Sensitivity analysis with χ^2-divergences

Vaishno Devi Makam* [1], Pietro Millossovich[1,2], and Andreas Tsanakas[1]

[1]The Business School (formerly Cass), City, University of London
[2]DEAMS, University of Trieste

January 7, 2021

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.

1 Introduction

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. (2015) 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

*Corresponding author. Email: vaishno.makam@cass.city.ac.uk.

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. Tsanakas and Millossovich, 2016). The complexity of quantitative risk models arises from the potential high-dimension and stochastic dependence of risk factors (e.g. Denuit et al., 2005; Arbenz et al., 2012), as well as the non-linearity of the aggregation function (e.g. Hong, 2009; Tsanakas and Millossovich, 2016), which may itself be numerically demanding in its evaluation at particular simulated scenarios (Risk and Ludkovski, 2016; Floryszczak et al., 2016).

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 (2002); Saltelli (2002); Saltelli et al. (2008); Borgonovo and Plischke (2016); Rabitti and Borgonovo (2020).

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 (Csiszár, 1967) 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 (Pesenti et al., 2019) offers a generalisation of the reverse stress testing approach often used in risk regulation (EIOPA, 2019). 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. 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 (Csiszár, 1975; Breuer and Csiszár, 2013). 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.

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 (2002); Saltelli (2002); Saltelli et al. (2008); Borgonovo and Plischke (2016). Major advances in sensitivity analysis can be accredited to Sobol (1993); Homma and Saltelli (1996); Saltelli et al. (2008). The range of sensitivity analysis methods available in the literature is substantial, with variance-based (Saltelli et al., 2008, 2010) and moment-independent methods (Borgonovo, 2007) 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 (Rabitti and Borgonovo, 2020). 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 (Liu et al., 2006). Efforts towards overcoming this shortcoming include the use of conditional Kullback-Leibler divergences, in order to quantify the importance of a model input (Auder and Iooss, 2008; Liu et al., 2006).

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. (2019), 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 (2013), who use Kullback-Leibler divergence in the context of model uncertainty. Model uncertainty is also addressed in Glasserman and Xu (2014), 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.

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 (2014), who use the α-divergence (of which the χ^2-divergence is a special case), when distributions are heavy-tailed. Similarly, Dey and Juneja (2010) minimise a related divergence measure under linear constraints in a portfolio selection problem.

1.3 Structure of the paper

The rest of the paper is organised as follows. In Section 2, we discuss the Kullback-Leibler and χ^2-divergences. In Section 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 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 5.

2 Preliminaries

Let P and Q be two probability measures defined on a common measurable space $(\Omega, \mathcal{A})$. $Q \ll P$ 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 (Ali and Silvey, 1966; Liese and Vajda, 2006; Cambou and Filipović, 2017), as measures of discrepancy between two probability measures.

Definition 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 (1951), and the χ^2-divergence (Csiszár, 1967; Liese and Vajda, 2006) are two special cases corresponding to $f(u) = u \log u$ and $f(u) = u^2 - 1$, respectively.

Definition 2.2. The KL-divergence of Q with respect to $\mathbb{P}$ 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 2.3. 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 (2014) 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 $\mathcal{A} = 2^\Omega$. 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:

$$D_{KL}(Q||P) = \sum_i q_i \log\left(\frac{q_i}{p_i}\right) = \sum_i p_i w_i \log w_i$$

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

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. (1999), Föllmer and Schied (2011)). We focus on the percentile-based risk measures Value-at-Risk (VaR) and Expected Shortfall (ES) (McNeil et al., 2015).

Definition 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 2.5. 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 (Rockafellar et al., 2000). Expected Shortfall is a coherent risk measure, whereas the Value-at-Risk in general is not (Artzner et al., 1999).

3 Stress testing models

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 $(\Omega, \mathcal{A})$ denote the random variables representing the input factors of the model under consideration. The aggregation function $g : \mathbb{R}^n \to \mathbb{R}$, when applied on the inputs, gives a one-dimensional output $Y = g(\mathbf{Z})$. 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 $\mathcal{P}$ denote the set of all probability measures on the measurable space $(\Omega, \mathcal{A})$ and, for a given $P \in \mathcal{P}$, 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 $Q \in \mathcal{P}$ 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 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 (2013). Here, we define the problem:

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

Remark 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 (2013) and Glasserman and Xu (2014). 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 (Csiszár, 1975; Breuer and Csiszár, 2013)

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

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 (2010) 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 4.4.

