Sensitivity analysis using risk measures*

Andreas Tsanakas [†1] and Pietro Millossovich[‡1]

[1]Faculty of Actuarial Science and Insurance, Cass Business School, City University London

*Accepted for publication in *Risk Analysis: An International Journal*. We thank two referees for constructive feedback. Earlier versions have been presented at the 17th *International Congress on Insurance: Mathematics and Economics*, the 6^{th} *International Conference of the ERCIM WG on Computational and Methodological Statistics*, the 6^{th} *International Conference on Mathematical and Statistical Methods for Actuarial Sciences and Finance*, the *FISC 2014 Conference*, and at research seminars in the Department of Mathematics, ETH Zurich, the Faculty of Economics, KU Leuven, and the Department of Mathematics and Statistics, University of Montreal. We thank Jens Perch Nielsen, María Dolores Miranda-Martínez and Mario Wüthrich for helpful discussions.

[†]andreas.tsanakas.1@city.ac.uk.

[‡]pietro.millossovich.1@city.ac.uk.

Abstract: In a quantitative model with uncertain inputs, the uncertainty of the output can be summarized by a risk measure. We propose a sensitivity analysis method based on derivatives of the output risk measure, in the direction of model inputs. This produces a global sensitivity measure, explicitly linking sensitivity and uncertainty analyses. We focus on the case of distortion risk measures, defined as weighted averages of output percentiles, and prove a representation of the sensitivity measure that can be evaluated on a Monte-Carlo sample, as a weighted average of gradients over the input space. When the analytical model is unknown or hard to work with, non-parametric techniques are used for gradient estimation. This process is demonstrated through the example of a non-linear insurance loss model. Furthermore, the proposed framework is extended in order to measure sensitivity to constant model parameters, uncertain statistical parameters, and random factors driving dependence between model inputs.

Keywords: Sensitivity analysis, risk measures, uncertainty analysis, aggregation, parameter uncertainty, dependence.

1 INTRODUCTION

Sensitivity analysis, alongside uncertainty analysis, is an activity of fundamental importance to risk analysts, especially in the presence of complex computational models with uncertain inputs. Reasons for performing sensitivity analyses include the need to rank the relative importance of input variables, to uncover relations between inputs and outputs in complex models, to perform quality assurance and reasonableness checks, and to identify areas of investigation and refinement in the model[1–5]. Sensitivity analysis comprises a wide range of methodologies, with *global sensitivity* approaches, reflecting the model behaviour over the whole of the input range[3,5,6], becoming the norm in recent years. Such methods have seen wide application in risk analysis; indicatively we mention safety assessment for nuclear waste disposal[7], food safety[2], microbial risk assessment[8], and climate change modelling[9].

Global sensitivity approaches are concerned with "how uncertainty in the output of a model [...] can be apportioned to different sources of uncertainty in the model input"[5]. In moment-based global sensitivity analyses[3,5], the variance of outputs is typically apportioned to contributions of input variable volatilities. While such analyses are revealing, it is known that using the variance as a measure of variability can lead to violations of standard dominance requirements[10]. Furthermore, application of variance-based sensitivity methods becomes involved when inputs are correlated. Such considerations have motivated the design of moment-free sensitivity analyses[6,11–13], which reflect the difference between the conditional and unconditional output density via a distance measure.

We consider models where a number of uncertain *model inputs* are mapped to a single *output* of interest via a known *model function*. Both inputs and outputs are observable. Uncertainty analysis allows determination of the output's probability distribution, e.g. via Monte-Carlo simulation. A summary of the output distribution can be produced using a *risk measure*, that is,

a statistical functional that maps a random variable to a real number[14,15]. Moments such as the mean and variance are examples of risk measures. A flexible and widely applicable class are *distortion risk measures*[14,16–18], arising as weighted averages of percentiles. The weighting function encodes preferences, in the sense that particular areas of the output probability distribution, corresponding to states of interest, can be focused on. The class of distortion risk measures includes most *coherent risk measures*[19] of interest. The sensitivity of optimization problems that involve such risk measures is studied in Ref.[20].

We propose a sensitivity analysis method that directly links to uncertainty analysis through the use of risk measures. Specifically, we argue in favor of measuring sensitivity by a *directional derivative of the risk measure applied on the output, in the direction of model inputs*. The resulting sensitivity measure satisfies the desirable properties of Saltelli[3]. In particular, it is global, in the sense that it is generated by the weighted average of local sensitivities over the whole of the input space. The proposed method is general enough to cope with measurement of sensitivity to various types of variables, such as (i) uncertain model inputs; (ii) statistical parameters of the distributions of inputs that are subject to uncertainty; (iii) known structural parameters; (iv) common factors driving dependence between inputs. The calculation of sensitivities can be carried out on a single Monte-Carlo sample of model inputs and outputs – computational speed can be further improved for high-dimensional models if information on the hierarchical (multi-level) construction of the model functions is available. While the method is in principle applicable to any type of risk measure (including moment-based ones), we focus on the case where the output distribution is summarized by a distortion risk measure.

In Section 2, the sensitivity to each model input is defined as the derivative of the risk measure applied on the output, with respect to the scale of a random shock applied to that model input. The sensitivity thus is a directional derivative, which makes the sensitivity measure a generalization of performance measurement methods used in finance and actuarial science[21–23]

to non-linear models. For distortion risk measures, an explicit formula for the sensitivity measure is derived, extending existing results[24,25]. The resulting sensitivity is expressed as an expectation involving the random shock, the corresponding gradient of the model function, and scenario weights generated by the risk measure. Thus the sensitivity measure reflects information about the statistical behaviour of the model inputs, the shape of the model function, and preferences. There is flexibility in relation to the shock chosen and we particularly focus on shocks that produce proportional changes to model inputs.

