular risk factor. Similarly, in Figure 3.8b, we show the sensitivities of Z_i for $b = 50$ bootstrap iterations when $m = 500$, keeping all other assumptions fixed in the baseline run. We see that when $m = 10$, there is high volatility in the sensitivities of the risk factors. Further, the overlapping of the sensitivities across risk factors also makes it difficult to discern the order of importance of risk factors. However, when $m = 500$, there is far less volatility in the sensitivities of all the risk factors and the ranking of risk factors is much more discernible. The order of ranking appears to be much more robust in this case, consistently with the reduction in parameter uncertainty arising from larger sample sizes.

In Figure 3.9, we plot sensitivities of risk factors, Z_i for $b = 50$ sets of bootstrap iterations when $m = 500$ when correlation matrix R is re-estimated. We observe a high volatility in the sensitivities of risk factors and the overlapping of the sensitivities implies that the order of rankings is not very robust even for a large sample of m. Comparing it to Figure 3.8b, it is evident that we lose the stability in the order of rankings when uncertainty in correlation parameters is taken into consideration, making it an important factor in understanding parameter uncertainty.

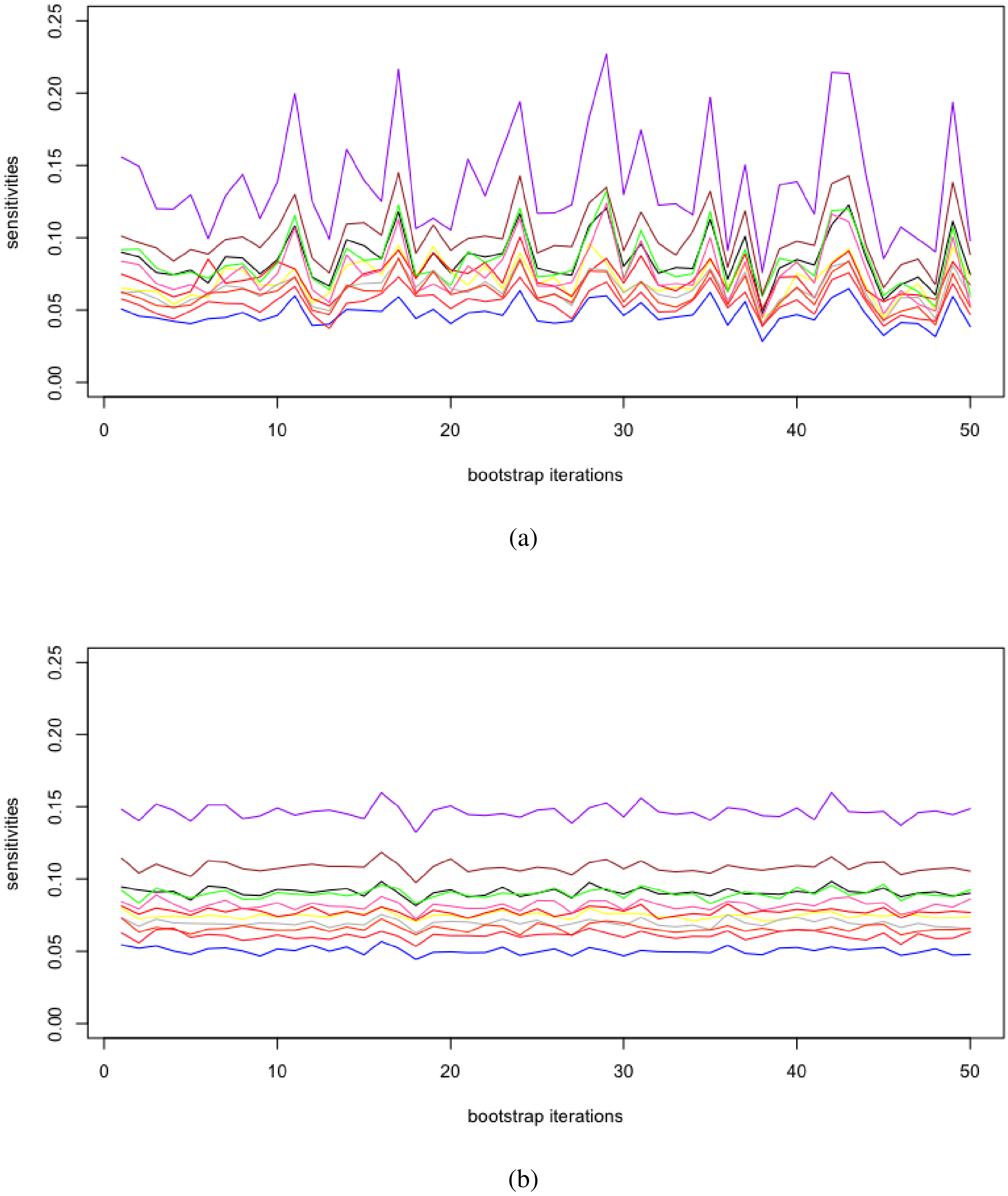


Fig. 3.8 (a) Sensitivities of risk factors for $b = 50$ bootstrap iterations when $m = 10$; (b) Sensitivities of risk factors for $b = 50$ bootstrap iterations when $m = 500$.

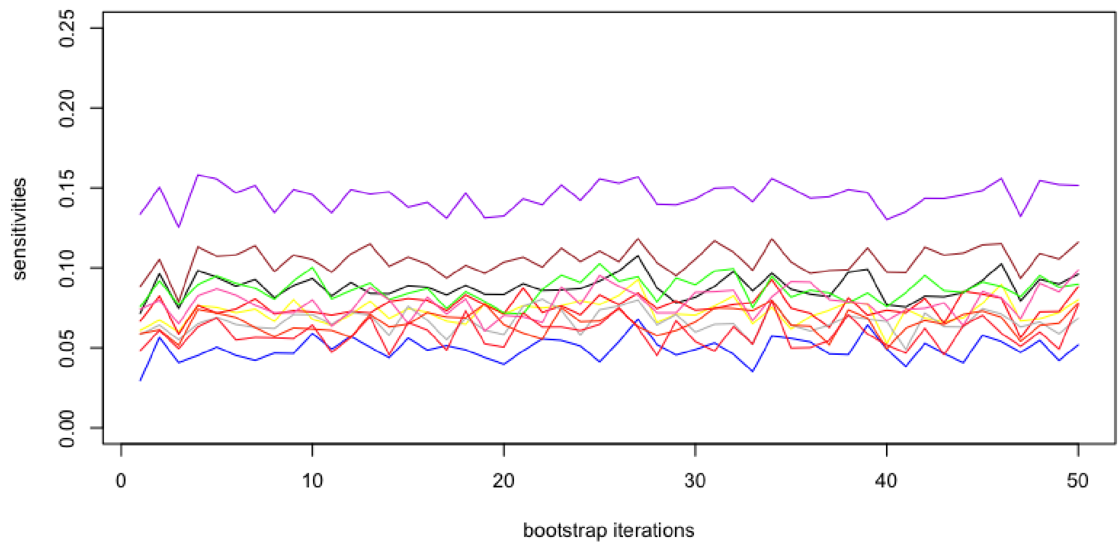


Fig. 3.9 Sensitivities of risk factors for $b = 50$ bootstrap iterations when $m = 500$ when correlation parameters are estimated.

3.4.7 Variations in simulation methodology

In this section, our aim is to examine the effect on parameter uncertainty when the re-weighting approach presented in Section 3.2.2 is implemented, compared to the re-simulation approach that was used in the baseline run. The two methodologies vary in how the alternative models of risk factors, as part of the stress-testing procedure, are generated. To make appropriate comparisons between the two methodologies, we set $m = 100$. The re-weighting methodology involves constructing a likelihood ratio using multivariate normal distributions. Having a smaller sample $m = 10$ as in the baseline run can lead to problems with estimation of the likelihood ratio, leading to unreliable alternate probabilities. This, in turn, would lead to difficulties in assessing the sensitivities of risk factors accurately. For that reason, we choose $m = 100$ to compare our results obtained under the two methodologies.

Figure 3.10 shows a box-plot of sensitivities of risk factors, Z_i when the re-weighting approach is implemented with $m = 100$. Comparing this to Figure 3.6a, which displayed a box-plot of sensitivities of Z_i using re-simulation when $m = 100$, we observe that the two methods give nearly identical results. As the two procedures are mathematically equivalent, it is expected that the change in simulation methodology would not have a significant impact on the resulting sensitivities. Nonetheless the two methods have somewhat different numerical features. In Figure 3.6a, we observe a larger inter-quartile range among the sensitivities across all risk factors, while in Figure 3.10 there are more outliers.

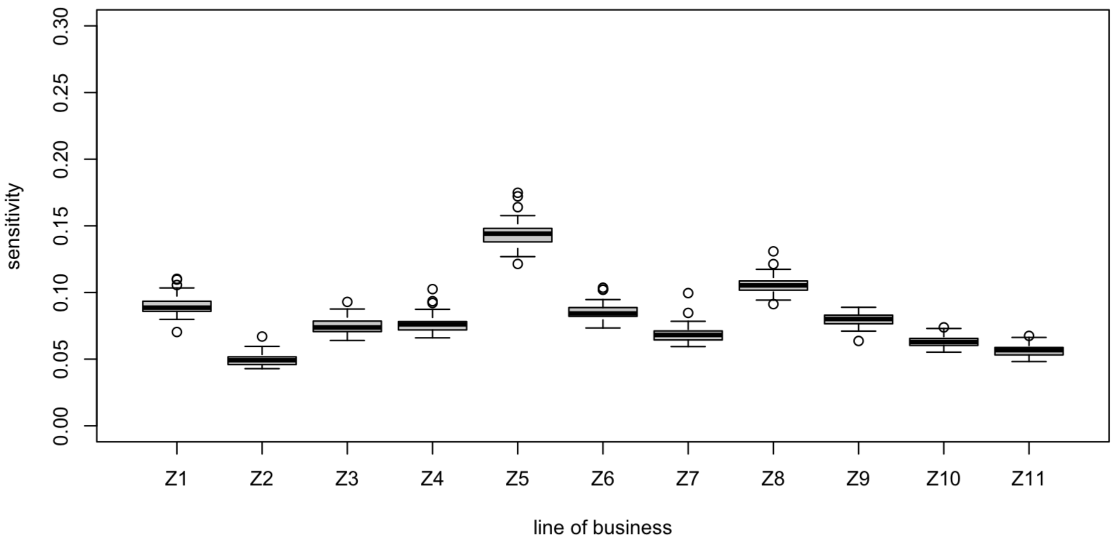


Fig. 3.10 Box plot of sensitivities of risk factors Z_i for $b = 50$ iterations of $n = 5 \times 10^5$ simulations under a re-weighting approach when $m = 100$.

3.5 Conclusion

We have proposed a stochastic model for non-life premium and reserve risk segments of Solvency II calculation which is consistent with the standard formula. Our objective is to investigate the impact of parameter uncertainty on the reverse sensitivity measures of the model's risk factors and, through this, the robustness of their importance ranking. We propose two bootstrapping techniques to simulate risk factors under alternative parametric assumptions and implement reverse and forward sensitivity analyses. We first established our baseline model and conducted sensitivity analyses by individually varying one parameter at a time to assess the impact of parameter changes on uncertainty. Through these analyses, we identified the risk factors that exhibited the highest and lowest sensitivity. The considerable presence of error bounds led to rankings of input factors that overlapped. However, when we stressed the mean by 10% instead of stressing the $\text{VaR}_{95\%}$, we obtained consistent rankings with fewer overlaps. This indicates that the robustness of importance measurement to parameter uncertainty depends on the loss statistic being stressed; in particular an excessive emphasis on the extreme tail may be counter-productive.

To model the dependence between random variables in our model, we initially used the Gaussian copula. As an alternative, we also employed the t-copula, which is comparable to the Gaussian copula but allows for tail dependence. We observed that increasing the degrees of freedom – hence reducing the level of tail dependence – led to a decrease in the average sensitivities, indicating a reduced impact of parameter variations on the portfolio loss. Further, we explored the impact of using a heavier tailed marginal distribution such as the Inverse Gamma. As expected, the sensitivities were slightly higher than the baseline model but maintained similar order of rankings.

Additionally, we investigated the effect of modifying our assumptions regarding the correlation matrix. In the baseline model, the correlation matrix of business lines remained constant across various model runs. However, by re-estimating the correlation matrix during each iteration of bootstrap procedure, we observe a substantial increase in the overlapping of error bounds. This suggests that uncertainty in the dependence structure has a dominant impact on the sensitivity of risk factors.

The sample size used in parameter estimation was set to 20 in the baseline model. By increasing the sample size, we observe greater stability in the rankings of input factors. However in the case where the correlation matrix was re-estimated, even for larger sample sizes, substantial uncertainty persists. This again highlights the lack of robustness to dependence uncertainty.

Lastly, we implemented different simulation methodologies. While the different methods, as expected, give consistent results, they appear to have somewhat different sampling properties. This is a numerical issue of practical importance which merits further investigation.

While changing model assumptions tends to have an impact on the values of the estimated sensitivities, this generally does not affect the relative importance ranking of risk factors, with the exception of the case where the correlation matrix is re-estimated. The study could have been further strengthened by exploring other conditions under which the rankings of input factors lack robustness. Overall we conclude that ignoring parameter uncertainty may not just limit an insurance company's understanding of its solvency capital requirements, but also on the reliability of the lines' relative performance and risk.

Chapter 4

Characterising minimal divergence stresses for general f-divergences

Abstract

We aim to provide a broader framework to risk analysts for deriving stressed probability measures using divergence minimisation problems. While the Kullback-Leibler and χ^2-divergence are popular choices, we present a more flexible approach that allows practitioners to choose from a wider class of f-divergences. In this chapter, we explore the possibility of accommodating a general f-divergence in the minimisation problem proposed in Makam et al. [62], where a χ^2-divergence was minimised subject to a moment constraint. Various results are presented in this chapter that classify f-divergences and provide a characterisation of the shape of the Radon-Nikodym derivative based on their first derivative value at zero. Additionally, we also present results to stochastically compare the stressed distributions of a random variable based on the behaviour of the Radon-Nikodym derivatives when one parameter in the optimisation problem has been altered.