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.

Proposition 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 $i = 1, \ldots, n$,

   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

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

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

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

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

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

Proposition 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 $\theta^* = \dfrac{1}{2} \sum_{i=1}^n p_i w_i^{*2}$. Then, $\mathbf{w}^*$ solves Problem (III) with $\theta = \theta^*$.

Finally, from the proof of Proposition 3.2 it can be seen that the χ^2-divergence constraint in Problem (III) is always binding at the optimum.

Remark 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 (Luenberger et al., 1984).

Remark 3.3. The increasingness of the optimal Radon-Nikodym derivative in Proposition 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., 2019, Prop. A.1). As a result the expectation of *any* increasing function of X is stressed upwards.

Remark 3.4. In Proposition 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:

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

and where $r_1 = -x_n, r_2 = -x_{n-1}, \ldots, r_n = -x_1$. This problem can be solved once again using Proposition 3.1, since $\bar{r} = -\bar{x} < -t \leq r_n = -x_1$.

Remark 3.5. In a Monte Carlo model where the states of the world are assumed to be equiprobable, the optimization Problem (II) simplifies to

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

The solution of Problem (II), as reported in Proposition 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 $\widetilde{n} = n - m$,

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

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

Remark 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 optimization problem

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

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

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

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 $\sum_{i=1}^n p_i v_i^2$.

3.3 Reverse and forward sensitivity analyses

Here we return to the problem definition of Section 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. (2019)), 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 represents a stress on the expected value of the output Y, and let $\frac{dQ_Y}{dP}$ the optimal Radon-Nikodym derivative. 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 $\frac{dQ_{Z_i}}{dP}$, $i = 1, \ldots, d$. 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 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 3.3 below. We remark that the sensitivity analysis framework we present, including Definition 3.1, can be altered as necessary to include functions of input factors to enable the assessment of different distributional characteristics.

Proposition 3.3. The sensitivity measures of Definition 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.

4 Case study of an insurance portfolio

Here we apply the framework of Section 3 to the example of a simplified insurance portfolio. In Section 4.1, we introduce the model, while in Sections 4.2 and 4.3 we, respectively, perform reverse and forward sensitivity analyses. Finally, in Section 4.4, we evaluate the sensitivity measures of Definition 3.1; furthermore, we examine their sampling performance, comparing them to similar measures that are constructed by replacing the χ^2- with Kullback-Leibler divergence.

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.

4.2 Reverse sensitivity analysis of the insurance model

Using the above model, we follow the sensitivity analysis process outlined in Section 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 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 1, the empirical distributions of Y under the baseline (black) and stressed (red) measure are shown. The stressed output distribution first-order stochastically dominates the output distribution under the baseline model, as remarked after Proposition 3.1.

Figure 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 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.

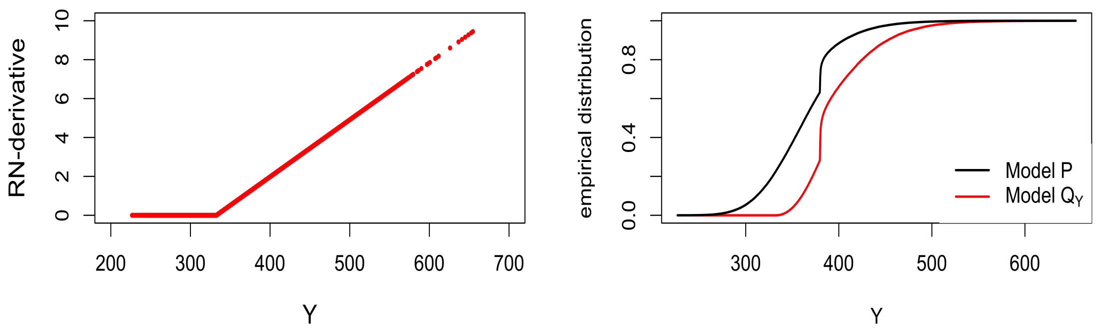

Figure 1: Left: Radon-Nikodym derivative of Q_Y against Y. Right: Stressed probability distributions of Y under models P, Q_Y.