The formal properties of the sensitivity measure are discussed in Section 2.3, where its relation to other sensitivity analysis approaches is also discussed. Conceptual parallels are identified between the properties our sensitivity measure and the concerns of standard methods such as Regional Sensitivity Analysis[26], elementary effects[27], variance-based sensitivity[3,5] and moment-free sensitivity[6].

In Section 3, the implementation and communication of the sensitivity analysis is demonstrated, using the example of a non-linear insurance loss model. The analysis is performed on a Monte-Carlo sample of the portfolios. We do not calculate the gradient of the model function analytically, but instead put ourselves in the shoes of an analyst or reviewer who has access to the simulated samples of the model inputs and output only, but no direct access to the model's analytical form. This corresponds to practical situations, where there is limited access to model documentation and/or the model is too complicated to make analytical calculations appealing. Then, the gradients of the model are estimated numerically from the simulated sample, borrowing tools from non-parametric statistics, in particular local linear regression[28,29].

In Section 4, the proposed methods are extended to the measurement of the sensitivity to parameters and the dependence structure. For this purpose, known model parameters, treated as degenerate model inputs, are artificially randomized. Sensitivity to statistical parameters is considered in a context where the distribution of an model input is modelled as a weighted

average of distributions (a mixture) over possible parameter values. We express an model input subject to parameter uncertainty as a function of two independent random variables, representing stochastic and parameter uncertainty respectively. The sensitivity to each of the two elements is then calculated.

Correlation and, more broadly, dependence between inputs is understood as a major driver of output variability[12], see also Ref.[30] for an extensive treatment. Many commonly used dependence structures can be expressed via conditional independence, given a number of latent or observable common factors. We proceed by measuring sensitivity to those common factors. This entails expressing the vector of dependent model inputs as a function of a vector of independent idiosyncratic and common factors; we present a probability transform that enables such a decomposition.

Brief conclusions are stated in Section 5; formal statements of results and proofs are given in the Appendix.

2 SENSITIVITY ANALYSIS

2.1 Preliminaries

For the purposes of this paper, a model consists of three elements: a vector $\mathbf{X}$ of inputs, a function g, and an output variable $Y = g(\mathbf{X})$. We call the elements of $\mathbf{X}$ *model inputs*. Model inputs are uncertain; they can be subject to both epistemic uncertainty and stochastic variability. Uncertainty is modelled by considering $\mathbf{X}$ a random vector taking values in an open convex set $\mathcal{X} \subset \mathbb{R}^d$ and with joint probability distribution F. We call $g : \mathcal{X} \to \mathbb{R}$ the *model function* and Y the *output*. For simplicity, we will typically assume that that high values of Y correspond to undesirable states of the world, for instance in the case that Y stands for a financial loss.

The probability distribution of the output is denoted F_Y. For simplicity of presentation, we

assume here and in the sequel that distributions are strictly increasing and continuous, though the results in this paper do not require such assumptions. Then, for $u \in (0,1)$, the inverse $F_Y^{-1}(u)$ gives the $100u^{th}$ percentile of Y. We call *output rank* the uniformly distributed random variable $U_Y = F_Y(Y)$, such that $Y = F_Y^{-1}(U_Y)$.[1]

A risk measure is a functional ρ mapping random variables to the real line. For example, $\rho(Y) = \mathbb{E}(Y)$ and $\rho(Y) = \mathbb{V}(Y)$, the mean and the variance, are risk measures. Here we consider the class of *distortion risk measures*, defined as weighted averages of percentiles. Consider a non-negative *weight function* ζ on $[0,1]$ with $\int_0^1 \zeta(u)du = 1$. The function ζ reflects preferences, in the sense that it allows assigning a higher weight to output ranks corresponding to (generally undesirable) states of the world that are of specific interest. The distortion risk measure, as applied to the random variable Y, is given by the relation

$$\rho(Y) = \int_0^1 F_Y^{-1}(u)\zeta(u)du = \mathbb{E}\big(Y\zeta(U_Y)\big). \tag{2.1}$$

To distinguish the risk measure ρ as an abstract functional, from the value $\rho(Y)$, we call the latter the *risk summary* of Y. The right-hand-side of equation (2.1) shows that the risk summary of Y may also be represented as its expected value, subject to a re-weighting of probabilities effected by ζ (note that $\mathbb{E}(\zeta(U_Y)) = 1$). Distortion risk measures were introduced in the actuarial literature by Wang et al.[16]. Such risk measures have found wide application in insurance pricing, financial risk management and risk analysis[14,17,18].

This class of risk measures is fairly flexible and admits commonly used measures as special cases. The constant weight function $\zeta(u) = 1$ gives the mean. The $100p^{th}$ percentile, $F_Y^{-1}(p)$, is derived for a trivial weight function $\zeta(u)$, with a mass concentrated at p, such that a weight of 1 is

[1]For a general distribution F_Y, the definitions are easily extended. The generalized inverse $F_Y^{-1}(u) = \inf\{y : \mathbb{P}(Y \leq y) \geq u\}$ is defined. There still exists a random variable U_Y such that $Y = F_Y^{-1}(U_Y)$, but it is no longer unique. A simple explicit construction of the output rank U_Y is given by the generalized distributional transform[31].

attached to $F_Y^{-1}(p)$ and a weight of zero to all other percentiles. In finance, $F_Y^{-1}(p)$ corresponds to the Value-at-Risk measure, denoted by VaR_p. Let $\mathbf{1}_A$ be the indicator function of an event (or set) A, taking value 1 when the event is true and zero otherwise. Then, $\zeta(u) = \frac{1}{1-p}\mathbf{1}_{\{u>p\}}$, applying a constant weight to percentiles more extreme than the $100p^{th}$, gives rise to Tail-Value-at-Risk (TVaR), $\text{TVaR}_p(Y) = \frac{1}{1-p}\int_p^1 F_Y^{-1}(u)du = \mathbb{E}(Y|Y > F_Y^{-1}(p))$, with the second equality holding when F_Y is continuous. The VaR and TVaR risk measures are extensively used in financial risk management[32], while other weighting functions considering the whole of the output distribution are available[18,30]. When the function ζ is non-decreasing (as in the case of TVaR), the corresponding risk measure ρ is a convex functional, thus satisfying the axioms of *coherent risk measures*[19]. For coherent risk measures, inconsistent orderings of probability distributions, such as those identified by Cox[10], are avoided.