4.1 Introduction and review of literature

In the field of Information theory, f-divergences have been extensively studied as a tool to quantify the distance between two probability distributions [1]. Among the various f-divergences available in the literature, the Kullback-Leibler divergence has been most commonly used by academics and modelers. Additionally, in optimisation problems where Kullback-Leibler divergence is minimised subject to a moment constraint, the corresponding Radon-Nikodym derivative has a tractable expression [55]. In Chapter 2, we highlighted some of the shortcomings of using Kullback-Leibler divergence and explored the idea of using χ^2-divergence as an alternative to address the issue of using Kullback-Leibler divergence in the context of model behaviour when heavy-tailed distributions are used. This motivates the study of a wider range of divergences, as applied to stress testing problems.

Practitioners may find it beneficial to choose from various f-divergences based on the objectives of their analysis. Applying multiple f-divergences in the analysis may result in a more comprehensive understanding of model behaviour. In this chapter, we aim to establish a broader framework for divergence minimisation problems where flexibility of selecting from different f-divergences is available to derive stressed probability distributions. In such a context, our intent is to study the characteristics of divergences and the solutions they produce to systematise them in a way to enable several comparisons.

Optimisation problems are often used in the field of actuarial science to assess and quantify model error and other stochastically driven uncertainties in the model. Csiszár [30] in his seminal paper had first studied minimisation divergence problem using the Kullback-Leibler divergence under linear constraints, giving rise to many practical applications in finance and information theory. Breuer and Csiszár [23] have developed an approach based on Csiszár [30] that is particularly relevant to stress testing. It involves searching for the worst case scenario against a reference distribution by measuring the plausibility of a scenario using Kullback-Leibler divergence. Pesenti et al. [67] and Makam et al. [62] have used optimisation problems in the context of stress testing and sensitivity analysis in risk management processes. Cambou and Filipović [24] solve a similar divergence minimisation problem for f-divergences, focusing the problem around internal models in the context of solvency assessment. Unlike Breuer and Csiszár [23], Pesenti et al. [67] and Makam et al. [62] where moment constraints were used while minimising the divergence, Cambou and Filipović [24] solve with respect to constraints on scenarios and show that an infinite-dimensional problem can be reduced to a finite-dimensional convex optimisation problem.

The applicability of minimising f-divergences also extends into the realm of finance. Pricing models are often considered in an incomplete market setting where selecting a price measure is not direct and trivial [10]. Minimal divergence martingale measures, defined to be the measures obtained after minimising an f-divergence over the set of all martingale measures, have been heavily relied on to derive a set of prices in the context of no-arbitrage argument [54]. Delbaen et al. [31] showed that maximising the utility function embedding preferences of investors is equivalent to minimising a pseudo-distance between probability measures. f-divergences such as relative entropy and Hellinger distance defined over all equivalent martingale measures are interpreted as pseudo-distances in optimisation problems. f-divergences arise naturally here as most often, options are represented by convex functions of the underlying asset. The conditions for which the convex order between equivalent martingale measures hold becomes relevant and this has been explored by [13]. Further, Goll and Rüschendorf [46] give a general characterisation of minimal distance martingale measure by providing necessary and sufficient conditions for all f-divergences which are strictly convex and differentiable. Generalisations of Csiszár [30]'s theorem have also been used in the pricing of options (see Goll and Rüschendorf [46]).

In Chapter 2, we have used a divergence minimisation problem where χ^2-divergence is minimised subject to a constraint on the mean of a random variable. Here, we generalise the problem such that a general f-divergence is minimised subject to a moment constraint. In addition, we also generalise the dual problem stated in Chapter 2 to accommodate a general f-divergence. Further, we present some results which allows us to categorise divergences based on the shape of the Radon-Nikodym derivative. By computing the derivative of any divergence at zero, we can determine a key property of the Radon-Nikodym derivative. The solution admits strictly positive weights when the derivative at zero is infinite; otherwise, the solution may also allow for null weights. Additionally, we present results concerning stochastic comparisons when one parameter in the optimisation problem is modified. We prove that a higher stress-induced probability measure stochastically dominates one that is induced by a lower stress. Furthermore, when solving the optimisation problem using different divergences for a given stress, we observe a convex ordering of distribution functions of a random variable.

The rest of the chapter is organised as follows: In Section 4.2, we discuss some preliminaries and define the optimisation problem which minimises a general f-divergence measure subject to a constraint on the mean. In Section 4.3, we present some of our main results that allows us to categorise f-divergences based on their derivative value at zero. In Section 4.4, we present some stochastic relations between distributions under different probability measures when one factor/parameter is altered and demonstrate the relations using a simple example. In Section 4.5, we define reverse and forward sensitivity measures which can be applied in a model irrespective of the divergence used. We use the insurance portfolio example introduced in Chapter 2 to demonstrate the impact of different divergences on the prioritisation of input factors. In Section 4.6, we state some conclusions.

4.2 Preliminaries and problem statement

Let $\mathscr{P}$ be the set of probability measures defined on $(\Omega, \mathscr{A})$. We consider two arbitrary measures P and $Q \ll P$ such that $\frac{dQ}{dP}$ is the Radon-Nikodym derivative of Q with respect to P. Here, we denote $Q \ll P$ to indicate absolute continuity of Q with respect to P. Further, we denote by $\mathbb{E}$ and $\mathbb{E}^Q$, the expectation operator under P and under Q, respectively.

The general definition of an f-divergence given in Liese and Vajda [58] is as follows:

Definition 4.2.1. Let $f : (0, \infty) \to \mathbb{R}$ be a convex function. The f-divergence of Q with respect to P, denoted by $D_f(Q||P)$, is defined as

$$D_f(Q||P) = \int_\Omega f\left(\frac{dQ}{dP}\right) dP = \mathbb{E}\left[f\left(\frac{dQ}{dP}\right)\right],$$

where $f(0) = \lim_{u \to 0} f(u)$ is assumed to be finite when $Q \ll P$. When Q is not absolutely continuous with P, then $D_f(Q||P) = +\infty$.

We restrict our choices to strictly convex f-divergences in this paper. Given that $f(1) = 0$, the f-divergence measure is non-negative, and $D_f(Q||P) = 0$ if and only if $Q = P$.

Moreover, adding a linear term to the function f or multiplying a positive constant does not change the ranking between different probability measures. For further reading on the properties of f-divergences, refer to Ali and Silvey [3], Liese and Vajda [58], Bellini [9], Shannon [80].

In this paper, we consider a model with $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_d)$ inputs on $(\Omega, \mathscr{A})$ such that an aggregation function, g, is used to associate the model inputs to a one-dimensional model output, Y. The aggregation function $g : \mathbb{R}^n \to \mathbb{R}$ when applied on the inputs gives $Y = g(\mathbf{Z})$. For a given $P, Q \in \mathscr{P}$, we can refer to the models $(\mathbf{Z}, g, P)$ and $(\mathbf{Z}, g, Q)$ as the *baseline model* and the *stressed model*. We interpret P to represent the best estimates of the modeler while Q represents an alternative view that stresses the model. We assume that $Q \ll P$, so that the stressed model must agree on zero probability events in the baseline model, and we formulate the change of measure via the Radon-Nikodym derivative $W = \dfrac{dQ}{dP}$. Here, Q is chosen such that the expectation of a random variable under Q is $\mathbb{E}^Q[X] = \mathbb{E}[WX] = t$, where $t \in \mathbb{R}$ is the stressed mean of X. We are essentially generating distortions to the baseline model by choosing an appropriate Q. Here, X is considered to be a generic random variable and depending on the purpose of the analysis, X can be set as either one of the inputs (Z_i), the output (Y) or a function of them.

Our aim is to minimise the dissimilarity between the stressed probability Q and the baseline probability P, that is $D_f(Q||P)$, subject to mean constraint stated above, so that the stressed model is close to the baseline model. For a given t, we then solve the optimisation problem

$$\begin{cases} \min_Q \mathbb{E}\left[f\left(\dfrac{dQ}{dP}\right)\right] \quad \text{st} \\ \mathbb{E}^Q[X] = t. \end{cases} \tag{I}$$

The divergence minimisation under moment constraints as defined in Problem (I) has been studied by multiple authors. Csiszár [29] and Breuer and Csiszár [23] have extensively used the Kullback-Leibler divergence in a minimisation problem within a model uncertainty perspective in a continuous space. Cambou and Filipović [24] studied the divergence minimisation problem using f-divergences on a finite probability space. Other papers by Ben-Tal et al. [11], Weber [84] have also worked on entropy minimisation problem under conditions on Expected Shortfall and average Value-at-Risk. Pesenti et al. [67] have used similar formulations of Problem (I) under VaR and ES constraints. In Makam et al. [62], the χ^2-divergence was minimised subject to a constraint on the mean of a random variable in a discrete space setting. In this paper, we build on the work from Makam et al. [62] to provide a more thorough characterisation to the solution of Problem (I) and its properties when different f-divergences are considered in a discrete setting.

4.3 Discrete setting

We consider models in a discrete space setting where realisations may be generated using Monte-Carlo simulation. In such a context, let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ with σ-algebra $\mathscr{A} = 2^{\Omega}$. Let $p_i = P(\omega_i) > 0$ and $q_i = Q(\omega_i)$ denote the probability of obtaining the state of the world $\omega_i \in \Omega$, under those two measures P and Q. In this discrete setting, we let $W = \dfrac{dQ}{dP}$ such that $W(\omega_i) = \dfrac{dQ}{dP}(\omega_i) = \dfrac{q_i}{p_i} = w_i$. Then the general expression for an f-divergence is given by

$$D_f(Q||P) = \sum_i f\left(\frac{q_i}{p_i}\right) p_i = \sum_i f(w_i)\, p_i.$$

The f-divergences which are under our purview are assumed to be continuously differentiable on $(0, \infty)$.

Some examples are shown in Table 4.1. Note that all divergences in Table 4.1 are chosen so that $f(1) = 0$. Further, the χ^2- and the Kullback-Leibler (KL) divergences essentially correspond to the α-divergence for $\alpha = 2$, respectively $\alpha = 1$.

Table 4.1 List of f-divergences along with their corresponding functions.

f-divergence	Notation	$f(u)$
χ^2	$D_{\chi^2}(Q\|\|P) = \sum_i \frac{q_i}{p_i}^2 - 1$	$u^2 - 1$
KL	$D_{KL}(Q\|\|P) = \sum_i q_i \log \frac{q_i}{p_i}$	$u \ln u$
Hellinger	$D_H(Q\|\|P) = \sum_i (\sqrt{q_i} - \sqrt{p_i})^2$	$(\sqrt{u} - 1)^2$
Triangular discrimination	$D_\Delta(Q\|\|P) = \sum_i \frac{(q_i - p_i)^2}{q_i + p_i}$	$\frac{(u-1)^2}{u+1}$
α	$D_\alpha(Q\|\|P) = \frac{\sum_i p_i \left(\frac{q_i}{p_i}\right)^\alpha - 1}{\alpha(\alpha - 1)}$	$\frac{u^\alpha - \alpha(u-1) - 1}{\alpha(\alpha-1)}$, $\alpha \in \mathbb{R}, \alpha \neq 0, \alpha \neq 1$
Le Cam	$D_{LC}(Q\|\|P) = \frac{1}{2} \sum_i \frac{p_i(p_i - q_i)}{p_i + q_i}$	$\frac{1-u}{2u+2}$

Under the current notation, Problem (I) can be written as

$$\begin{cases} \min_{\mathbf{w}} \sum_{i=1}^n p_i f(w_i) \quad \text{st} \\ \sum_{i=1}^n p_i w_i = 1, \\ \sum_{i=1}^n p_i w_i x_i = t, \\ w_i \geq 0 \quad \text{for all} \quad i = 1, \dots, n. \end{cases} \tag{II}$$

Here, we denote by $\mathbf{w} = (w_1, \dots, w_n)$, with $w_i = W(\omega_i)$, the vector of Radon-Nikodym derivative values, such that $Q(\omega_i) = q_i = p_i w_i$. We will refer to w_i as weights since they allow to re-scale the probabilities p_i to their new values q_i. Furthermore, let $X(\omega_i) = x_i$ and denote $\mathbf{x} = (x_1, \dots, x_n)$. We assume without loss of generality $x_1 < \dots < x_n$. We refer to Chapter 2 for details on how to handle ties in $\mathbf{x}$. For a special case of a Monte-Carlo setting, $p_i = 1/n$ and the states of world have an equal probability of occurrence.

Problem (II) admits a unique solution for all $x_1 \leq t \leq x_n$ since the objective function is strictly convex, continuous and the feasible region is a non-empty, compact, convex subset of $\mathbb{R}^n$.

When $t = \bar{x} = \sum p_i x_i$, the solution of Problem (II) is constant (i.e., $w_i = 1$ for all i so that the solution is $Q = P$) and, when $t = x_n$, the solution is $w_i = 0$ for $i < n$ and $w_n = n$.

The Karush-Kuhn-Tucker (KKT) conditions for $i = 1, \dots, n$ of Problem (II) with $\bar{x} < t < x_n$, are necessary and sufficient for optimality of the solution and are as follows: $\mathbf{w}$ solves Problem (II) if and only if, there exist multipliers λ_1, λ_2 and μ_i for $i = 1, \dots, n$ such that for $i = 1, \dots, n$

$$\begin{aligned} p_i f'(w_i) = p_i \lambda_1 + p_i \lambda_2 x_i + \mu_i, &\quad \sum_{i=1}^{n} p_i w_i = 1, \\ w_i \mu_i = 0, &\quad \sum_{i=1}^{n} p_i w_i x_i = t, \\ \mu_i \geq 0, &\quad w_i \geq 0. \end{aligned} \tag{KKT II}$$

Note that these conditions hold for all i such that $w_i > 0$ or for $w_i = 0$ when $f(0)$ is finite. When $f'(0) = -\infty$, the solution is strictly positive and (KKT II) still hold, see Proposition 4.3.1. This finiteness of $f'(0)$ will play a role in the next results.

In Breuer and Csiszár [23], the minimisation of KL divergence resulted in a solution that was exponential, leading to strictly positive values, whereas, in Makam et al. [62], the minimisation of the χ^2-divergence under constraints on the stressed mean resulted in a solution where null weights could be admitted for large t. We aim to study conditions under which we may get either of the two classes of solutions for any f-divergence used. A solution containing null weights implies a stressed model which gives zero probability to events that the baseline deemed possible. This is typically the case when there is a substantial stress on the mean of X. Such a type of stress may be undesirable for some modellers. Further, note that any calculation/analysis performed in such a stressed model will ignore the zero-weighted states.

In Proposition 4.3.1, we show that the derivative of the f-divergence at zero determines whether the solution is strictly positive for every meaningful value of t. This result is not surprising as the slope of the f-divergence at zero must be finite if a solution with null weights must be obtained. The converse of Proposition 4.3.1 will be proved later in Proposition 4.3.3.