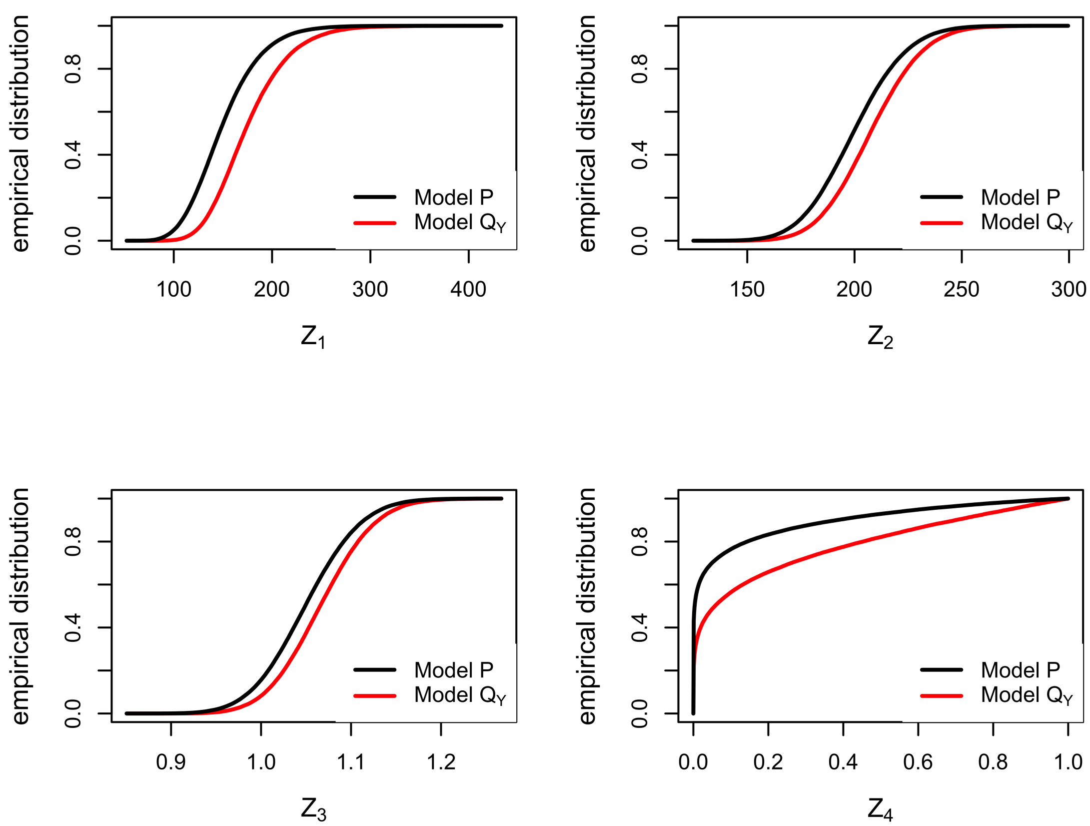

Figure 2: Empirical distributions of the input factors under the baseline and stressed models P, Q_Y.

Table 1: Percentage increase in statistics of input factors Z_i under the stressed model Q_Y, with respect to the baseline model P.

Input factors	Mean	St. Dev.	$\mathrm{VaR}_{0.95}$	$\mathrm{ES}_{0.95}$
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

4.3 Forward sensitivity analysis of the insurance model

Now we carry out forward sensitivity analysis, as discussed in Section 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 4.2. Figure 3 displays the Radon-Nikodym derivative of the stressed probability measures $Q_{Z_i},\ i = 1, \ldots, 4$. 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.