It is not obvious how to best choose a risk measure. Sometimes, particular risk measures are prescribed by regulation. For example, in an insurance context, capital requirements under the European Solvency II regime are calculated by $\text{VaR}_{0.995}$[33], while under the Swiss Solvency Test a $\text{TVaR}_{0.99}$ measure is used[34]. The choice of these measures is dictated by their regulatory use, as they both are simple and focus on extreme adverse events. For internal use, which is related to sensitivity analysis, one need not restrict to those or indeed to any single risk measure. The choice of risk measure will be discussed further through Examples 1 and 2.

2.2 Sensitivity measures

We measure the sensitivity of the output risk summary to an individual model input, by stressing that model input with a random shock. Specifically, consider a random variable Z and $\epsilon > 0$. Denote by $\mathbf{e_i}$ the vector in $\mathbb{R}^d$ with 1 in the i^{th} position and 0 elsewhere. Then, the vector of model inputs with a shock ϵZ applied to the i^{th} element is written as $\mathbf{X} + \epsilon Z\mathbf{e_i}$. Additionally, we denote the $(d-1)$-dimensional random vector containing all elements of $\mathbf{X}$ except X_i as $\mathbf{X}_{-i}$.

The application of the shock ϵZ corresponds to a change in the distribution of the random variable X_i (and more generally the random vector $\mathbf{X}$), effected by adding to it another source of uncertainty. The selection of Z depends on the purpose of the sensitivity analysis. If one wants to study for example the impact of increasing the variability of X_i, then Z can be chosen to have a mean of zero. It will be seen below that it is possible without loss of generality to restrict attention to shocks Z that are functions of the random vector $\mathbf{X}$, such that Z and X_i will generally be dependent random variables. We note that this way of changing the distribution of the inputs, by perturbing the random inputs directly, is distinct from a perturbation of the probability distribution itself as practised in Bayesian[35] and robust[36] statistics.

The sensitivity to the shock Z is now defined as the derivative of the risk summary of the stressed output $\rho(g(\mathbf{X} + \epsilon Z\mathbf{e_i}))$ with respect to ϵ, evaluated at $\epsilon = 0$.[2] Hence, it is a directional derivative in the direction of Z. Such derivatives have been considered in finance, where generally only linear model functions are considered[21–23], while non-linear models are only studied in the case of VaR[25].

Definition 1. *Let the function $\epsilon \mapsto \rho\big(g(\mathbf{X} + \epsilon Z\mathbf{e_i})\big)$ be differentiable at $\epsilon = 0$. Then the sensitivity of $\rho\big(g(\mathbf{X})\big)$ to the shock Z applied on X_i is defined as*

$$S_i(Z; \mathbf{X}, g) = \frac{\partial}{\partial \epsilon} \rho\big(g(\mathbf{X} + \epsilon Z\mathbf{e_i})\big)\bigg|_{\epsilon=0}.$$

For $Z = X_i$, we denote $S_i(\mathbf{X}, g) \equiv S_i(X_i; \mathbf{X}, g)$ and call this the sensitivity of $\rho\big(g(\mathbf{X})\big)$ to X_i.

In general it is not straightforward to obtain $S_i(Z; \mathbf{X}, g)$ by direct differentiation and numerical evaluation may be challenging for complex models. The following result gives a formula that is generally be easier to evaluate and also provides additional insight into the definition of sensitivity introduced.

[2]The support of $\mathbf{X}$ is a subset of $\mathcal{X}$ and throughout the paper we assume that evaluations of g at points not belonging to the support of $\mathbf{X}$ are meaningful.

Proposition 2. *Assume that*

i) g is differentiable with $\nabla g = (g_1, \ldots, g_d)$;

ii) g is invertible in at least one argument, say the k^{th}, and X_k has a conditional density given $\mathbf{X}_{-k}$;

iii) the function $\epsilon \mapsto F^{-1}_{g(\mathbf{X}+\epsilon Z\mathbf{e_i})}(u)$ is differentiable at $\epsilon = 0$ for all $u \in (0,1)$.

Then

$$S_i(Z;\mathbf{X},g) = \mathbb{E}\big(Z g_i(\mathbf{X})\zeta(U_Y)\big)$$

$$S_i(\mathbf{X},g) = \mathbb{E}\big(X_i g_i(\mathbf{X})\zeta(U_Y)\big).$$

The proof of Proposition 2, its relation to results already in the literature, and a discussion of its assumptions are given in Appendix A. Proposition 2 shows that the sensitivity $S_i(Z;\mathbf{X},g)$ can be viewed as the expected value of the shock Z, multiplied by the derivative of the model function in the i^{th} element (evaluated at random points $\mathbf{X}$) and the weights $\zeta(U_Y)$. Taking the expectation reflects the scale of the shock Z; multiplying by the derivative considers the (local) sensitivity of the model function to such shock at each state of the world; using the weight function places emphasis on those scenarios that are of most interest for the calculation of the the output risk summary. Thus, the sensitivity appropriately reflects the joint distribution of $(Z,\mathbf{X})$, the properties of g, and the preferences encoded in ζ. By considering the derivative g_i at random points $\mathbf{X}$ and placing emphasis on regions of Y that are of interest, a *global* sensitivity measure is obtained. Thus the first three desirable properties of sensitivity analysis methods specified by Saltelli[3] are satisfied.