Proposition 4.3.1. Assume that $f'(0+) = -\infty$. Then for all $\bar{x} \leq t < x_n$, the solution $\mathbf{w}^*$ of Problem (II) is strictly positive.

For the divergences listed in Table 4.1, we can provide a full characterisation in terms of their derivative value at zero. Those divergences for which $f'(0) = -\infty$ are as follows: The KL with derivative $f'(u) = 1 + \ln u$ and Hellinger with derivative $f'(u) = 1 - \dfrac{1}{\sqrt{u}}$ result in solutions which are always strictly positive as their derivative at zero is $-\infty$. On the other hand, divergences such as χ^2, Triangular discrimination and Le Cam with respective derivatives $f'(u) = 2u$, $f'(u) = \dfrac{(u-1)(u+3)}{(u+1)^2}$ and $f'(u) = \dfrac{-1}{(x+1)^2}$ have a finite derivative value at zero and consequently may admit zero weights depending on the stress applied. Note that for the α-divergence, the derivative at zero depends on the value of α as $f'(u) = \dfrac{u^{\alpha-1}-1}{\alpha-1}$. More specifically, the derivative is infinite for $\alpha < 1$ and finite for $\alpha > 1$.

In this paper, we only discuss results based on upward stresses on the mean of a random variable regardless of our choice of divergence used in Problem (II). In Proposition 4.3.2, we show that for an upward stress in the mean of X, the solution is always non-decreasing. This implies that a heavier weight is placed on larger realisations of X, in order to be able to achieve such upward stress. In the case of divergences for which $f'(0) > -\infty$, this may result in sacrificing some states by setting their weights to 0 in order for the constraint to be satisfied. Similar analyses can be done for a downward stress of the mean of a random variable. The optimisation problem can be formulated accordingly using a substitution detailed in Makam et al. [62].

Proposition 4.3.2. Let $\bar{x} \leq t \leq x_n$. Then, the solution $\mathbf{w}^*$ of optimisation Problem (II) is non-decreasing in x_i.

From Proposition 4.3.2 and using Proposition A.1 of Pesenti et al. [67], we can infer the stochastic dominance of the distribution of X under measure Q corresponding to the solution $\mathbf{w}^*$ of Problem (II) with respect to its baseline distribution. A further result on stochastic dominance is stated in Proposition 4.4.1 where we discussed the ranking of the solutions for different stresses on the mean of X.

The next result provides a converse to Proposition 4.3.1.

Proposition 4.3.3. If for all $t \in (\bar{x}, x_n)$ Problem (II) has a strictly positive solution $\mathbf{w}^*$, then $f'(0) = -\infty$.

Consider the case where all the weights are positive i.e., suppose $w_1 > 0$. From the KKT conditions, we obtain $\mu_i = 0$ for all i, implying $f'(w_i) = \lambda_1 + \lambda_2 x_i$. More generally, a semi-explicit solution of Problem (II) holds in terms of the inverse of the divergence function's first derivative.

Lemma 4.3.1. Let $\bar{x} < t < x_n$. If $\mathbf{w}^*$ is the solution of Problem (II), then for $i = 1, \ldots, n$,

$$\begin{aligned} w_i^* &= g(\max(f'(0), \lambda_1 + \lambda_2 x_i)) \\ &= \begin{cases} 0 & \text{if } \lambda_1 + \lambda_2 x_i \leq f'(0) \\ g(\lambda_1 + \lambda_2 x_i) & \text{if } \lambda_1 + \lambda_2 x_i > f'(0), \end{cases} \end{aligned}$$

where $g = (f')^{-1}$ and λ_1, λ_2 are KKT multipliers.

We also consider an alternative optimisation problem, "dual" of Problem (II), by maximising the expectation of the random variable X, while constraining the f-divergence to a particular upper bound. Therefore, we seek for the worst case scenario in terms of expected loss which is consistent with a given divergence bound.

$$\begin{cases} \max_{\mathbf{v}} \sum_{i=1}^n p_i v_i x_i \quad \text{s.t} \\ \sum_{i=1}^n p_i v_i = 1, \\ \sum_{i=1}^n p_i f(v_i) \leq \theta, \\ v_i \geq 0 \quad \text{for all} \quad i = 1, \ldots, n. \end{cases} \tag{III}$$

Note that Problem (III) is feasible if and only if $\theta \geq f(1)$. In this case, the feasible region is non-empty, compact and convex and Problem (III) admits a solution.

Similar problems analogous to Problem (III) have been studied by Breuer and Csiszár [23] using the KL-divergence and in Makam et al. [62] using the χ^2-divergence. Blanchet et al. [18] also look at similar formulations to Problem (III) within the context of model error in a non-parametric framework. In Proposition 4.3.4 we show that the solutions of Problem (II) and Problem (III) coincide.

Proposition 4.3.4. Let $\bar{x} < t < x_n$. Denote the solution of Problem (II) by $\mathbf{w}^*$ and the optimal value of the objective function by $\theta^* = \sum_{i=1}^n p_i f(w_i^*)$. Then, $\mathbf{w}^*$ solves Problem (III) with $\theta = \theta^*$. Conversely, for a feasible $\theta \geq f(1)$, denote the solution of Problem (III) by $\mathbf{v}^*$ and let $t^* = \sum_i p_i v_i^* x_i$ be the optimal value. Then, $\mathbf{v}^*$ solves Problem (II) with $t = t^*$.

4.4 Stochastic comparisons

Stochastic relations are useful to understand the relative ranking of the inputs and output of a given model. Here, we analyse how the distribution of a random variable X under a stressed probability solution of Problem (II) changes when one of the parameters of the problem is altered.

We introduce some familiar definitions for stochastic order relations that are considered in this paper (see Denuit et al. [32] for a detailed overview). Denote $F_X^{Q'}, F_X^{Q''}$ to be the distribution function of X under measures Q', Q'' respectively. We write $F_X^{Q'} \preceq_{\text{st}} F_X^{Q''}$ if $F_X^{Q''}$ stochastically dominates $F_X^{Q'}$ in first-order, that is if $F_X^{Q''}(x) \leq F_X^{Q'}(x)$ for all $x \in \mathbb{R}$. Alternatively, $F_X^{Q'} \preceq_{\text{st}} F_X^{Q''}$ if and only if $\mathbb{E}^{Q'}[h(X)] \leq \mathbb{E}^{Q''}[h(X)]$ for all increasing functions h such that the expectations exist. First-order stochastic dominance allows us to compare random variables under different probability measures, indicating which probability measure makes a variable stochastically larger.

We say that $F_X^{Q''}$ strictly stochastically dominates $F_X^{Q'}$ in first-order, denoted by $F_X^{Q'} \prec_{\text{st}} F_X^{Q''}$, if $F_X^{Q''}(x) \leq F_X^{Q'}(x)$ and in addition, $F_X^{Q''}(x) < F_X^{Q'}(x)$ for at least one $x \in \mathbb{R}$. Alternatively, $F_X^{Q'} \prec_{\text{st}} F_X^{Q''}$ if and only if $F_X^{Q'} \preceq_{\text{st}} F_X^{Q''}$ and $\mathbb{E}^{Q'}[h(X)] < \mathbb{E}^{Q''}[h(X)]$, for all strictly increasing functions h such that the expectations exist.

Furthermore, we denote by $F_X^{Q'} \preceq_{\text{cx}} F_X^{Q''}$ the convex ordering of $F_X^{Q'}$ and $F_X^{Q''}$, that is if $\mathbb{E}^{Q'}[h(X)] \leq \mathbb{E}^{Q''}[h(X)]$ for all convex functions h such that the expectations exist.

Through convex ordering of a random variable under different probability measures, the volatility of the distribution can be compared. For more on stochastic orders see Kaas et al. [52], Denuit et al. [32], Shaked and Shanthikumar [79].

In Proposition 4.4.1, we first consider the effect of a change in the stressed mean of X, that is, the parameter t in Problem (II). The distribution of X under the resulting measure increases in stochastic dominance with respect to t.

Proposition 4.4.1. Let $\bar{x} < t' < t'' < x_n$. Denote by $\mathbf{w}', \mathbf{w}''$ the solutions of Problem (II) with $t = t'$ and $t = t''$ respectively. Further, let Q' and Q'' be the probability measures induced by $\mathbf{w}', \mathbf{w}''$ respectively. Then, $F_X^{Q'} \prec_{\text{st}} F_X^{Q''}$.

For example, if X is interpreted as a random variable representing losses, Proposition 4.4.1 implies that a stress on the mean of X can induce an overall stress in the entire loss distribution. In particular, losses at the right tail of the distribution will be assigned higher probability of occurrence.

We now investigate the role played in Problem (II) by the divergence function f in determining the shape of the Radon-Nikodym derivative W, and in turn, the stressed probability measure Q, for a given change in the mean of X. In particular, as the new mean of X is constrained, it is interesting to understand under what conditions the distribution of X becomes more variable, that is increases in convex order, after a divergence f is replaced by another function $\tilde{f}$. See Chan et al. [26], Kaas et al. [52], Gupta and Aziz [47], Roch and Valdez [73] for relevant applications of convex ordering in mathematical science, economics and actuarial science.

Proposition 4.4.2. Let $f, \tilde{f}$ be divergence functions, such that $f(u) = u^2 - 1$ leads to the χ^2-divergence and $\tilde{g} = (\tilde{f}')^{-1}$ is strictly convex. Denote by $\mathbf{w}, \tilde{\mathbf{w}}$ the solutions of Problem (II), corresponding to $f, \tilde{f}$ respectively, for a given $t \in (\bar{x}, x_n)$. Further, let Q and $\tilde{Q}$ be the probability measures induced by $\mathbf{w}, \tilde{\mathbf{w}}$ respectively. Then, $F_X^Q \preceq_{\text{cx}} F_X^{\tilde{Q}}$.

Proposition 4.4.2 implies that the stressed distribution of X, derived with a divergence corresponding to a strictly convex $\tilde{g}$, dominates in convex order the stressed distribution of X

under the χ^2 divergence. In the former case, the stressed model induces higher probabilities in both tails of the distribution, resulting in a higher variability, even though the stressed mean is the same.

Example. To demonstrate the results in this section, we consider a random variable X that follows a Log-Normal distribution with mean 150 and standard deviation 35. In Figure 4.1, we depict the effect on the distribution functions when different stresses are used. We solve Problem (II) using the KL-divergence to obtain Radon-Nikodym derivatives. We use a $5\%, 10\%, 15\%$ stress on the mean to induce a change of measure. On the left, we plot the Radon-Nikodym derivatives obtained under each of the stresses and on the right, we plot the corresponding stressed probability functions. We see that a stressed probability function stochastically dominates the baseline function. Furthermore, as the stress t increases the distribution of X is seen to stochastically increase, consistently with Proposition 4.4.1.

In Figure 4.2 we show similar figures to demonstrate stochastic dominance under varying stresses this time using the χ^2-divergence to solve Problem (II).

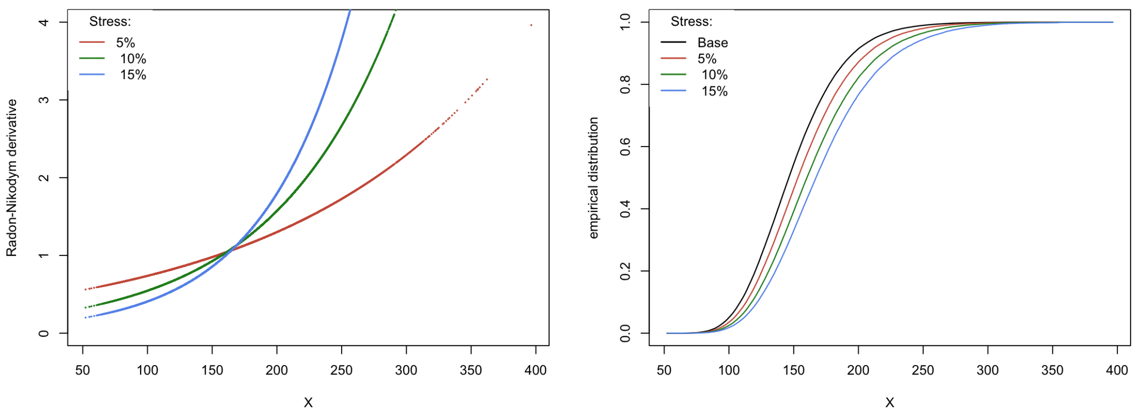


Fig. 4.1 Left: Radon-Nikodym derivatives of stressed models using KL-divergence. Right: Stressed probability distributions of X under different stresses.

To demonstrate the convex order of the distribution functions under different probability measures as in Proposition 4.4.2, we solve Problem (II) using different divergences but for a fixed stressed mean, t. We let $\tilde{f}$ represent KL-divergence. Hence, a comparison between the χ^2- and KL-divergence is shown in Figure 4.3. On the left, we plot the Radon-Nikodym derivatives obtained when Problem (II) is solved while minimising KL-divergence and subsequently solved when χ^2-divergence is minimised for a fixed t corresponding to stressing the mean by 15%. On the right, we plot the empirical distribution functions under the baseline (black) and stressed distribution under χ^2- and KL-divergence (green, red resp).

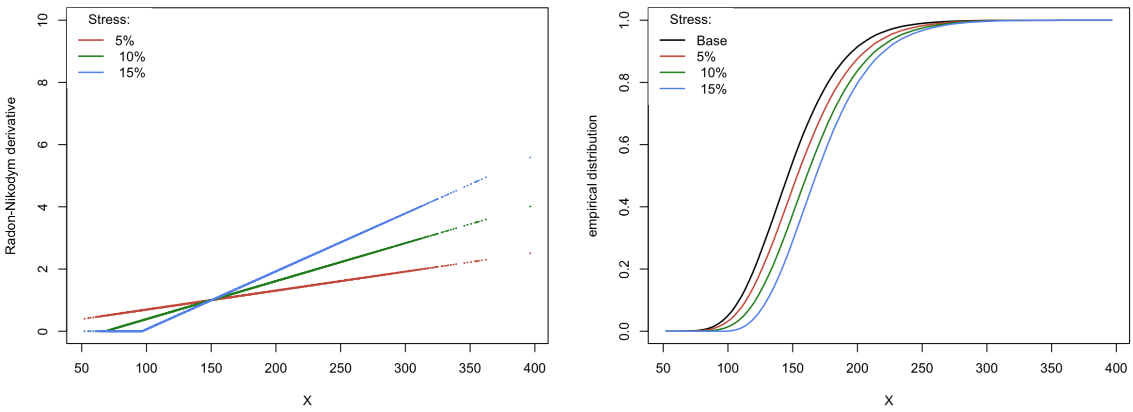


Fig. 4.2 Left: Radon-Nikodym derivatives of stressed models using χ^2-divergence. Right: Stressed probability distributions of X under different stresses.