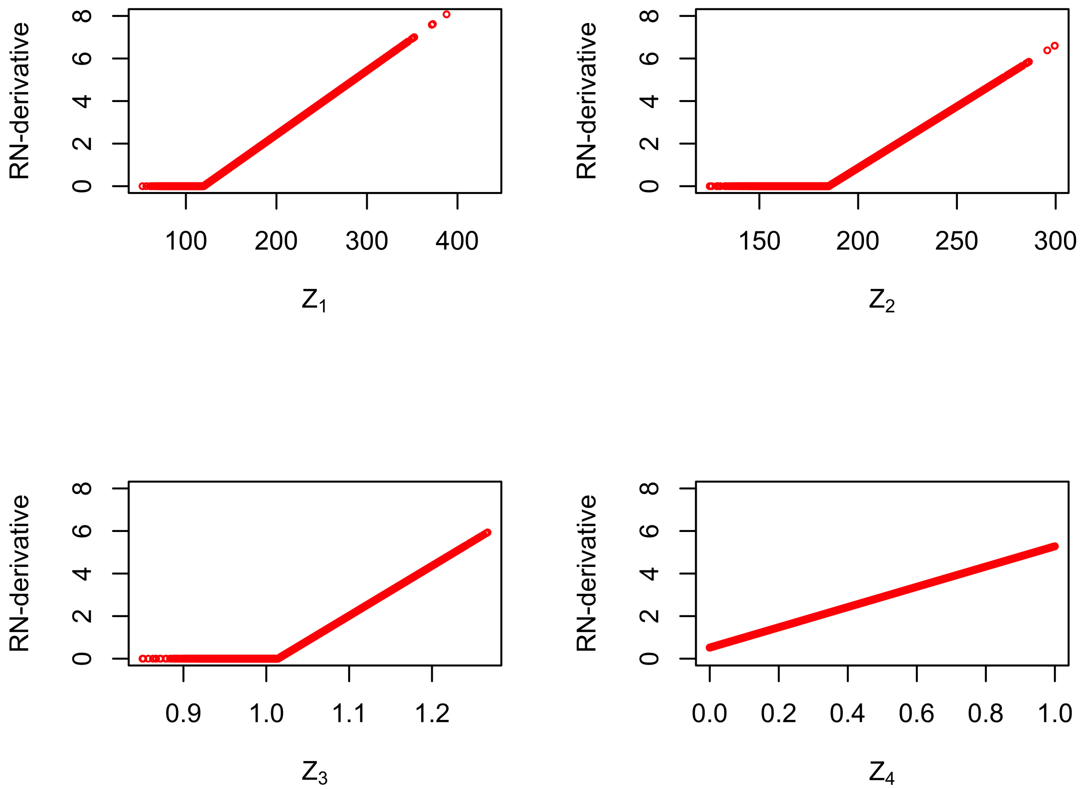

Figure 3: Radon-Nikodym derivatives of stressed models Q_{Z_i} for $i = 1, \ldots, 4$

.