There is flexibility in the potential choice of Z. The particular choice $Z = X_i$, on which we generally focus, effects a proportional stress on the i^{th} model input. When Z is statistically independent of $\mathbf{X}$ one obtains $S_i(Z;\mathbf{X},g) = \mathbb{E}(Z)\mathbb{E}\big(g_i(\mathbf{X})\zeta(U_Y)\big) = S_i(\mathbb{E}(Z);\mathbf{X},g)$, thus

the variability of independent shocks is of no consequence, with only their mean driving the corresponding sensitivity. This observation motivates the choice of a shock Z that is dependent on the vector of model inputs $\mathbf{X}$, $Z = X_i$ being one such case. More generally, for any choice Z, application of the tower property reveals that $S_i(Z; \mathbf{X}, g) = S_i(\mathbb{E}(Z|\mathbf{X}); \mathbf{X}, g)$, showing that one may meaningfully restrict attention to shocks Z that are functions of $\mathbf{X}$.

One can perform the decomposition

$$S_i(Z; \mathbf{X}, g) = S_i^{\mathcal{M}}(Z; \mathbf{X}, g) + S_i^{\mathcal{D}}(Z; \mathbf{X}, g), \tag{2.2}$$

where

$$S_i^{\mathcal{M}}(Z; \mathbf{X}, g) = \mathbb{E}\big(\mathbb{E}(Z) g_i(\mathbf{X}) \zeta(U_Y)\big) = S_i(\mathbb{E}(Z); \mathbf{X}, g) \tag{2.3}$$

$$S_i^{\mathcal{D}}(Z; \mathbf{X}, g) = \mathbb{E}\big((Z - \mathbb{E}(Z)) g_i(\mathbf{X}) \zeta(U_Y)\big) = S_i(Z - \mathbb{E}(Z); \mathbf{X}, g). \tag{2.4}$$

Hence the sensitivity to the shock Z can be viewed as the sum of two distinct effects: a *mean* element $S_i^{\mathcal{M}}(Z; \mathbf{X}, g)$, reflecting sensitivity to a non-random shock $\mathbb{E}(Z)$, and a *deviation* element $S_i^{\mathcal{D}}(Z; \mathbf{X}, g)$, reflecting sensitivity to a zero-mean shock $Z - \mathbb{E}(Z)$. Considering such a breakdown can be useful for further understanding the nature of the reported sensitivity $S_i(Z; \mathbf{X}, g)$; for example when $Z = X_i$, for a high mean / low variability model input X_i one would expect the mean element to dominate, while in a low mean / high variability example the deviation element would be more important. In particular, if Z is independent of $\mathbf{X}$ (with Z constant a special case), it is $S_i^{\mathcal{D}}(Z; \mathbf{X}, g) = 0$.

Finally, we note that the proposed sensitivity measure can also be calculated on moment-based risk measures. For example, from elementary calculations, we have for the mean and standard deviation:

$$\rho(Y) = \mathbb{E}(Y) \implies S_i(Z; \mathbf{X}, g) = \mathbb{E}(Z g_i(\mathbf{X}))$$

$$\rho(Y) = \sqrt{\mathbb{V}(Y)} \implies S_i(Z; \mathbf{X}, g) = \mathbb{C}\left(Z g_i(\mathbf{X}), Y\right) / \sqrt{\mathbb{V}(Y)},$$

where $\mathbb{C}$ is the covariance operator. Moment-based risk measures are not further considered in this paper.

2.3 Properties and relation to the literature

While the sensitivity measure proposed is different from other measures in the risk analysis literature[3,5,11], at the conceptual level several parallels emerge. First note that, though defined by a partial derivative, this is a global rather than a local sensitivity measure, since it is defined as the directional derivative of a statistical functional (risk measure) and not a derivative of the model function g at a given point. This is reflected in Proposition 2, showing that that evaluation of the gradient of g is necessary *for all* possible states of of the input vector $\mathbf{X}$. Considering derivatives of g across the range of $\mathbf{X}$ is something already encountered in the Morris method of Elementary Effects[5,27,37], however here we consider a scenario-weighted average that is explicitly linked to the risk measure used in the uncertainty analysis.

The sensitivity measure, while global, is still based on differentiation, a characteristic shared with many local sensitivity analysis methods. Indeed, if one is primarily interested in changes to $\rho(g(\mathbf{X}+\epsilon Z\mathbf{e}_i)$ (as opposed to $g(\mathbf{X}+\epsilon Z\mathbf{e}_i)$), the problem can be recast as a local sensitivity analysis with the risk summary as the output and ϵ as the input. In that case the arguments for (and against) local differential methods apply, as outlined in Ref.[38] Sec. 2.5.

The sensitivity $S_i(Z;\mathbf{X},g)$ depends on the weighting function ζ, a dependence suppressed in the notation. A way of expressing sensitivities without reference to ζ follows from considering the function

$$s_i(u;Z;\mathbf{X},g) = \mathbb{E}\big(Zg_i(\mathbf{X})|U_Y = u\big). \tag{2.5}$$

The quantity $s_i(u;Z;\mathbf{X},g)$ is the sensitivity of the output percentile $F_Y^{-1}(u)$ to the shock Z applied on the i^{th} model input; hence it provides a sensitivity measure specific to a particular confidence level of the distribution of the output Y. The sensitivity for a more general distortion

risk measure can be simply recovered by averaging over u,

$$S_i(Z;\mathbf{X},g) = \int_0^1 s_i(u;Z;\mathbf{X},g)\zeta(u)du. \tag{2.6}$$

While $S_i(Z;\mathbf{X},g)$ summarizes sensitivity in a single number, it will be seen in examples that it can be informative to plot the whole curve $s_i(u;Z;\mathbf{X},g)$.