The Radon-Nikodym derivative obtained using χ^2 and KL-divergence intersect twice, while the distribution functions cross each other once [79]. Therefore by using Proposition 4.4.2, we have $F_X^Q \preceq_{\text{cx}} F_X^{\tilde{Q}}$, where $Q, \tilde{Q}$ are the probability measures induced when χ^2-, KL-divergence is minimised in Problem (II) respectively. Hence, though the mean of the random variable is kept fixed, the KL-divergence produces a more dispersed stressed probability distribution than χ^2-divergence implying that a change of divergence can result in greater fluctuations.

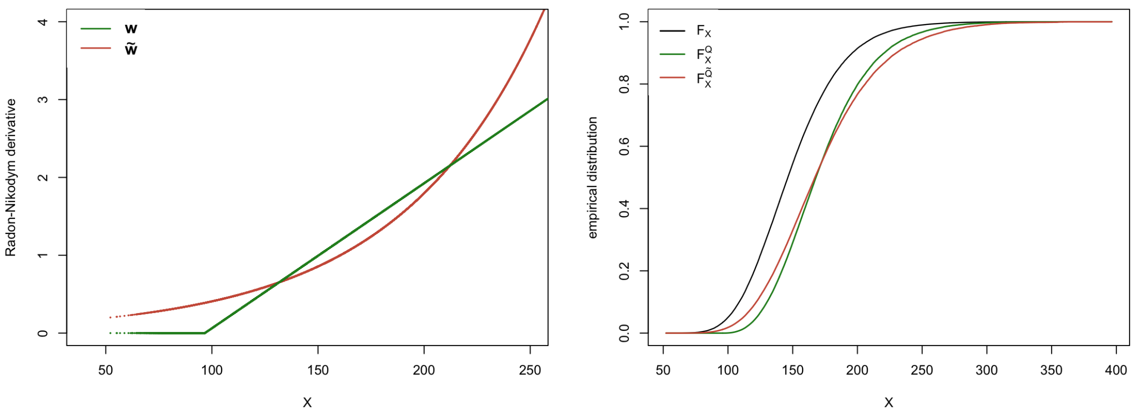


Fig. 4.3 Left: Radon-Nikodym derivatives of stressed models using χ^2- and KL-divergence. Right: Stressed probability distributions of X under χ^2- and KL-divergence; Base probability distribution shown in black.

In Table 4.2, we display the standard deviation, Value-at-Risk and Expected Shortfall at 95% to corroborate our findings. We can see that while the stressed means are equal for both divergences, the standard deviation for the KL-divergence is larger and has a higher impact on the tail.

Table 4.2 Table of statistics of X under stressed probability measures for KL- and χ^2-divergence with respect to baseline probability when t is varied.

	Base	Stress 1 (5%)		Stress 2 (10%)		Stress 3 (15%)	
		KL	χ^2	KL	χ^2	KL	χ^2
Mean	150.06	157.56	157.56	165.06	165.06	172.56	172.56
Sd	35.04	37.76	36.86	40.58	37.10	43.45	36.18
$\text{VaR}_{95\%}$	213.07	225.86	224.19	239.55	232.30	253.02	239.39
$\text{ES}_{95\%}$	235.99	250.99	248.13	266.22	256.68	281.07	263.36

We note that in Chapter 2, the use of χ^2-divergence in quantitative models when heavy-tailed distributions were employed was motivated by the limitations encountered when using KL-divergence. KL-divergence can effectively capture tail behaviour, but due to the convex nature of the resulting Radon-Nikodym derivative, it is particularly susceptible to extreme values. This can result in excessive weights on extreme states, which in turn can lead to numerical instability. Consequently, the use of χ^2-divergence offers a more balanced approach as it is less sensitive to extreme values and can effectively capture tail behaviour without placing excessive emphasis on tails.

4.5 Application to importance measurement

In this section, we aim to study the effect of using different divergence measures on importance rankings of inputs in a model. By understanding the influence of various inputs on the output of a model, we can determine the relative importance of model inputs. This is done by evaluating the change in the output caused as a result of modifying one input factor at a time.

4.5.1 Model specification

We borrow the example from Chapter 2 that was used as case study for an insurance portfolio. Briefly, the insurance portfolio consists of inputs factors Z_1, Z_2, Z_3, Z_4 and the output is denoted by Y. In our model, we take into account losses from two distinct lines of business, Z_1, Z_2. Specifically, the distribution of Z_1 follows a Log-normal distribution with mean 150 and standard deviation 35 while Z_2 follows a Gamma distribution with mean 200 and standard deviation 20. We subject the two losses to a multiplicative (inflation) factor Z_3 which follows a Log-normal distribution with mean 1.05 and standard deviation 0.05. Finally Z_4 represents the percentage of the recovery lost when the reinsurer defaults and follows a Beta distribution with mean 0.1 and standard deviation 0.2. The total portfolio loss is given by $Y = L - (1 - Z_4)\min\{(L-d)_+, l\}$, where $L = (Z_1 + Z_2)Z_3$. The insurance company buys reinsurance with a deductible d and limit l, with chosen values $d = 380, l = 30$. Furthermore, we let Z_1, Z_2, Z_3 to be independent with each other while Z_4 is dependent on L through a Gaussian copula with a correlation of 0.6. We use a Monte Carlo simulation to generate a sample of 10^5 scenarios for $(\mathbf{Z}, Y)$.

4.5.2 Methodology and results

We choose the family of α-divergences [9] listed in Table 4.1 to implement our approach. The following steps are used to implement sensitivity analysis to determine the rankings of input factors based on the evaluation of expected shortfall at 95% confidence level.

We choose a value of α for the divergence and a value for θ. We note here that the choice of θ is at the discretion of the modeller and reflect for example the extent of perceived model uncertainty. Constraining θ, we solve Problem (III) to maximise the mean of each input factor $Z_i, i = 1,\dots,4$ individually. The Radon-Nikodym derivative obtained as a result of maximising the mean of input Z_i subject to a value of θ is denoted by $W^{\theta}_{Z_i,\alpha}, i = 1,\dots,4$. Using the Radon-Nikodym derivatives obtained, we evaluate the expected shortfall of output Y at 95% confidence level. We rank the model inputs based on their impact on expected shortfall of Y. In Figure 4.4, we show distortions of Y using $W^{\theta}_{Z_i,\alpha}, i = 1,\dots,4; \alpha = 1.5; \theta = 0.5$ against the baseline probability distribution. We see that the distortions are more pronounced in Figures 4.4a and 4.4d especially near the tails. Therefore, we expect Z_1, Z_4 are more influential inputs in the model.

We repeat the process for $\alpha = 1, 1.5, 2, 2.5, 3, 3.5$ to observe the changes in expected shortfall in Y as α increases. We note here that $\alpha = 1$ corresponds to KL-divergence and $\alpha = 2$ corresponds to the χ^2-divergence. In Figure 4.5, we plot the expected shortfall of Y at 95% confidence level using weights $W^{\theta}_{Z_i,\alpha}$ for $i = 1,\dots,4; \alpha = 1, 1.5, 2, 2.5, 3, 3.5; \theta = 0.5$.

From Figure 4.5, we see that Z_1 and Z_4 have the greatest influence on Y as the expected shortfall of Y is highest under $W^{\theta}_{Z_1,\alpha}, W^{\theta}_{Z_2,\alpha}$ for all values of α. Z_2 and Z_3 have the least impact on Y and the expected shortfall of Y under $W^{\theta}_{Z2,\alpha}$ and $W^{\theta}_{Z_3,\alpha}$ are fairly similar. We observe that as the value of α increases, the expected shortfall decreases. The decrease is more significant for Z_1 followed by Z_4. The sensitivities of Z_i also converge to each other as α increases. This is because the Radon-Nikodym derivative transitions from a convex function to a concave function as α increases, leading to a less impactful weighting of tail

(a) Distortion of Y using $W^{0.5}_{Z_1,1.5}$

(b) Distortion of Y using $W^{0.5}_{Z_2,1.5}$

(c) Distortion of Y using $W^{0.5}_{Z_3,1.5}$

(d) Distortion of Y using $W^{0.5}_{Z_4,1.5}$

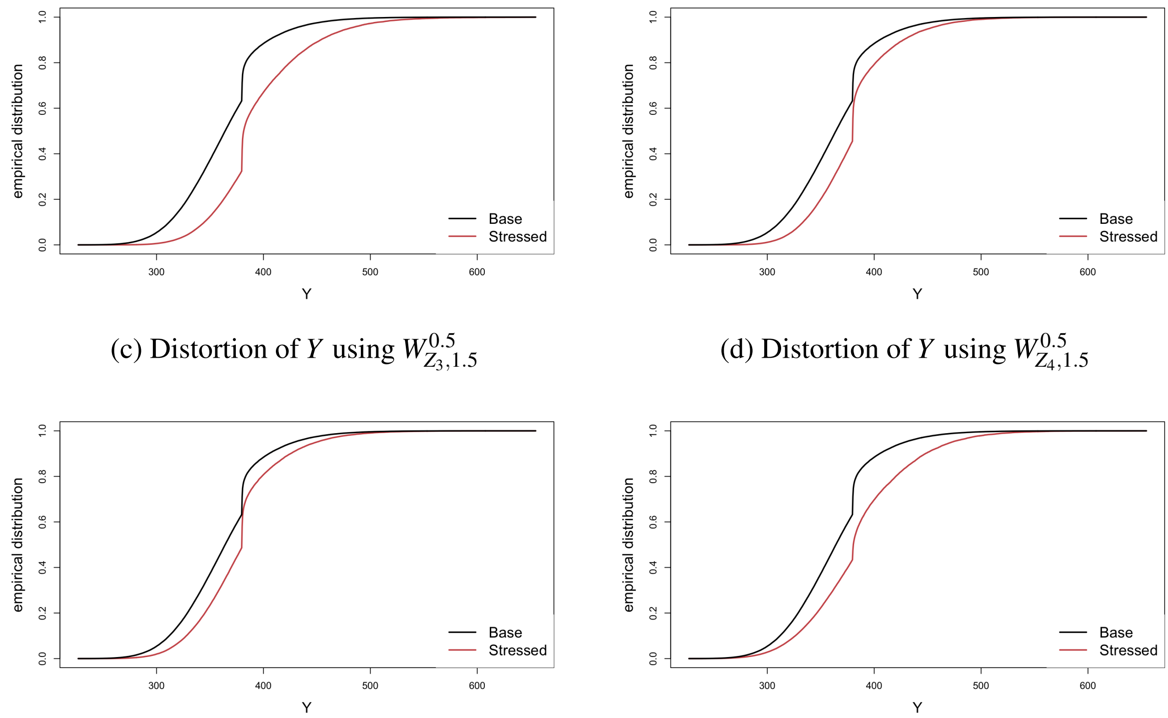


Fig. 4.4 Distortions of Y using Radon-Nikodym derivatives $W^{\theta}_{Z_i,\alpha}$ for $i = 1,\dots,4; \alpha = 1.5; \theta = 0.5$.

probabilities. To illustrate the changes in Radon-Nikodym derivatives as α is varied, we plot in Figure 4.6 the Radon-Nikodym derivatives of Z_1 obtained when Problem (III) is solved for different values of $\alpha = 1.5, 2, 2.5$ when $\theta = 0.5$. For $\alpha = 1.5$, the Radon-Nikodym derivative is a convex function with increasing weights. At $\alpha = 2$, we obtain a piece-wise linear function and for $\alpha = 2.5$, the Radon-Nikodym derivative becomes concave. Hence, the weights increase less steeply as α increases, implying that larger realisations of Y are assigned a weight which is less severe. As a result, the expected shortfall of Y decreases as α increases.

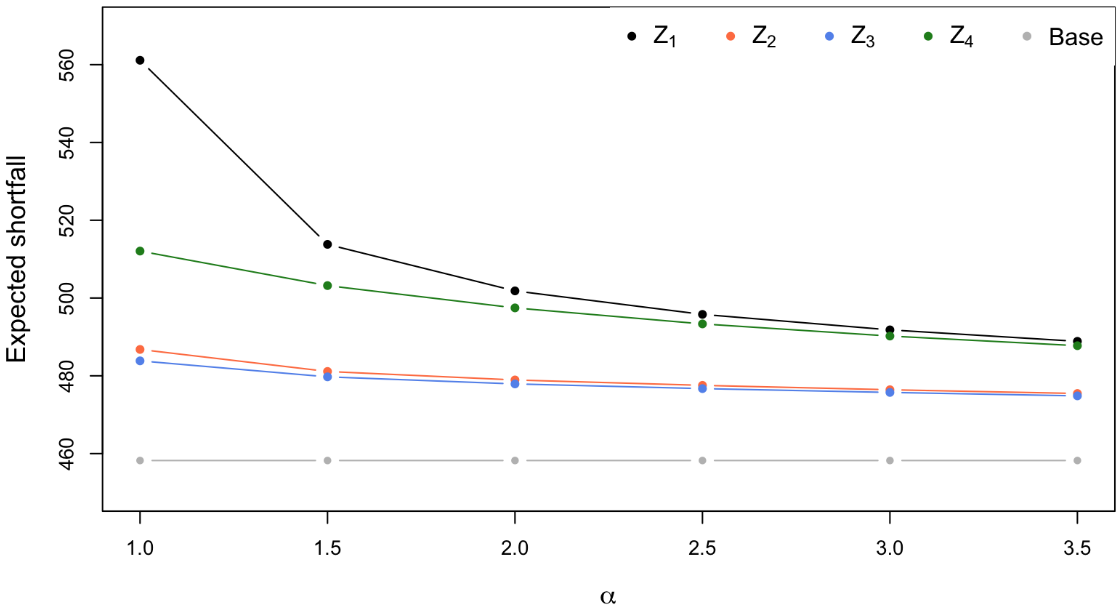


Fig. 4.5 $\mathrm{ES}_{95\%}(Y)$ using Radon-Nikodym derivatives $W^{\theta}_{Z_i,\alpha}$ for $i = 1,\dots,4; \alpha = 1, 1.5, 2, 2.5, 3, 3.5; \theta = 0.5$.