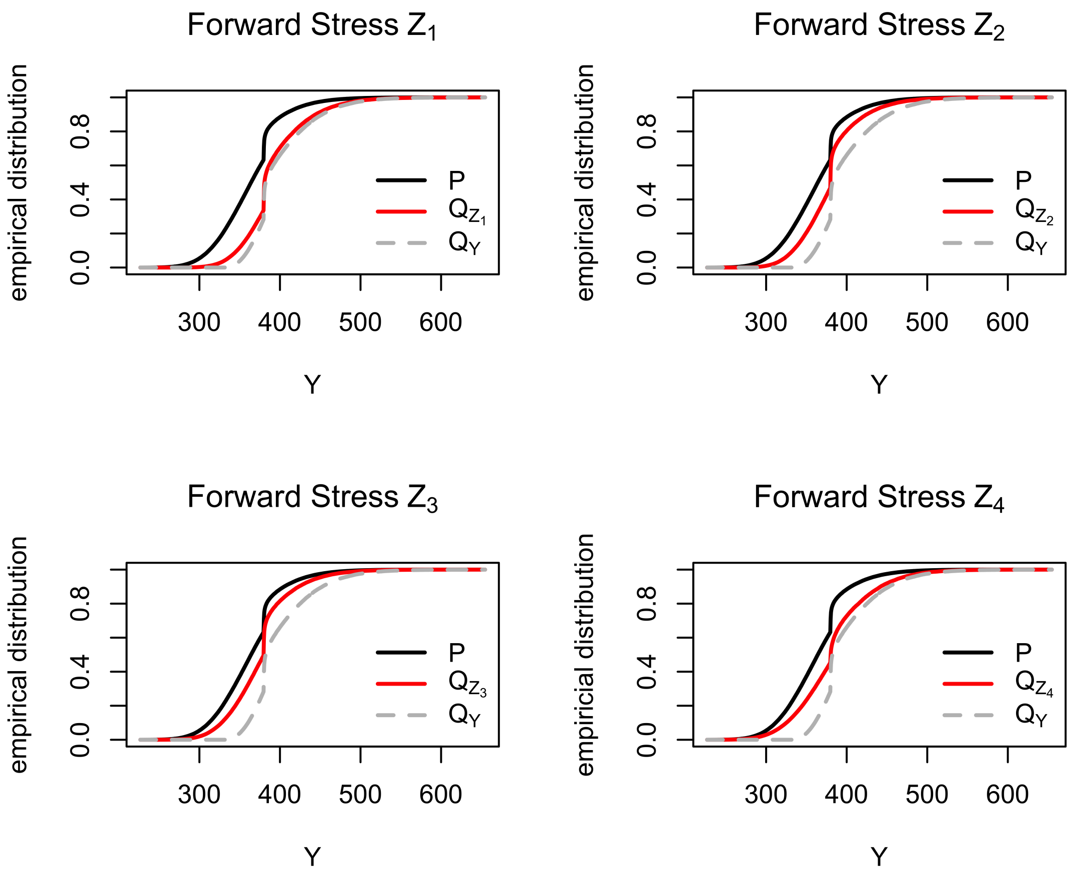

Figure 4: Stressed probability distributions of Y under the baseline model P and the stressed models Q_Y and Q_{Z_i} for $i = 1, \ldots, 4$.

In Figure 4, the empirical distributions of Y under the baseline (P, black) and all stressed (Q_{Z_i}, red; Q_Y, dashed grey) 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 4.2. This is evident from Figure 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 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 4.2. In Table 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.

Table 2: Percentage increase in statistics of Y under the stressed models Q_{Z_i} for $i = 1, \dots, 4$, with respect to the baseline model P.

Variables stressed	Mean	St. Dev.	$\mathrm{VaR}_{0.95}$	$\mathrm{ES}_{0.95}$
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

4.4 Evaluation of sensitivity measures

The aim of this section is to evaluate the sensitivity measures defined in Section 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 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 4.2 and 4.3.

Furthermore, for comparison purposes, in the third and fifth column of Table 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 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. Breuer and Csiszár (2013) and the numerical implementation is carried out via the **R** package SWIM by Pesenti et al. (2020). We observe that a change in the divergence measure does not impact the relative importance of input factors.

Table 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 $t^{(k)} = 1.1\dfrac{1}{n}\sum_{j=1}^{n} y_j^{(k)}$ 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, $\theta^{(k)} = \dfrac{1}{n}\sum_{j=1}^{n}(\mathbf{w}_j^{(k)})^2$.
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 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 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 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 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.