A particular choice of the weight function ζ induces an interpretation of the sensitivity measure that is consistent with the ideas of Regionalized Sensitivity Analysis (RSA)[26,39,40]. In such approaches, the distribution of each model input is compared, conditional on the system being or not in states of a predefined (adverse) behavior. When the conditional distribution differs across adverse and non-adverse states, then the model input is classified as important. In our setting, we can consider an event A that includes states of particular interest. If those states correspond to a range of Y, then we can choose confidence levels $0 \leq \alpha < \beta \leq 1$ and define $A = \{F_Y^{-1}(\alpha) < Y < F_Y^{-1}(\beta)\}$ ($\beta = 1$ is often chosen to reflect the full tail of the probability distribution, giving rise to the TVaR risk measure[18]). Then, if we let $\zeta(u) = \frac{1}{\beta-\alpha}\mathbf{1}_{u\in(\alpha,\beta)}$, we obtain $\rho(Y) = \mathbb{E}(Y|A)$, $S_i(\mathbf{X},g) = \mathbb{E}(X_i g_i(\mathbf{X})|A)$. Hence the risk measure is the expected value of the output Y conditional on the adverse event A and the sensitivity $S_i(\mathbf{X},g)$ is the expected value of $X_i g_i(\mathbf{X})$ over the same states.

In global sensitivity methods, it is typically required that sensitivity measures reflect the dependence between X_i and Y and in particular the extent to which knowledge of X_i will reduce the variability of Y[3,5,11]. To see how this concern is reflected in the current framework, observe that $S_i^{\mathcal{D}}(\mathbf{X},g) = \mathbb{C}(X_i, g_i(\mathbf{X})\zeta(U_Y))$, such that the dependence between X_i and (the rank-transformation of) Y is captured. When the function g is linear in the i^{th} element, $g_i(\mathbf{X}) = g_i$ is constant and the expression simplifies to $g_i\mathbb{C}(X_i, \zeta(U_Y))$. Furthermore, we can reflect a marginal reduction in the variability of X_i, by changing X_i to $X_i^\epsilon = (1-\epsilon)X_i + \epsilon\mathbb{E}(X_i)$ for some small

$\epsilon > 0$. By the definition of the sensitivity measure it then is

$$\rho(g(\mathbf{X})) - \rho(g(X_i^\epsilon, \mathbf{X}_{-1})) \approx -\epsilon S_i^{\mathcal{D}}(\mathbf{X}, g). \tag{2.7}$$

Thus the deviation component of the sensitivity reflects the reduction in $\rho(g(\mathbf{X}))$ induced by a small reduction in the variability of X_i.

In sensitivity analysis, stressing a group of factors at the same time is often desired (the fourth desirable property of Saltelli[3]). Let such a group be indexed by a set $G \subset \{1, \ldots, d\}$. Consider a vector of random shocks $(Z_1, \ldots, Z_d)$ and let $\mathbf{Z}^G$ be a random vector such that $Z_i^G = Z_i$ if $i \in G$ and $Z_i^G = 0$ otherwise. Then $\rho\big(g(\mathbf{X} + \epsilon \mathbf{Z}^G)\big)$ is the output risk summary with all model inputs in the group G stressed at the same time. Straightforward calculus leads to

$$\frac{\partial}{\partial \epsilon} \rho\big(g(\mathbf{X} + \epsilon \mathbf{Z}^G)\big)\Big|_{\epsilon=0} = \sum_{i \in G} S_i(Z_i; \mathbf{X}, g). \tag{2.8}$$

Therefore, sensitivity to the group G of model inputs decomposes into a sum of single-factor sensitivities. This form of additivity is shared with the (local) *differential importance measure*[41,42], which calculates sensitivity as the ratio of a partial derivative over the total differential. In this respect, the sensitivity method proposed here can be seen as an extension of the differential importance measure, applying to functionals.

A further feature that sometimes occurs in sensitivity analyses is an additive decomposition of the output variability to sensitivities[3,5,43]. As we focus on ranking the importance of different model inputs rather than a variability allocation, we do not pose such an additivity requirement, consistently with other moment-independent sensitivity analysis methods[6,11]. However, in the special case when g is linear, the first derivatives $g_i(\mathbf{x}) = g_i$ are constant and we can write $g(\mathbf{x}) = g(\mu) + \sum_{i=1}^d g_i(x_i - \mu_i)$, for any vector $\mu \in \mathbb{R}^d$. If we let μ stand for the vector of means of $\mathbf{X}$, it follows that $\rho(g(\mathbf{X})) - g(\mu) = \sum_{i=1}^d S_i^{\mathcal{D}}(\mathbf{X}, g)$, i.e. an additive apportioning of the risk summary is effected. The case of linear g is the one typically considered in financial applications, where linearity corresponds to a standard portfolio structure[21–23]. An additive apportioning

of the risk summary $\rho(g(\mathbf{X}))$ may be obtained if one uses a Sobol'/Hoeffding expansion of the function g[43,44] (rather than the variance of $g(\mathbf{X})$). Such a process is outside the scope of the present paper.

Further formal properties of the sensitivity measure are now briefly discussed. Precise statements and proofs are given in Appendix B. A first consideration relates to the possible sign of the sensitivity measure. This is of natural interest, since by its definition $S_i(Z;\mathbf{X},g)$ enables approximation of the change in the risk summary due to a marginal change in exposure, by $\rho(g(\mathbf{X}+\epsilon\mathbf{e}_i Z)) - \rho(g(\mathbf{X})) \approx \epsilon S_i(Z;\mathbf{X},g)$ for small ϵ. When the model function g is convex in the i^{th} argument and the weight function ζ is increasing, then $\epsilon S_i(Z;\mathbf{X},g)$ actually provides a lower bound for the incremental change in the risk summary (Lemma 4). Restricting to the case $Z = X_i$, it furthermore holds that convexity of g and positive dependence of the random vector $\mathbf{X}$ leads to positive values of the deviation component $S_i^{\mathcal{D}}\big(\mathbf{X},g\big)$ (Lemma 5).