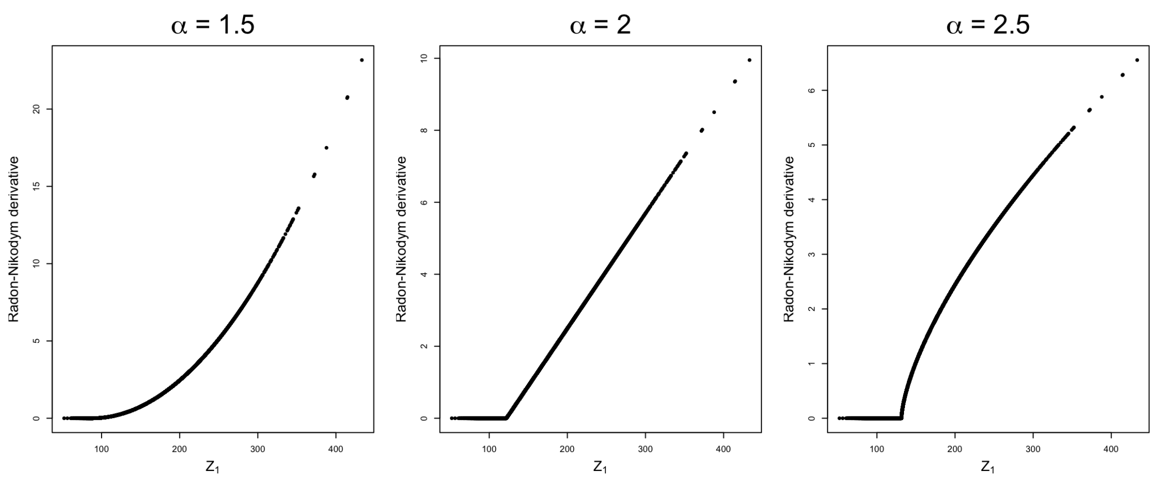


Fig. 4.6 Radon-Nikodym derivatives $W^{0.5}_{Z_1,1.5}, W^{0.5}_{Z_1,2}, W^{0.5}_{Z_1,2.5}$.

In Table 4.3, we show the mean, standard deviation, Value-at-Risk and Expected Shortfall at 95% confidence level for model input factors under baseline model and under Radon-Nikodym derivatives $W^{0.5}_{Z_i,1.5}$. The displayed numbers corroborate our findings. However, note that the impact on the standard deviation of Y is somewhat different to the impact on the expected shortfall, with Z_4 being the most important risk factor with respect to the standard deviation.

We note here that a stress does not always result in an increase in the standard deviation. The impact of a stress depends on various factors, including the shape of the distribution. Here, the standard deviations of Z_2, Z_3 do not change significantly. This could also be due to a concentration of data points around the new mean. As a consequence, the spread or variability of the distribution becomes smaller, leading to a lower standard deviation.

Table 4.3 Statistics of Y under corresponding $W^{\theta}_{Z_i,\alpha}$ for $\alpha = 1.5$ and $\theta = 0.5$.

	Base	$W^{0.5}_{Z_1,1.5}$	$W^{0.5}_{Z_2,1.5}$	$W^{0.5}_{Z_3,1.5}$	$W^{0.5}_{Z_4,1.5}$
Mean	361.64	393.50	379.07	375.96	384.57
Sd	39.42	44.84	38.49	39.88	48.33
$\text{VaR}_{95\%}$	428.48	480.07	451.58	449.41	472.39
$\text{ES}_{95\%}$	458.22	513.79	481.12	479.73	503.20

In Figure 4.7, we plot the expected shortfall of Y under $W^{\theta}_{Z_i,\alpha}$ when $\alpha = 1.5$ and θ is varied. We solve Problem (III) by varying θ from 0.5 to 2.5. We observe that the expected shortfall of all Z_i increases as θ increases. The expected shortfall increases more substantially for Z_1 as compared to other input factors owing to the fact that the tail statistics are more sensitive for Z_1. Hence, a larger divergence value induces higher stresses on the risk factors which subsequently affects the output Y.

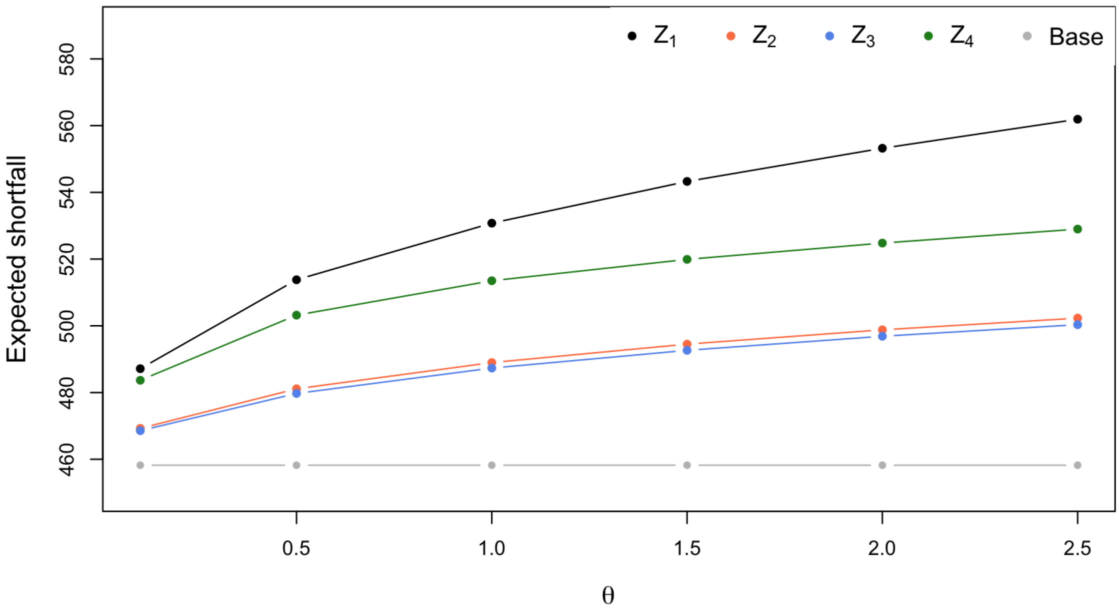
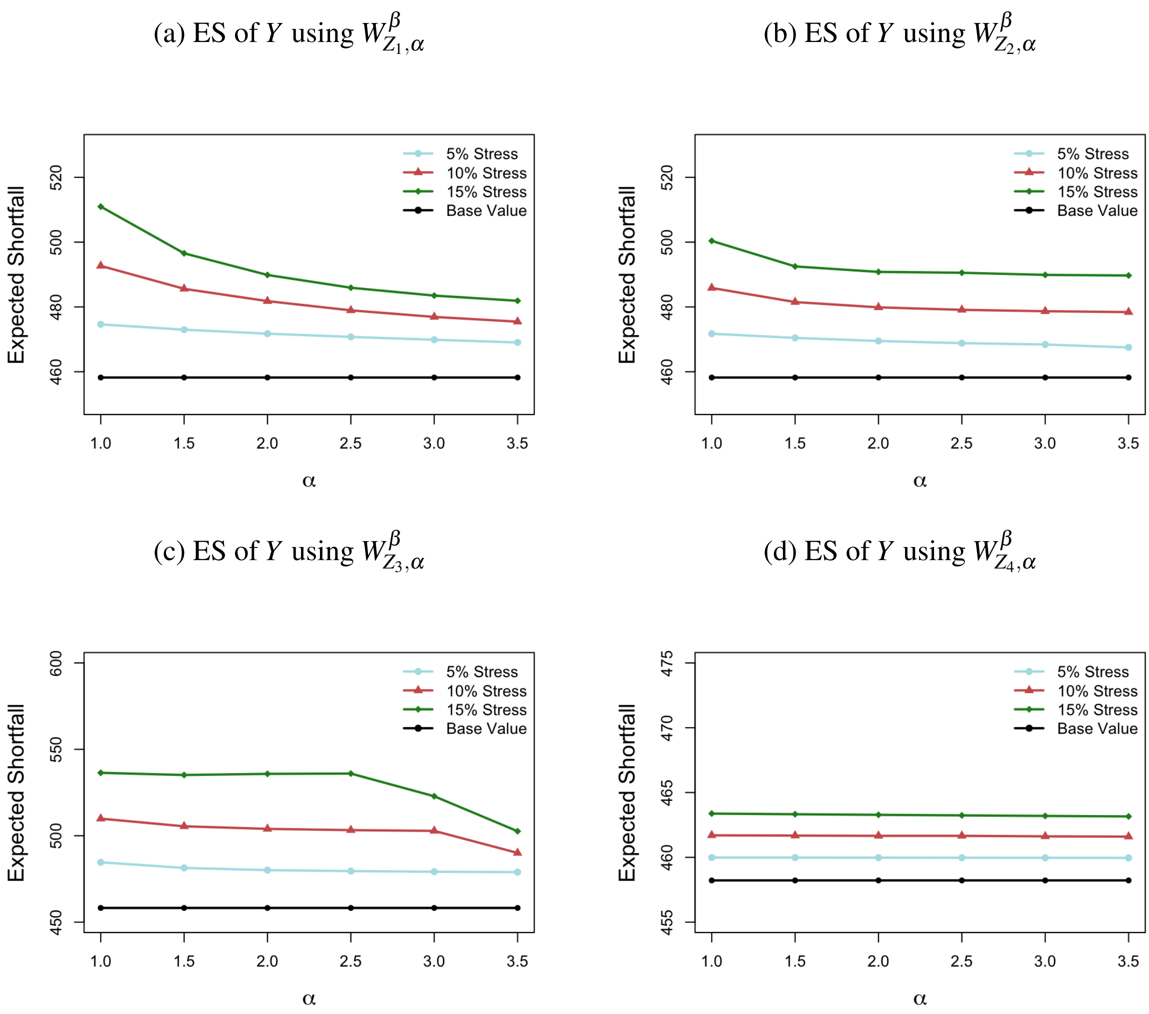


Fig. 4.7 $\mathrm{ES}_{95\%}(Y)$ under $W^{\theta}_{Z_i,\alpha}$ for $\alpha = 1.5$ and $\theta = 0.1, 0.5, 1, 1.5, 2, 2.5$.

The above methodology could potentially introduce inconsistencies in the derivation of stressed distributions. This is because adjusting the parameter α leads to different divergences and hence the metric used (end the interpretation of θ) is itself changing. We present an alternate way of examining the impact of the parameter α. We choose a value of α and we separately stress the mean of inputs by a factor of β in Problem (II) by setting $X = Z_i$, and $t = (1+\beta)\mathbb{E}(Z_i), i = 1, \dots, 4$. We repeat this process for $\alpha = 1, 1.5, 2, 2.5, 3, 3.5$ and $\beta = 5\%, 10\%, 15\%$. The Radon-Nikodym derivatives obtained as a result are denoted by $W^{\beta}_{Z_i,\alpha}$. Then we evaluate the stressed expected shortfall of output Y at 95% confidence level.

In Figure 4.8a, we plot the expected shortfall of Y under $W^{\beta}_{Z_1,\alpha}$ for varying values of α and β. The baseline expected shortfall is represented by black points. As expected, for all α values, the expected shortfall of Y increases with higher stress levels. However, as α increases, the stressed expected shortfall of Y decreases, reflecting a lower emphasis of the stress on the extreme tail. Subsequently, we plot the expected shortfall of Y under $W^{\beta}_{Z_i,\alpha}, i = 2, 3, 4$. We find similar patterns in expected shortfall of Y when stressing the means of Z_2, Z_3. Note though that for Z_4 the change in α does not impact the stressed expected shortfall of Y.



(a) ES of Y using $W^{\beta}_{Z_1,\alpha}$
(b) ES of Y using $W^{\beta}_{Z_2,\alpha}$
(c) ES of Y using $W^{\beta}_{Z_3,\alpha}$
(d) ES of Y using $W^{\beta}_{Z_4,\alpha}$

Fig. 4.8 $\mathrm{ES}_{95\%}(Y)$ under $W^{\beta}_{Z_i,\alpha}$ for $\alpha = 1,\ldots,3.5$ and $\beta = 0.05, 0.1, 0.15$.

4.6 Conclusion

We have provided a framework in a discrete setting to obtain stressed probability measures using a divergence minimisation problem under a moment constraint. We present a generalised version of the problem as stated in Chapter 2 to consider a general f-divergence instead of the χ^2-divergence. We present results that characterise the Radon-Nikodym derivative based on the shape of the derivative of the divergence and the existence of null weights.

We also present results that focus on stochastic comparisons when one parameter in the problem is altered. Specifically, we show that there is a stochastic dominance between distribution functions when two different stresses are used to solve the divergence minimisation problem. The probability measure induced by a higher stress would stochastically dominate the measure induced by a lower stress. Additionally, we also show a convex ordering of distribution functions when different divergences are used to solve the problem for a particular stress, comparing a strictly convex divergence with the χ^2-divergence.

Lastly, we demonstrate the robustness of rankings of input factors using sensitivity measures when different divergences are used. We observe consistent importance rankings of input factors when α-divergence with different α values is used. Nonetheless, using a divergence with higher α, makes the respective impacts of stressing different input factors less pronounced.

4.7 Appendix: Proofs

Proposition (4.3.1). The result is obvious for $t = \bar{x}$. Assume by contradiction that, for some $\bar{x} < t < x_n$, the solution $\mathbf{w}^*$ is not strictly positive. We distinguish two cases, depending on the number of positive terms in $\mathbf{w}^*$.

Suppose there is a single positive term $w_j^* > 0$ and $w_i^* = 0$ for $i \neq j$. By the constraints of Problem (II), it must be the case that $w_j^* = 1/p_j, x_j = t$ and $1 < j < n$. Fix then l, k such that $1 \leq h < j < k \leq n$ and define, for all $\varepsilon \geq 0$, $\mathbf{w}(\varepsilon)$ by

$$w_h(\varepsilon) = \varepsilon,\, w_i(\varepsilon) = w_i^* = 0 \text{ for all } i \neq h, j, k,$$

and $w_j(\varepsilon), w_k(\varepsilon)$ as the unique solution of

$$\sum_{i=1}^{n} p_i w_i(\varepsilon) = 1,\, \sum_{i=1}^{n} p_i w_i(\varepsilon) x_i = t.$$

Note that $w_j(\varepsilon) = A_1 + B_1\varepsilon$ and $w_k(\varepsilon) = B_2\varepsilon$ with $B_2 = \dfrac{p_h(x_j - x_h)}{p_k(x_k - x_j)} > 0$. As $\mathbf{w}(0) = \mathbf{w}^*$, there exists ε_0 such that $w_j(\varepsilon) > 0$ for all $0 \leq \varepsilon < \varepsilon_0$ and the vector $\mathbf{w}(\varepsilon)$ is feasible for Problem (II).