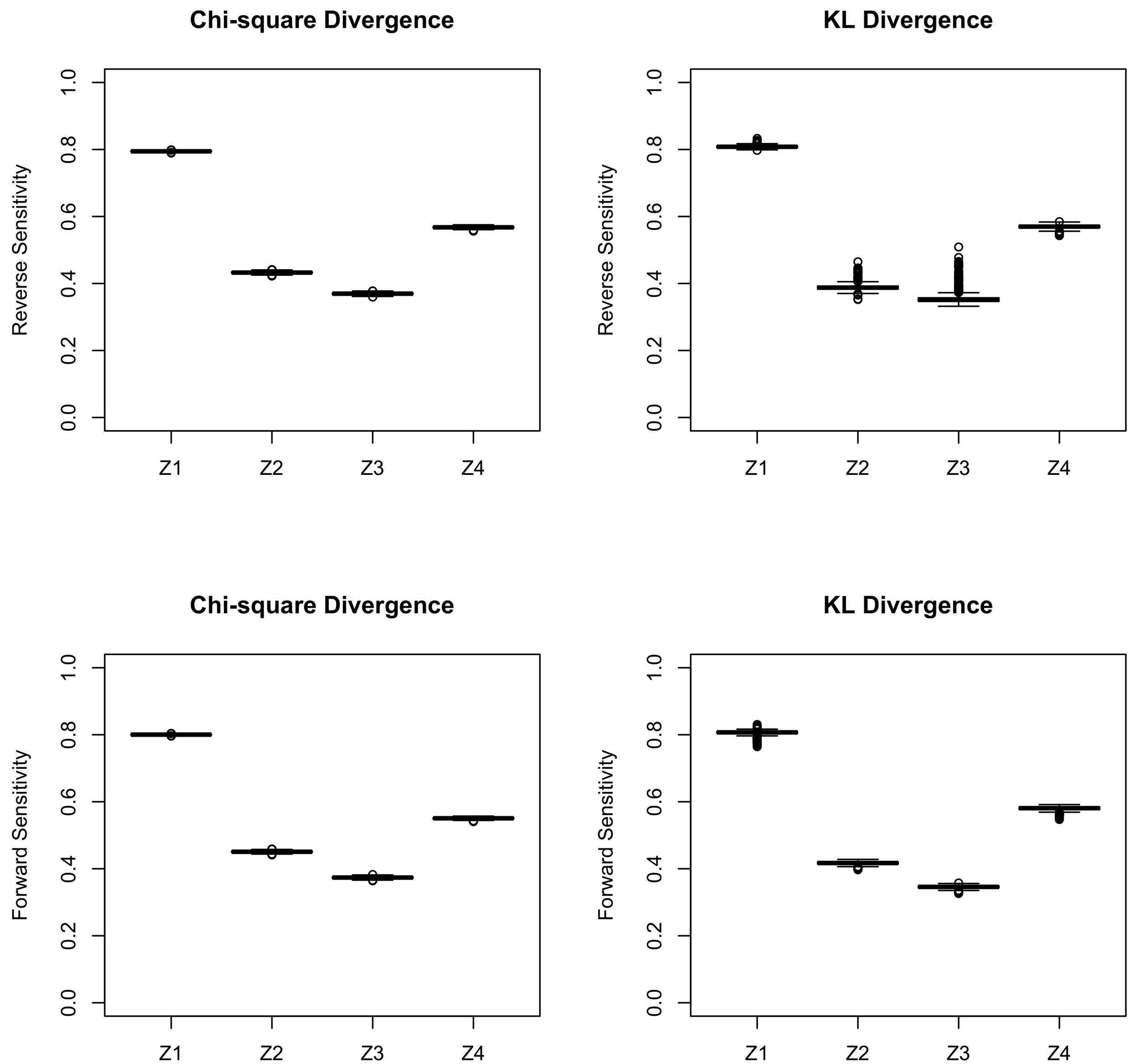

Figure 5: Box plots of reverse and forward sensitivities of input factors under χ^2- and Kullback-Leibler divergences, for $m = 1000$ sets of $n = 10^5$ simulated scenarios.

In Table 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 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

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.

A Proofs

Proposition (3.1). To solve Problem (II), we first define some notation. 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:

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

For any integer $j \in \{1, \ldots, n-2\}$, 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:

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

If $t \leq x_n$, (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 $\mathbf{w}$ (Luenberger et al., 1984).

The KKT conditions are for $i = 1, \ldots, n$:

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

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 $t < \bar{x} + \dfrac{s^2}{\bar{x} - x_1}$.

Case $1 < k^* < n$:

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

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

Therefore,

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

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:

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

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.

$k^* = n$:
If $k^* = n$, we get $w_n = n$ and $w_i = 0$ for $i = 1, \ldots, n-1$. In such a case, it is clear from the second constraint of Problem II that $t = x_n$. □

Proposition 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 $\eta_2^* = \dfrac{-1}{\lambda_2}, \eta_1^* = \dfrac{\lambda_1}{\lambda_2}$ and $\epsilon_i^* = \dfrac{\mu_i}{\lambda_2}$, where λ_1, λ_2 and μ_i are the Lagrangian multipliers in Problem (II).

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

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

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

Proposition (3.3).
1. The denominator of F_i is strictly positive, by assumption. The denominator of R_i is strictly positive by Proposition 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 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 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.

$\square$

References

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.

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

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

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.

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

Liu, H., Chen, W., and Sudjianto, A. (2006). Relative entropy based method for probabilistic sensitivity analysis in engineering design.

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

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

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.*

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.

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.

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

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

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

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.

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.

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.

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

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