Quantitative models are often built hierarchically via graphical methods: a number of lower level model inputs are specified and then combined into model inputs at a higher level, which in turn determine the output. Such hierarchical structures can be exploited in order to reduce the dimensionality of the model function whose gradient is required for sensitivity analysis (Lemma 6). Mathematically this is an elementary argument, but the computational benefits it produces may be substantial. The same approach is extendible to situations where common model inputs impact on different branches of the hierarchical structure (Lemma 7). Finally, note the potential arbitrariness in the choice of model inputs, as a non-linear change of scale for a particular model input will tend to change the measured sensitivity. However, the deviation component of sensitivity is invariant to linear transformations of model inputs (Corollary 8). This means that the sensitivity measure introduced here does not satisfy the property of invariance under monotonic transformations possessed by moment-independent importance measures[9,11], unless the transformation is linear.

2.4 Treatment of discrete inputs

It is possible that one or more model inputs are modelled as discrete random variables. In such a case, Proposition 2 is still applicable, but some further comment is necessary. For simplicity, consider that X_1 takes only the values 0 or 1. Then, values of the function $g(\mathbf{x})$ are only observed for $x_1 = 0$ or $x_1 = 1$; g is therefore not uniquely defined for $x_1 \in (0, 1)$. Hence, differentiation of g in the first argument is not meaningful, unless g is extended in order to take values for $x_1 \in (0, 1)$. Such an interpolation is not unique and should be chosen in a way that is interpretable. Unavoidably, the choice of g will then be subjective.

As a particular choice, one may let the function g be linear in its first argument, such that for $x_1 \in (0, 1]$ it is

$$g(x_1, \mathbf{x}_{-1}) = g(0, \mathbf{x}_{-1}) + x_1(g(1, \mathbf{x}_{-1}) - g(0, \mathbf{x}_{-1})) \tag{2.9}$$

Since $[X_1 g_1(\mathbf{X})|X_1 = 0] \equiv 0$ the value of $g_1(0, \mathbf{x}_{-1})$ is not of primary interest for the sensitivity; as a matter of convention one may set it to zero. Such a choice is presented in Example 1 of Section 3.

3 APPLICATION TO SIMULATION MODEL

In practical applications, quantitative risk models are typically implemented in a Monte-Carlo simulation environment, since determining the probability distribution of Y analytically is often unfeasible. After specifying the joint distribution of $\mathbf{X}$ and the model function g, n independent observations of $(\mathbf{X}, Y = g(\mathbf{X}))$ are simulated. The risk summary $\rho(Y)$ is then estimated from the empirical distribution of Y thus obtained.

While specification of the distribution of $\mathbf{X}$ and of the model g is necessary to implement the simulation, a risk analyst or model reviewer does not necessarily have unencumbered access to this information. The exact specification of the model may be buried in thousands of pages

of documentation. Moreover the function g is often so complex that, even if well documented, may be cumbersome to use in explicit form. (Model complexity is often the motivation behind performing sensitivity analyses in the first place.) While some information on the model model can of course be glanced from documentation (e.g. hierarchical structure), here we simplify the discussion by considering a *black box* model. In the current context, that means that a risk analyst has only access to the simulated sample $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, where $\mathbf{x}_j = (x_{j1}, \ldots, x_{jd})$, and (for now) nothing else.

From Proposition 2, the sensitivity to the i^{th} model input may be estimated as

$$\hat{S}_i(\mathbf{X}, g) = \frac{1}{n} \sum_{j=1}^{n} x_{ji} \hat{g}_i(\mathbf{x}_j) \zeta(u_j), \tag{3.1}$$

where u_j are obtained via the empirical distribution $\hat{F}_Y$, such that $u_j = \hat{F}_Y(y_j)$, and $\hat{g}_i(\mathbf{x}_j)$ is an estimate of $g_i(\mathbf{x}_j)$.

Estimating the gradient of g from the sample provided is not trivial. One possible method is to use tools from non-parametric statistics, in particular *local linear regression* [28,29,45,46]. In that framework, smoothing methods are used to provide non-parametric estimates of g and its gradient, by formulating a statistical model of the form $\mathbb{E}\big(Y_j \big| \mathbf{X}_j\big) = g(\mathbf{X}_j)$. A detailed review of local linear regression is beyond the scope of this paper; the description of local linear regression in the rest of the present section follows Li and Racine [29]. For simplicity assume, as in Section 2.4, that the first model input is a Bernoulli variable and all others are continuous. Denote by $\mathbf{x}_{j,-1} = (x_{j,2}, \ldots, x_{j,d})$ the vector of observations for model inputs $X_2, \ldots, X_d$ from the j^{th} simulated scenario. Estimates of the model function and its gradient for continuous model inputs, $(\hat{g}(\mathbf{z}), \hat{g}_2(\mathbf{z}), \ldots, \hat{g}_d(\mathbf{z}))$ at some point $\mathbf{z} \in \mathcal{X}$ are obtained by solving

$$(\hat{g}(\mathbf{z}), \hat{g}_2(\mathbf{z}), \ldots, \hat{g}_d(\mathbf{z})) = \arg\min_{\alpha,\beta} \sum_{j=1}^{n} \big(y_j - \alpha - \beta'(\mathbf{x}_{j,-1} - \mathbf{z}_{-\mathbf{1}})\big)^2 K_{\boldsymbol{\delta}}(\mathbf{x}_j, \mathbf{z}) \tag{3.2}$$

where $K_{\boldsymbol{\delta}}(\mathbf{x}_j, \mathbf{z}) = \prod_{i=1}^{d} k_{i,\delta_i}(x_{ji}, z_i)$ is a *product kernel*. Thus the estimates $(\hat{g}(\mathbf{z}), \hat{g}_2(\mathbf{z}), \ldots, \hat{g}_d(\mathbf{z}))$ are produced by fitting a hyperplane in a neighborhood of $\mathbf{z}$. For continuous variables, each func-

tion $k_{i,\delta_i}(x_{ji}, z_i),\ i = 2, \ldots, d$ is a kernel function with *bandwidth* δ_i, that is, a univariate symmetric density (e.g. Gaussian) with scale parameter δ_i. For the indicator X_1, a kernel of the type $k_{1,\delta_1}(x_{j1}, z_1) = \mathbf{1}_{\{x_{j1}=z_1\}} + \delta_1 \mathbf{1}_{\{x_{j1}\neq z_1\}}$ is used, a limiting case of which is $k_{1,0}(x_{j1}, z_1) = \mathbf{1}_{\{x_{j1}=z_1\}}$, such that no smoothing over X_1 takes place. In-sample optimal selection of the vector of bandwidths $\boldsymbol{\delta}$ can be carried out by methods like leave-one-out cross-validation, though such bandwidth selection procedures can be computationally very intensive.