Define now $F(\varepsilon) = \sum_{i=1}^{n} p_i f(w_i(\varepsilon))$ the f-divergence corresponding to the weights $\mathbf{w}(\varepsilon)$, and note that $F(0) = \sum_{i=1}^{n} p_i f(w_i^*)$. A straightforward calculation gives, for $\varepsilon > 0$,

$$F'(\varepsilon) = p_h f'(\varepsilon) + p_j f'(w_j(\varepsilon))B_1 + p_k f'(w_k(\varepsilon))B_2.$$

As $\varepsilon \to 0, w_j(\varepsilon) \to w_j^* = 1/p_j > 0$ and $w_k(\varepsilon) \to 0$. Since $B_2 > 0$, it follows that $F'(0+) = \lim_{\varepsilon \downarrow 0} F'(\varepsilon) = -\infty$, contradicting the optimality of $\mathbf{w}^*$.

Suppose now there are at least two positive terms $w_{j_1}^* > 0, w_{j_2}^* > 0$ for some $1 \leq j_1 < j_2 \leq n$. Let h be such that $w_h^* = 0$. The proof follows similarly to the previous case. For all

$\varepsilon \geq 0$, define $\mathbf{w}(\varepsilon)$ by

$$w_h(\varepsilon) = \varepsilon,\, w_i(\varepsilon) = w_i^* \text{ for all } i \neq h, j_1, j_2,$$

and $w_{j_1}(\varepsilon), w_{j_2}(\varepsilon)$ as the unique solution of

$$\sum_{i=1}^{n} p_i w_i(\varepsilon) = 1,\, \sum_{i=1}^{n} p_i w_i(\varepsilon) x_i = t.$$

Note that $w_{j_1}(\varepsilon) = A_1 + B_1\varepsilon$, $w_{j_2}(\varepsilon) = A_2 + B_2\varepsilon$ for some $A_1, A_2, B_2, B_2 \in \mathbb{R}$. As $\mathbf{w}(0) = \mathbf{w}^*$, there exists $\varepsilon_0 > 0$ such that for all $0 \leq \varepsilon < \varepsilon_0$, $w_{j_1}(\varepsilon) > 0$, $w_{j_2}(\varepsilon) > 0$, and the vector $\mathbf{w}(\varepsilon)$ is feasible for problem (II).

Define now $F(\varepsilon) = \sum_{i=1}^{n} p_i f(w_i(\varepsilon))$ and note that $F(0) = \sum_{i=1}^{n} p_i f(w_i^*)$. A straightforward calculation gives, for $\varepsilon > 0$,

$$F'(\varepsilon) = p_h f'(\varepsilon) + p_{j_1} B_1 f'(w_{j_1}(\varepsilon)) + p_{j_2} B_2 f'(w_{j_2}(\varepsilon)).$$

As $\varepsilon \to 0, w_{j_1}(\varepsilon) \to w_{j_1}^* > 0, w_{j_2}(\varepsilon) \to w_{j_2}^* > 0$, and it follows that $F'(0+) = \lim_{\varepsilon \downarrow 0} F'(\varepsilon) = -\infty$, contradicting the optimality of $\mathbf{w}^*$. □

Proposition (4.3.2). Recall that the KKT conditions are necessary and sufficient for optimality of a candidate solution $\mathbf{w}^*$.

We use the method of contradiction to prove $\lambda_2 > 0$. Let us suppose $\lambda_2 \leq 0$. Consider $1 \leq h < j \leq n$ and assume $w_j^* > 0$. We have

$$\begin{aligned} p_h f'(w_h^*) &= p_h\lambda_1 + p_h\lambda_2 x_h + \mu_h \\ &\geq p_h\lambda_1 + p_h\lambda_2 x_h && (\mu_h \geq 0) \\ &\geq p_h(\lambda_1 + \lambda_2 x_j) && (\lambda_2 \leq 0, x_h < x_j) \\ &= p_h f'(w_j^*) && (\mu_j = 0) \end{aligned}$$

so that $w_h^* \geq w_j^*$. The same conclusion clearly holds if $w_j^* = 0$ and, necessarily, $f'(0)$ is finite. This implies that there is a counter-monotonic relationship between X and W^* as the w_i^* are non-increasing in i, contradicting the assumption that $t > \bar{x}$: by Chebyshev's Sum Inequality,

$$t = \sum_{i=1}^{n} p_i x_i w_i^* \leq \sum_{i=1}^{n} p_i w_i^* \sum_{i=1}^{n} p_i x_i = \bar{x}.$$

Hence, we conclude that $\lambda_2 > 0$. To check that the weights are non-decreasing, consider $1 \leq h < j \leq n$ with $w_h^* > 0$. Proceeding similarly as before,

$$\begin{aligned} p_j f'(w_j^*) &= p_j\lambda_1 + p_j\lambda_2 x_j + \mu_j \\ &\geq p_j\lambda_1 + p_j\lambda_2 x_j \\ &> p_j(\lambda_1 + \lambda_2 x_h) \\ &= p_j f'(w_h^*). \end{aligned}$$

Hence, $w_h^* \leq w_j^*$. The same conclusion clearly holds if $w_h^* = 0$. This implies that w_i is non-decreasing with i.

□

Proposition (4.3.3). Consider a sequence of mean stresses (t_h) with $\bar{x} < t_h < x_n$ and such that $t_h \uparrow x_n$. For each h, let $\mathbf{w}^*(h)$ be the optimal solution of Problem (II) with $t = t_h$. In particular $\mathbf{w}^*(h)$ is feasible for Problem (II), so $(\mathbf{w}^*(h))$ is a bounded sequence and there exists a converging sub-sequence, denoted again $(\mathbf{w}^*(h))$ for simplicity, such that $\mathbf{w}^*(h) \to \mathbf{w}^*$ as $h \to +\infty$ for some $\mathbf{w}^*$ in $\mathbb{R}^n$. Taking the limit in the constraints of Problem (II) one gets

$$\sum_{i=1}^{n} p_i w_i^* = 1, \qquad \sum_{i=1}^{n} p_i w_i^* x_i = x_n,$$

from which it follows that $w_i^* = 0$ for $i < n$ and $w_n^* = \frac{1}{p_n} > 0$.

Since each solution $\mathbf{w}^*(h)$ is strictly positive, it satisfies the equation

$$f'(w_i^*(h)) = \lambda_1(h) + \lambda_2(h)x_i,\, i = 1, \ldots, n,$$

where $\lambda_1(h), \lambda_2(h) > 0$ are for all h, the corresponding KKT multipliers. Taking the limit as $h \to +\infty$ in the last expression, one finds that

$$\lambda_1(h) + \lambda_2(h)x_i \to f'(0), \text{ for } i < n \tag{i}$$

$$\lambda_1(h) + \lambda_2(h)x_n \to f'(1/p_n). \tag{ii}$$

If $f'(0) < +\infty$ then, from (i), for all $1 \leq i < j < n$,

$$\lambda_2(h)(x_j - x_i) \to 0 \text{ as } h \to +\infty,$$

so that $\lambda_2(h) \to 0$. From (i) and (ii) it now follows that for $1 \le i < n$,

$$\lambda_2(h)(x_n - x_i) \to f'(1/p_n) - f'(0) > 0,$$

a contradiction. □

Lemma (4.3.1). If $f'(0) = -\infty$, by Proposition 4.3.1 $w_i^* > 0$ for all i and by (KKT II),

$$f'(w_i^*) = \lambda_1 + \lambda_2 x_i.$$

So $w_i^* = g(\lambda_1 + \lambda_2 x_i)$ and the result follows.

Assume now that $f'(0) > -\infty$. If $w_i^* > 0$ then

$$f'(w_i^*) = \lambda_1 + \lambda_2 x_i > f'(0)$$

so that $w_i^* = g(\lambda_1 + \lambda_2 x_i)$ and the conclusion follows. If $w_i^* = 0$ then by (KKT II),

$$p_i f'(w_i^*) = p_i\lambda_1 + p_i\lambda_2 x_i + \mu_i \ge p_i\lambda_1 + p_i\lambda_2 x_i.$$

Hence, $f'(w_i^*) = f'(0) \ge \lambda_1 + \lambda_2 x_i$ so that $w_i^* = g(\max(f'(0), \lambda_1 + \lambda_2 x_i)) = 0$. □

Proposition (4.3.4). The KKT conditions for Problem (III) are as follows: if $\mathbf{v}$ is optimal for Problem (III), there exists multipliers η_1, η_2 and $\varepsilon_i, i = 1, \ldots, n$ such that

$$\begin{aligned} \eta_2 p_i f'(v_i) &= -\eta_1 p_i - p_i x_i - \varepsilon_i, \quad v_i\varepsilon_i = 0, \\ \eta_2\left(\sum_i p_i f(v_i) - \theta\right) &= 0, \qquad \varepsilon_i \ge 0, \\ \sum_i p_i v_i &= 1, \qquad \eta_2 \le 0, \\ \sum_i p_i f(v_i) &\le \theta, \qquad v_i \ge 0. \end{aligned} \tag{KKT III}$$

It can be confirmed that if $\mathbf{w}^*$ is the unique solution of Problem (II), it satisfies (KKT III) with $\theta = \theta^* := \sum p_i f(w_i^*)$ after setting $\eta_2^* = -\frac{1}{\lambda_2}, \eta_1^* = \frac{\lambda_1}{\lambda_2}$ and $\varepsilon_i^* = \frac{\mu_i}{\lambda_2}$, where λ_1, λ_2 and μ_i are the Lagrangian multipliers of Problem (II), see (KKT II). As III is a convex problem, $\mathbf{w}^*$ solves Problem (III).

Conversely, if $\mathbf{v}^*$ solves Problem (III) for some $\theta \ge f(1)$, then $\mathbf{v}^*$ satisfies (KKT III) and it can be verified that it satisfies (KKT II) with $t = t^* := \sum p_i v_i^* x_i$ after setting $\lambda_1 =$

$\dfrac{-\eta_1}{\eta_2}, \lambda_2 = \dfrac{-1}{\eta_2}, \mu_i = \dfrac{-\varepsilon_i}{\eta_2}$ where η_1, η_2 and ε_i are the Lagrangian multipliers of Problem (III). Hence $\mathbf{v}^*$ is optimal for Problem (II).

□

The next Theorem follows directly from standard results; for detail see Kaas et al. [52], Denuit et al. [32], Shaked and Shanthikumar [79], Karlin and Novikoff [53]. We add a proof in the interest of completeness, because the case of strict stochastic dominance is usually not explicitly covered.

Remark 4.7.1. Note that in the proof of Theorem 4.7.1, we use throughout the fact that if two measures Q', Q'' corresponding to Radon-Nikodym derivative weights $\mathbf{w}'$, $\mathbf{w}''$ are not identical, there exists some i, j such that $w'_i > w''_i$ and $w'_j < w''_j$. Furthermore if the conditions of Theorem 4.7.1(2) are satisfied, then within each of the ranges of indices considered, there will be at least one index where the stated inequalities is strict.

Theorem 4.7.1. *Consider two probabilities Q', Q'', with corresponding Radon-Nikodym derivatives, $\mathbf{w}'$, $\mathbf{w}''$.*

1. *Assume there is an $i^* \in \{1, \dots, n-1\}$ such that for each $i \leq i^*$, it holds that $w'_i \geq w''_i$ and for each $i > i^*$, it holds that $w'_i \leq w''_i$. Then $F_X^{Q'} \preceq_{\mathrm{st}} F_X^{Q''}$. If in addition $Q' \neq Q''$, then $F_X^{Q'} \prec_{\mathrm{st}} F_X^{Q''}$.*
2. *Let $\mathbb{E}_{Q'}[X] = \mathbb{E}_{Q''}[X]$ and assume there are i^*, i^{**} with $1 \leq i^* < i^{**} < n$ such that*
   - *for each $i \leq i^*$, it holds that $w'_i \leq w''_i$;*
   - *for each $i^* < i \leq i^{**}$, it holds that $w'_i \geq w''_i$;*
   - *for each $i > i^{**}$, it holds that $w'_i \leq w''_i$.*

   Then $F_X^{Q'} \preceq_{\mathrm{cx}} F_X^{Q''}$.

Proof. 1. For all $x \leq x_i^*$:

$$F_X^{Q'}(x) = \sum_{i:x_i \leq x} w'_i p_i \geq \sum_{i:x_i \leq x} w''_i p_i = F_X^{Q''}(x).$$

Similarly, for all $x > x_i^*$:

$$F_X^{Q'}(x) = 1 - \sum_{i:x_i > x} w'_i p_i \geq 1 - \sum_{i:x_i > x} w''_i p_i = F_X^{Q''}(x).$$

This proves the first-order stochastic dominance.

If in addition $Q' \neq Q''$, there exists $j \leq i^*$ such that $w'_j > w''_j$ then

$$F_X^{Q'}(x_j) = \sum_{i \leq j} w'_i p_i > \sum_{i \leq j} w''_i p_i = F_X^{Q''}(x_j).$$

If instead there exists $j > i^*$ such that $w'_j < w''_j$ then

$$F_X^{Q'}(x_{j-1}) = 1 - \sum_{i \geq j} w'_i p_i > 1 - \sum_{i \geq j} w''_i p_i = F_X^{Q''}(x_{j-1}).$$

Therefore, strict stochastic dominance holds.

2. To prove $F_X^{Q'} \preceq_{\text{cx}} F_X^{Q''}$, it is enough to show that $\mathbb{E}^{Q'}[X-x]_+ \leq \mathbb{E}^{Q''}[X-x]_+$ for all $x \in \mathbb{R}$, see Denuit et al. [32]. Note that

$$\mathbb{E}^{Q'}([X-x]_+) = \sum_{x_i > x} p_i w'_i (x_i - x),$$

A similar expression holds for $\mathbb{E}^{Q''}([X-x]_+)$.