In the following example, drawn from insurance loss modelling, it is shown how the above ideas can be applied. Of course the model considered here is not a black box, but the example shows that the sensitivities of even standard models can display complexities that the proposed methodology elucidates. The model function is a simple function containing additive, multiplicative and non-linear components, thus similar in nature to the test models often used in to demonstrate sensitivity analysis methods[6]. Notwithstanding its simplicity, the example displays standard features found in more complex models used in risk analysis: namely the presence of uncertain correlated inputs joined by a highly non-linear function. This is the case for models used e.g. in safety assessment for nuclear waste disposal[7], financial engineering and the study of chemical reactions (Sec. 6.2-6.3 in Ref.[5]). In particular, the use of a discrete Bernoulli variable is also observed in the construction of composite indices combining different input averaging methods (Sec. 6.1 in Ref.[5]).

Example 1. Consider a simple insurance portfolio of short-term liabilities in two areas (*lines*) of business. The first and second lines yield total claims X_1 and X_2 respectively. The claims from X_1 and X_2 are statistically independent, but are both subject to adjustment by the same multiplicative factor X_3 (e.g. due to inflation). Hence the total cost of insurance claims is given by

$$C = X_3(X_1 + X_2). \tag{3.3}$$

The insurer has bought protection via a reinsurance contract. In particular, a reinsurance layer with *deductible* λ and *limit* l has been purchased. This means that the reinsurer agrees to pay claims in excess of λ, but only up to a maximum payment of l, corresponding to a claim of $C = \lambda + l$. However there is always the chance that the reinsurer does not fulfil his contractual obligation, e.g. because of a default. Let then X_4 be the *default indicator*, a Bernoulli variable taking values 0 and 1 in the respective cases of non-default and default respectively. Thus, the total loss is given by

$$Y = C - (1 - X_4) \min\{(C - \lambda)_+, l\}, \tag{3.4}$$

which with (3.3) defines fully the function g.[3]

The statistical assumptions and parameters used in the example are:

- $X_1 \sim LogNormal$, with $\mathbb{E}(X_1) = 153,\ \mathbb{V}(X_1) = 44^2$;
- $X_2 \sim Gamma$, with $\mathbb{E}(X_2) = 200,\ \mathbb{V}(X_1) = 10^2$;
- $X_3 \sim LogNormal$ with $\mathbb{E}(X_3) = 1.05,\ \mathbb{V}(X_3) = 0.01^2$;
- $X_4 \sim Bernoulli(0.1)$;
- X_1, X_2, X_3 are mutually independent;
- C, X_4 are dependent by a Gaussian copula with correlation parameter 0.6. Conditional on C, X_4 is independent of X_1, X_2, X_3;
- $\lambda = 380,\ l = 30$. The values of λ and $\lambda + l$ correspond to the 63^{th} and 82^{th} percentiles of C respectively.

[3]Following the discussion of Section 2.4, for $x_4 \in (0, 1)$, values of the function $g(\mathbf{x})$ are interpretable; in those scenarios we have partial default, were only a fraction of $\min\{(C-\lambda)_+, l\}$ is paid by the reinsurer. Note that only values $X_4 = 0$ or 1 are simulated; the above argument is only used in the interpretation of the derivative $g_4(\mathbf{X})$.

The statistical assumptions imply that X_4 and C are positively dependent, reflecting the possibility that the reinsurer is under financial distress under those scenarios where performance of the contract is most likely needed.

A Monte-Carlo sample of size $n = 20000$ was obtained from $(\mathbf{X}, Y)$. To this sample, a local linear regression model was fitted, using the NP package in $\mathbf{R}$[47]. For X_1, X_2, X_3, Gaussian kernels with bandwidths $4, 4, 0.01$ respectively were used. For the discrete variable X_4, we used $k_4(x, z) = \mathbf{1}_{\{x=z\}}$.[4]

In Figure 1, the quantities $x_{ji}\hat{g}_i(\mathbf{x}_j)$ are plotted as black circles, against $u_j = \hat{F}_Y(y_j)$, such that large values of u in the horizontal axis correspond to high outcomes of the output Y. The grey line represents an estimate of the function $u \mapsto s_i(u; \mathbf{X}, g) = \mathbb{E}\big(X_i g_i(\mathbf{X}) | U_Y = u\big)$, obtained again by local regression with a bandwidth of 0.01. As discussed in Section 2.2, averaging over the curve $s_i(u; \mathbf{X}, g)$ with respect to weights $\zeta(u)$ yields $S_i(\mathbf{X}, g)$.

It is seen that for X_1, X_2, X_3, the plotted curves dip, representing a reduced sensitivity under those scenarios where a payment from the reinsurer is received (the black circles that do not follow that pattern correspond to observations $x_{j4} = 1$ where reinsurance default has taken place). Thus, once the reinsurance layer is activated, small changes in those model inputs have no effect on the value of Y. Once the layer is exhausted (i.e. for scenarios with $y_j > \lambda + l$) the sensitivity rises again.