We now show that $G(x) \leq 0$ for all $x \in \mathbb{R}$, where

$$G(x) = \sum_{x_i > x} p_i (w'_i - w''_i)(x_i - x) \quad = \sum_{x_i > x} p_i (w'_i - w''_i) x_i - x \sum_{x_i > x} p_i (w'_i - w''_i). \tag{iii}$$

We first check that $G(x_j) \leq 0$ for all $j = 1, \ldots, n-1$. Letting $\Delta x_h = x_h - x_{h-1}$, we have

$$\begin{aligned} G(x_j) &= \sum_{i > j} p_i (w'_i - w''_i)(x_i - x_j) \\ &= \sum_{i > j} p_i (w'_i - w''_i) \sum_{j < h \leq i} \Delta x_h \\ &= \sum_i \mathbb{1}_{i > j} p_i (w'_i - w''_i) \sum_h \mathbb{1}_{j < h} \mathbb{1}_{h \leq i} \Delta x_h \\ &= \sum_h \mathbb{1}_{j < h} \Delta x_h \sum_i \mathbb{1}_{i > j} \mathbb{1}_{i \geq h} p_i (w'_i - w''_i) \\ &= \sum_{h > j} \Delta x_h K_h, \end{aligned} \tag{iv}$$

where $K_h = \sum_{i \geq h} p_i (w'_i - w''_i)$.

Note that

$$K_h = p_h(w'_h - w''_h) + K_{h+1},$$

hence K_h is increasing in h on $(1,\dots,i^*)$ and $(i^{**},\dots,n)$ and decreasing in h on $(i^*+1,\dots,i^{**}-1)$. Further,

$$K_1 = \sum_{i\geq 1} p_i(w'_i - w''_i) = 0,$$

so that $K_h \geq 0$ for $1 \leq h \leq i^*$. Similarly,

$$K_n = p_n(w'_n - w''_n) \leq 0$$

for $i^{**} \leq h \leq n$.

Therefore, there exists $h^* \in (i^*+1, i^{**}-1)$ such that $K_h \geq 0$ for $h \leq h^*$ and $K_h \leq 0$ for $h > h^*$.

From (iv), it follows that $G(x_j) = \Delta x_{j+1} K_{j+1} + G(x_{j+1})$. Hence, $G(x_j)$ is decreasing for $j \in (1, h^*-1)$ and increasing for $j \in (h^*, n)$. From the assumption $\mathbb{E}^{Q'}(X) = \mathbb{E}^{Q''}(X)$ it follows that $G(x_1) = G(x_n) = 0$, which implies that $G(x_j) \leq 0$ for all j.

Finally, we can state that $G(x) \leq 0$ for all x from (iii) and $G(x)$ is linear in $[x_i, x_{i+1}]$.

□

Proposition (4.4.1). First we show $\lambda''_2 > \lambda'_2$.

Let λ'_1, λ_2' and λ_1'', λ_2'' be the corresponding Lagrangian multipliers for Problem (II) when $t = t', t = t''$ respectively.

The proof starts by contradiction. We assume $\lambda_2'' \leq \lambda_2'$ i.e., $\lambda_2'' = \lambda_2'$ or $\lambda_2'' < \lambda_2'$.

If $\lambda_2'' = \lambda_2'$, then either $\lambda_1' = \lambda_1''$ which is not possible as $t'' > t'$ or $\lambda_1' \neq \lambda_1''$ which is also not feasible as one of $\mathbf{w}', \mathbf{w}''$ would dominate the other.

If $\lambda_2'' < \lambda_2'$, we prove that $F_X^{Q'} \succeq F_X^{Q''}$ which is inconsistent with $t'' > t'$.

Recall from Lemma (4.3.1),

$$w_i' = g(\max(f'(0), \lambda_1' + \lambda_2' x_i))$$

and

$$w_i'' = g(\max(f'(0), \lambda_1'' + \lambda_2'' x_i))$$

where $g = (f')^{-1}$. Note that there must be at least one i such that $w_i' < w_i''$. We show that for any such i, it holds $w_j' \leq w_j''$ for all $j < i$.

We consider 4 cases here:

(i) $f'(0) \geq \lambda_1' + \lambda_2' x_i, f'(0) \geq \lambda_1'' + \lambda_2'' x_i,$

(ii) $f'(0) < \lambda_1' + \lambda_2' x_i, f'(0) \geq \lambda_1'' + \lambda_2'' x_i,$

(iii) $f'(0) \geq \lambda_1' + \lambda_2' x_i, f'(0) < \lambda_1'' + \lambda_2'' x_i,$

(iv) $f'(0) < \lambda_1' + \lambda_2' x_i, f'(0) < \lambda_1'' + \lambda_2'' x_i.$

We reject the possibility of having cases (i) and (ii) as they violate the assumption $w_i' < w_i''$.

Case (iii):

We get $w_i' = 0 < w_i'' = g(\lambda_1'' + \lambda_2'' x_i)$. For $j < i$, we have $x_j < x_i$. Hence, $\lambda_1' + \lambda_2' x_j < \lambda_1' + \lambda_2' x_i \leq f'(0)$ since $\lambda_2' > 0$, hence $w_j' = 0$. Therefore, $w_j' \leq w_j''$ for any $j < i$.

Case (iv):

We have $g(\lambda_1' + \lambda_2' x_i) = w_i' < w_i'' = g(\lambda_1'' + \lambda_2'' x_i)$.

$\implies \lambda_1' + \lambda_2' x_i < \lambda_1'' + \lambda_2'' x_i \iff (\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_i > 0.$

Now, for $j < i, x_j < x_i$ and since $\lambda_2' > \lambda_2''$,

$(\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_j > (\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_i > 0 \implies w_j' \leq w_j''.$

Let $i^* = \max\{i : w_i' < w_i''\} < n$, then for $i \leq i^*, w_i' \leq w_i''$ and for $i > i^*, w_i' \geq w_i''$. This follows from part 1 of Theorem 4.7.1.

$\implies F_X^{Q'} \succeq F_X^{Q''}$ contradicting the assumption $\mathbb{E}^{Q'}(X) = t' < t'' = \mathbb{E}^{Q''}(X)$.

Hence, we conclude that $\lambda_2'' > \lambda_2'$.

Now, we prove that $F_X^{Q''} \succeq F_X^{Q'}$. Note that there is at least one i such that $w_i'' > w_i'$. We show that for any such i, we have $w_j'' > w_j'$ for all $j > i$. We consider the same four cases as above. Again, cases (i), (ii) are not feasible.

Case (iii): $f'(0) \geq \lambda_1' + \lambda_2' x_i, f'(0) < \lambda_1'' + \lambda_2'' x_i$

From above, we get $(\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_i > 0$.

Hence $(\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_j > (\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_i > 0$.

It follows that $\lambda_1'' + \lambda_2'' x_j > \lambda_1' + \lambda_2' x_j \implies w_j'' = g(\lambda_1'' + \lambda_2'' x_j) \geq g(\max\{f'(0), \lambda_1' + \lambda_2' x_j\}) = w_j'$.

Case (iv): $f'(0) < \lambda_1' + \lambda_2' x_i, f'(0) < \lambda_1'' + \lambda_2'' x_i$.

$w_i' = g(\lambda_1' + \lambda_2' x_i) < g(\lambda_1'' + \lambda_2'' x_i) = w''_i$.

From above, we get $\lambda_1'' + \lambda_2'' x_i > \lambda_1' + \lambda_2' x_i$.

It follows that for $j > i, x_j > x_i \implies \lambda_1'' + \lambda_2'' x_j > \lambda_1'' + \lambda_2'' x_i$.

Hence, $(\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_j > (\lambda_1'' - \lambda_1') + (\lambda_2'' - \lambda_2') x_i > 0$. We conclude that $w_j'' > w_j'$.

Let $i^* = \min\{i : w_i{}'' > w_i{}'\} > 1$.

Hence, for $i < i^*, w_i{}'' \leq w_i{}'$ and for $i \geq i^*, w_i{}'' \geq w_i{}'$ implying $F_X^{Q''} \succeq F_X^{Q'}$.

In addition, given that $\lambda_2'' \neq \lambda_2'$, it follows that the corresponding Radon-Nikodym derivatives can not be identical. Hence, strict stochastic ordering holds $F_X^{Q''} \succ F_X^{Q'}$. □

Proposition (4.4.2). We know that the solution of Problem (II) for χ^2-divergence obtained in [62] is of the form $w_i = \max(0, \lambda_1 + \lambda_2 x_i)$ where λ_1, λ_2 are the KKT multipliers. Further, using Lemma (4.3.1), $\tilde{w}_i = \tilde{g}(\max(\tilde{f}'(0), \tilde{\lambda}_1 + \tilde{\lambda}_2 x_i))$ where $\tilde{\lambda}_1, \tilde{\lambda}_2$ are the corresponding KKT multipliers when the $\tilde{f}$-divergence is used to solve Problem (II). Define the continuous functions $\mathscr{W}(x) = \max(0, \lambda_1 + \lambda_2 x)$, $\tilde{\mathscr{W}}(x) = \tilde{g}(\max(\tilde{f}'(0), \tilde{\lambda}_1 + \tilde{\lambda}_2 x))$. Denote by $\beta = \sup\{x | \mathscr{W}(x) = 0\}$ the last point where $\mathscr{W}$ is 0.

To prove $F_X^Q \preceq_{\text{cx}} F_X^{\tilde{Q}}$, we consider two cases here.

<u>Case (a)</u>: Let $\tilde{f}'(0) = -\infty$. Then, $\tilde{w}_i = \tilde{g}(\tilde{\lambda}_1 + \tilde{\lambda}_2 x_i)$.

First, we show that it is not possible for there to be a i^*, such that $w_i \leq (\geq)\ w_i^*$ for $i \leq i^*$ and $w_i \geq (\leq)\ w_i^*$ for $i > i^*$. If such an i^* did exist, then we would have by Theorem 4.7.1.1 strict stochastic dominance between F_X^Q and $F_X^{\tilde{Q}}$, implying inequality of the mean of X under the two measures, which is a contradiction.

Second, we show that Theorem 4.7.1.2 applies. To do this we consider the functions $\mathscr{W}, \tilde{\mathscr{W}}$. Note that $\tilde{\mathscr{W}}(x) = \tilde{g}(\tilde{\lambda}_1 + \tilde{\lambda}_2 x)$ is strictly convex as $\tilde{\lambda}_2 > 0$. We now show that the functions $\mathscr{W}(x), \tilde{\mathscr{W}}(x)$ cross no more than twice. By $\tilde{f}'(0) = -\infty$, it holds that $\tilde{\mathscr{W}}(x) > 0$ for all x. Hence, $\mathscr{W}, \tilde{\mathscr{W}}$ cannot intersect at a point $x \leq \beta$. Now assume that there exist z_1, z_2, z_3 with $\beta < z_1 < z_2 < z_3$ such that $\mathscr{W}(z_i) = \tilde{\mathscr{W}}(z_i), i = 1,2,3$. We select a $\lambda^* \in (0,1)$ such that $z_2 = \lambda^* z_1 + (1-\lambda^*) z_3$. Then, we obtain the contradiction:

$$
\begin{aligned}
\tilde{\mathscr{W}}(z_2) &< \lambda^* \tilde{\mathscr{W}}(z_1) + (1-\lambda^*) \tilde{\mathscr{W}}(z_3) \\
&= \lambda^* \mathscr{W}(z_1) + (1-\lambda^*) \mathscr{W}(z_3) \\
&= \mathscr{W}(z_2),
\end{aligned}
$$

where the inequality results from the strict convexity of $\tilde{\mathscr{W}}$ and the second equality from the linearity of $\mathscr{W}(x)$ for $x > \beta$. Therefore, there will be $\beta < z_1 < z_2$ such that

$$
\begin{cases}
\tilde{\mathscr{W}}(x) \geq \mathscr{W}(x) & \text{for} \quad x \leq z_1, \\
\tilde{\mathscr{W}}(x) \leq \mathscr{W}(x) & \text{for} \quad z_1 < x < z_2, \\
\tilde{\mathscr{W}}(x) \geq \mathscr{W}(x) & \text{for} \quad x \geq z_2.
\end{cases}
$$

Since the w_i and $\tilde{w}_i$ are points on the curves $\mathscr{W}$ and $\tilde{\mathscr{W}}$ respectively, it follows that the sequence of differences $w_i - \tilde{w}_i$ cannot change sign more times than the functions $\mathscr{W}$ and

$\tilde{\mathscr{W}}$ cross. Hence, the differences $w_i - \tilde{w}_i$ change sign at most twice. We know that they cannot change sign only once (contradicting equality of means) or not change sign at all (contradicting the $\mathbf{w}, \tilde{\mathbf{w}}$ being Radon-Nikodym densities). Hence, we find ourselves in the scenario of Theorem 4.7.1.2 and $F_X^Q \preceq_{\text{cx}} F_X^{\tilde{Q}}$ holds.

Case (b): Let $\tilde{f}'(0) > -\infty$.

Define $\tilde{\beta} = \sup\{x | \tilde{\mathscr{W}}(x) = 0\}$ as the first point where $\tilde{\mathscr{W}}$ is positive. If $\tilde{\mathscr{W}}(x_1) > 0 \iff \tilde{\beta} < x_1$, then the argument is identical to the case (a).

Consider now the situation that $\tilde{\beta} \geq x_1$, such that $\tilde{w}_i = 0$ for $x_i \leq \tilde{\beta}$. If $\beta < \tilde{\beta}$, then for $x > \beta$, the functions $\mathscr{W}, \tilde{\mathscr{W}}$ can only cross once, as $\mathscr{W}$ is linear in that range. Hence the sequence $w_i - \tilde{w}_i$ changes sign at most once. This in turn contradicts with the assumption that $\mathbf{w}, \tilde{\mathbf{w}}$ are Radon-Nikodym derivatives leading to $\mathbb{E}^Q[X] = \mathbb{E}^{\tilde{Q}}[X]$.

Consequently $\beta > \tilde{\beta}$. If there are any $x_i \leq \tilde{\beta}$, the corresponding weights will be $w_i = \tilde{w}_i = 0$. Such points do not impact the change in sign of the sequence $w_i - \tilde{w}_i$ and hence can be ignored in the application of Theorem 4.7.1. For $x_i > \tilde{\beta}$, we find ourselves once more in a similar situation as case (a), such that $F_X^Q \preceq_{\text{cx}} F_X^{\tilde{Q}}$ holds. $\square$

Chapter 5

Conclusion and future work

5.1 Summary

The overarching goal of this thesis was to develop a computationally effective sensitivity approach that can be applied to simulation models in insurance and risk management. We present a reverse and forward sensitivity approach where the discrepancy between the baseline probability distribution and the alternate probability distribution is evaluated using f-divergences. This difference is determined by subjecting the expectation of the output to a stress. Constraining the value of the divergence, the forward analysis is conducted involving stressing model input factors. The χ^2-divergence was explicitly used in Chapter 2 to derive alternate probability distributions. Our contribution falls within the domain of optimisation problems where f-divergences are minimised to produce alternative probability measures. While the sensitivity analysis approach we present in this thesis builds upon reverse sensitivity testing introduced in Pesenti et al. [67], we add another layer by interlinking both reverse and forward sensitivity analyses in a consistent manner. The interlinking of the two analyses provides a more comprehensive understanding of the relation between model inputs and output. The sensitivity analysis approach developed in Chapter 2 was applied specifically to determine the rankings of model inputs of a simple insurance portfolio model and later applied in Chapter 3 to test for the effects of model and parameter uncertainty in Solvency II Standard model for non-life premium and reserve sub-module. Further, the optimisation problem was generalised to accommodate a broader class of f-divergences. This allows a modeller the freedom to choose between various f-divergences and to make appropriate comparisons without being limited to one specific f-divergence. By understanding the attributes of solutions obtained using different divergences, an analyst can make better informed choices.

5.2 Future Work

There are several possible extensions of the research presented in this thesis. The first would be to extend the research presented in Chapter 4. In Chapter 4, some conditions were presented to investigate stochastic dominance and convex ordering among different divergences. However, further conditions can be explored to establish convex ordering between two divergences such that one is more convex than the other.

One can also establish conditions on model components that detect (in)consistent rankings among input factors when reverse and forward sensitivity approaches are used. An inconsistency in the rankings is referred to as a *dissonance* by Cooke and van Noortwijk [28]. The reverse and forward approaches were interlinked so as to detect any dissonance that may surface and it is important to detect anomalies as the critical points of importance in a sensitivity analysis may differ.

Another potential area for future research in the application of sensitivity analysis approach is to address problems involving multiple constraints in a finite space. I have so far only dealt with optimisation problems where f-divergences have been minimised subject to a constraint only on one moment. One can generalise the optimisation problem further, such that f-divergences are minimised subject to multiple moment constraints. This would open up several potential avenues for a more comprehensive understanding of model uncertainty. Plausible scenarios can be constructed where moments of several input factors are simultaneously modified providing a more thorough assessment of the interactions between inputs and output. Alternatively, several moments of a single factor could be stressed. Karush-Kuhn-Tucker conditions [61] can be used to obtain the characterisation of the solution for the Radon-Nikodym derivative for optimisation problems with multiple constraints. While a full analytical solution will not be feasible for such a general optimisation problem however, partial theoretical results can be developed which would allow for an easy implementation of a sensitivity analysis using Monte-Carlo simulation and reduce the computational cost to an extent.

Further, robust optimisation can be used as a tool to investigate more deeply the relationship between sensitivity and model uncertainty. While our sensitivity analysis methodology shares some parallels with robust optimisation, there are notable differences. In robust optimisation, a set of probability measures is considered, and decision-making is based on the worst-case scenario. An example of this approach can be observed in optimal decision making with maxmin preferences, as demonstrated by Gilboa and Schmeidler [43]. In such cases, uncertainty is often defined as a ball around the probability distribution P, utilizing a specific distance metric. In contrast, our approach involves explicitly deriving the probability measures, which introduces a predefined distortion to our model. By explicitly deriving the probability measures, we can gain insights into the uncertainties present in the system.

Defining an appropriate uncertainty region for a robust optimisation problem is important and Ben-Tal et al. [11] has constructed uncertainty regions for various f-divergences. The uncertainty set would be finite if the optimisation problem is subjected to a finite number of constraints but, in general, it can be infinite and therefore literature on general theory of robust optimisations would need to be considered to solve such kind of problems. Formulating robust optimisation problems that concurrently address model uncertainty and stress testing would be a key output for such a strand of research. Papers that can be considered for optimisation problems in the context of robust optimisation are Ben-Tal et al. [12, 11], Bertsimas et al. [15], Lam [56].

References

[1] Adhikari, B. and Joshi, D. (1956). Distance, discrimination et résumé exhaustif. *Publications de l'Institut Statistics*, University de Paris 5:57–74.

[2] AIG Holdings Europe (2020). Aig holdings europe ltd 2019 solvency and financial condition report. https://www.aig.co.uk/content/dam/aig/emea/united-kingdom/documents/Report-and-Accounts/2019-group-solvency-report.pdf. Accessed: February 21, 2023.

[3] Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142.

[4] Arbenz, P., Hummel, C., and Mainik, G. (2012). Copula based hierarchical risk aggregation through sample reordering. *Insurance: Mathematics and Economics*, 51(1):122–133.

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

[6] Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Stat. 9*, page 2023–2050.

[7] Auder, B. and Iooss, B. (2008). Global sensitivity analysis based on entropy. In *Safety, reliability and risk analysis-Proceedings of the ESREL 2008 Conference*, pages 2107–2115.

[8] Bedford, T., Cooke, R., et al. (2001). *Probabilistic risk analysis: foundations and methods*. Cambridge University Press.

[9] Bellini, F. (2012). Convex comparison of minimal divergence martingale measures in discrete time models. *Available at SSRN 2113976*.

[10] Bellini, F. and Frittelli, M. (2002). On the existence of minimax martingale measures. *Mathematical Finance*, 12(1):1–21.

[11] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

[12] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*, volume 28. Princeton University Press.

[13] Bergenthum, J. and Rüschendorf, L. (2006). Comparison of option prices in semimartingale models. *Finance and Stochastics*, 10(2):222–249.

[14] Bermúdez, L., Ferri, A., and Guillén, M. (2013). A correlation sensitivity analysis of non-life underwriting risk in solvency capital requirement estimation. *ASTIN Bulletin: The Journal of the IAA*, 43(1):21–37.

[15] Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, 53(3):464–501.

[16] Bignozzi, V. and Tsanakas, A. (2014). Model uncertainty in risk capital measurement. *Journal of Risk, Forthcoming*.

[17] Bignozzi, V. and Tsanakas, A. (2016). Parameter uncertainty and residual estimation risk. *Journal of Risk and Insurance*, 83(4):949–978.

[18] Blanchet, J., Lam, H., Tang, Q., and Yuan, Z. (2017). Applied robust performance analysis for actuarial applications. In *Technical Report*. the Society of Actuaries (SOA).

[19] Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784.

[20] Borgonovo, E., Hazen, G. B., Jose, V. R. R., and Plischke, E. (2021). Probabilistic sensitivity measures as information value. *European Journal of Operational Research*, 289(2):595–610.

[21] Borgonovo, E. and Plischke, E. (2016). Sensitivity analysis: a review of recent advances. *European Journal of Operational Research*, 248(3):869–887.

[22] Borowicz, J. M. and Norman, J. P. (2006). The effects of parameter uncertainty in the extreme event frequency-severity model. In *28th International Congress of Actuaries, Paris*, volume 28. Citeseer.

[23] Breuer, T. and Csiszár, I. (2013). Systematic stress tests with entropic plausibility constraints. *Journal of Banking & Finance*, 37(5):1552–1559.

[24] Cambou, M. and Filipović, D. (2017). Model uncertainty and scenario aggregation. *Mathematical Finance*, 27(2):534–567.

[25] CEIOPS (2010). Qis 5 technical specifications. *European Commission, Annex to Call for Advice from CEIOPS on QIS5*.

[26] Chan, W., Proschan, F., and Sethuraman, J. (1990). Convex-ordering among functions, with applications to reliability and mathematical statistics. *Lecture Notes-Monograph Series*, pages 121–134.

[27] Christopher Frey, H. and Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis*, 22(3):553–578.

[28] Cooke, R. M. and van Noortwijk, J. M. (1999). Local probabilistic sensitivity measures for comparing form and monte carlo calculations illustrated with dike ring reliability calculations. *Computer Physics Communications*, 117(1-2):86–98.

[29] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318.

[30] Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158.

[31] Delbaen, F., Grandits, P., Rheinländer, T., Samperi, D., Schweizer, M., and Stricker, C. (2002). Exponential hedging and entropic penalties. *Mathematical finance*, 12(2):99–123.

[32] Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons.

[33] Dey, S. and Juneja, S. (2010). Entropy approach to incorporate fat tailed constraints in financial models. *Available at SSRN 1647048*.

[34] Diers, D., Eling, M., and Linde, M. (2013). Modeling parameter risk in premium risk in multi-year internal models. *The Journal of Risk Finance*.

[35] EIOPA (2014). The underlying assumptions in the standard formula for the solvency capital requirement calculation. *The European Insurance and Occupational Pensions Authority*.

[36] EIOPA (2019). Methodological Principles of Insurance Stress Testing. Technical report, European Insurance and Occupational Pensions Authority.

[37] Eling, M., Schmeiser, H., and Schmit, J. T. (2007). The solvency ii process: Overview and critical analysis. *Risk management and insurance review*, 10(1):69–85.

[38] Floryszczak, A., Le Courtois, O., and Majri, M. (2016). Inside the solvency 2 black box: net asset values and solvency capital requirements with a least-squares monte-carlo approach. *Insurance: Mathematics and Economics*, 71:15–26.

[39] Föllmer, H. and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter.

[40] Fröhlich, A. and Weng, A. (2015). Modelling parameter uncertainty for risk capital calculation. *European Actuarial Journal*, 5(1):79–112.

[41] Georgescu, D. I., Higham, N. J., and Peters, G. W. (2018). Explicit solutions to correlation matrix completion problems, with an application to risk management and insurance. *Royal Society open science*, 5(3):172348.

[42] Gerrard, R. and Tsanakas, A. (2011). Failure probability under parameter uncertainty. *Risk Analysis: An International Journal*, 31(5):727–744.

[43] Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics*, 18(2):141–153.

[44] Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58.

[45] Goldfarb, D. and Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming*, 27(1):1–33.

[46] Goll, T. and Rüschendorf, L. (2001). Minimax and minimal distance martingale measures and their relationship to portfolio optimization. *Finance and Stochastics*, 5(4):557–581.

[47] Gupta, A. K. and Aziz, M. A. (2010). Convex ordering of random variables and its applications in econometrics and actuarial science. *European Journal of Pure and Applied Mathematics*, 3(5):779–785.

[48] Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2-3):327–367.

[49] Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17.

[50] Hong, L. J. (2009). Estimating quantile sensitivities. *Operations research*, 57(1):118–130.

[51] Hornberger, G. M. and Spear, R. C. (1981). Approach to the preliminary analysis of environmental systems. *J. Environ. Mgmt.*, 12(1):7–18.

[52] Kaas, R., Van Heerwaarden, A. E., and Goovaerts, M. J. (1994). Ordering of actuarial risks. *Institute for Actuarial Science and Econometrics, University of Amsterdam, Amsterdam.*

[53] Karlin, S. and Novikoff, A. (1963). Generalized convex inequalities. *Pacific Journal of Mathematics*, 13(4):1251–1279.

[54] King, A. J. and Korf, L. A. (2001). *Martingale pricing measures in incomplete markets via stochastic programming duality in the dual of* L^∞. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Institut für Mathematik.

[55] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

[56] Lam, H. (2019). Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research*, 67(4):1090–1105.

[57] Leamer, E. (1990). Let's take the con out of econometrics, and sensitivity analysis would help. *Modelling economic series*, 73:31–43.

[58] Liese, F. and Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412.

[59] Liu, H., Chen, W., and Sudjianto, A. (2006). Relative entropy based method for probabilistic sensitivity analysis in engineering design. *Journal of Mechanical Design*, 128(2):326–336.

[60] Lloyd's (2022). Standard formula for the solvency capital requirement (scr). https://www.lloyds.com/resources-and-services/capital-and-reserving/capital-guidance/standard-formula-scr/. Accessed: February 21, 2023.

[61] Luenberger, D. G., Ye, Y., et al. (1984). *Linear and Non-linear Programming*, volume 2. Springer.

[62] Makam, V. D., Millossovich, P., and Tsanakas, A. (2021). Sensitivity analysis with χ2-divergences. *Insurance: Mathematics and Economics*, 100:372–383.

[63] Mata, A. (2000). Parameter uncertainty for extreme value distributions. In *GIRO Convention Papers*, pages 151–173.

[64] McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press.

[65] Ohlsson, E. and Lauzeningks, J. (2009). The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, 45(2):203–208.

[66] Pesenti, S. M., Bettini, A., Millossovich, P., and Tsanakas, A. (2020). Scenario weights for importance measurement (swim)–an r package for sensitivity analysis. *Available at SSRN*.

[67] Pesenti, S. M., Millossovich, P., and Tsanakas, A. (2019). Reverse sensitivity testing: What does it take to break the model? *European Journal of Operational Research*, 274(2):654–670.

[68] Pesenti, S. M., Millossovich, P., and Tsanakas, A. (2021). Cascade sensitivity measures. *Risk Analysis*, 41(12):2392–2414.

[69] Pinheiro, P. J., Andrade e Silva, J. M., and de Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, 70(4):701–714.

[70] Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *The Review of Financial Studies*, 17(2):581–610.

[71] Rabitti, G. and Borgonovo, E. (2020). Is mortality or interest rate the most important risk in annuity models? a comparison of sensitivity analysis methods. *Insurance: Mathematics and Economics*, 95:48–58.

[72] Risk, J. and Ludkovski, M. (2016). Statistical emulators for pricing and hedging longevity risk products. *Insurance: Mathematics and Economics*, 68:45–60.

[73] Roch, O. and Valdez, E. A. (2011). Lower convex order bound approximations for sums of log-skew normal random variables. *Applied Stochastic Models in Business and Industry*, 27(5):487–502.

[74] Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.

[75] Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk analysis*, 22(3):579–590.

[76] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer physics communications*, 181(2):259–270.

[77] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.

[78] Saraswat, R. N. (2014). Chi square divergence measure and their bounds. In *3rd International Conference on "Innovative Approach in Applied Physical, Mathematical/Statistical"*, Chemical Sciences and Emerging Energy Technology for Sustainable Development, page 55.

[79] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer.

[80] Shannon, M. (2020). Properties of f-divergences and f-gan training. *arXiv preprint arXiv:2009.00757*.

[81] Sobol, I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical modelling and computational experiment*, 1:407–414.

[82] Tsanakas, A. and Millossovich, P. (2016). Sensitivity analysis using risk measures. *Risk Analysis*, 36(1):30–48.

[83] Tsanakas, A. and Smith, A. D. (2007). High dimensional modelling and simulation with asymmetric normal mixtures. *Available at SSRN 1005894*.

[84] Weber, S. (2007). Distribution-invariant risk measures, entropy, and large deviations. *Journal of Applied Probability*, 44(1):16–40.