Comparing the curves for X_1 and X_2, it is seen that (ignoring the dip in the curves) the sensitivity to X_2 is higher for low u, close to its higher mean of 200. However, the sensitivity to X_1 shows an increasing trend, ending up dominating for large confidence levels of the output, i.e. for $u > 0.9$. This reflects its higher volatility and the heavier tail of the LogNormal distribution

[4] Due to the small number of simulated scenarios in which Y is very high, it is possible that the local regression algorithm fails to produce reasonable gradients for some states. In this implementation, we did not use the 30 samples (out of 20000) that correspond to the highest values of Y, when estimating the gradient of g.

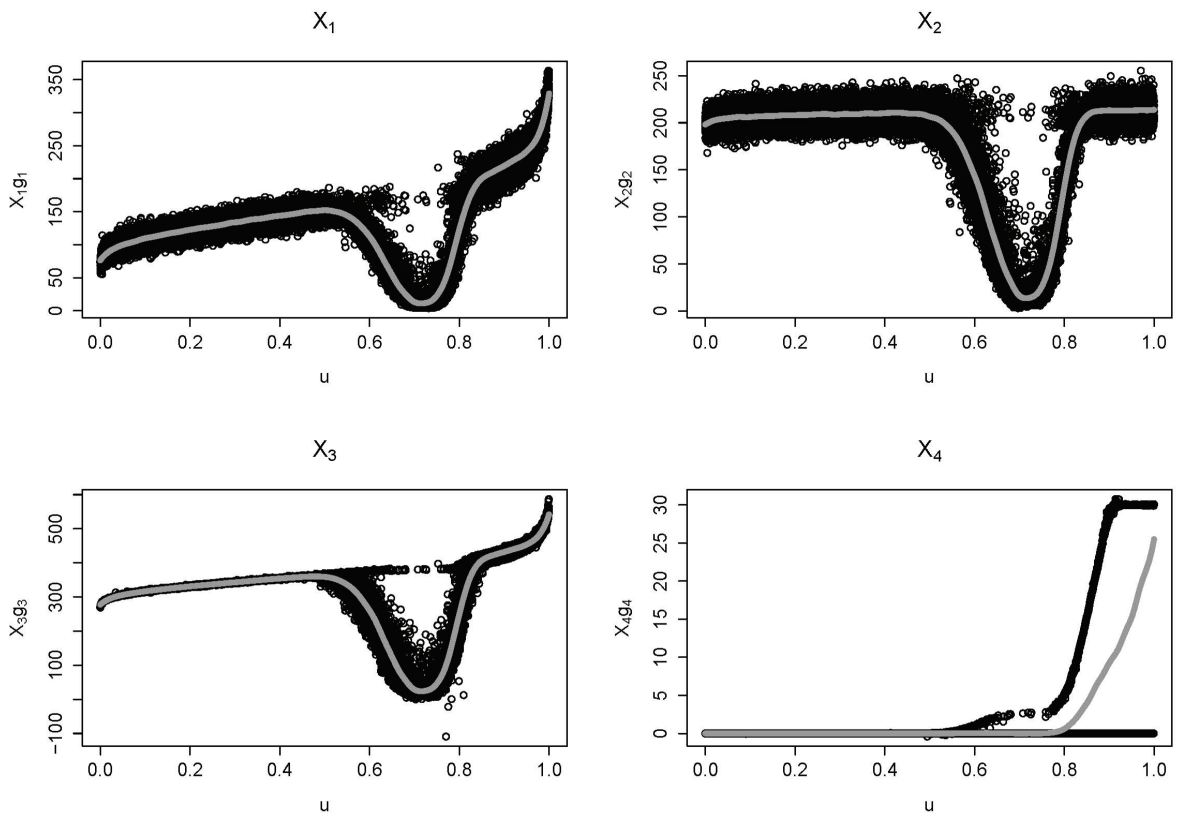

Figure 1: Sensitivity estimates in example 1. Each black circle represents an observation $x_{ji}\hat{g}_i(\mathbf{x}_{ij})$; the grey line represents a non-parametric estimate of $u \mapsto s_i(u; \mathbf{X}, g) = \mathbb{E}\big(X_i g_i(\mathbf{X}) | U_Y = u\big)$.

used to model X_1. The sensitivity to X_3 also shows an increasing trend in u, even though the volatility of X_3 is low. This reflects the impact of the multiplicative effect of X_3.

On the other hand, it is seen that the sensitivity curve $s_4(u; \mathbf{X}, g)$ for X_4 is zero up until approximately $u = 0.6$. This corresponds to scenarios where the reinsurance layer is not activated, and thus reinsurance default is inconsequential. Then the black circles bifurcate. The part that is stuck at zero corresponds to non-default events $\{X_4 = 0\}$, while the increasing part corresponds to default events $\{X_4 = 1\}$. The curve increases up to $l = 30$ which is the maximum reinsurance recovery that may be obtained. At the most extreme scenarios for the output,

Table I: Scaled sensitivities $S_i(\mathbf{X}, g)/\rho(Y)$ for a range of risk measures.

	TVaR($p = 0.99$)	TVaR ($p = 0.95$)	ED (γ=20)	ED (γ=5)
$\rho(Y)$	516.0	471.4	455.8	409.5
X_1	0.593	0.567	0.550	0.425
X_2	0.415	0.454	0.460	0.416
X_3	1.001	1.019	1.009	0.834
X_4	0.050	0.043	0.038	0.019

the highest sensitivity is obtained, reflecting the fact that under those scenarios the maximum recovery is due and will therefore be lost in the case of default.

In Table I, we report the sensitivities $S_i(\mathbf{X}, g)$ to model inputs for two risk measures, TVaR and Exponential Distortion (ED), at different levels of conservativism. The respective weight functions are

$$\zeta_{TVaR}(u; p) = \frac{1}{1-p}\mathbf{1}\{u > p\}, \quad \zeta_{ED}(u; \gamma) = \frac{\gamma \exp(\gamma u)}{\exp(\gamma) - 1}. \tag{3.5}$$

The TVaR risk measure places a constant positive weight to extreme loss scenarios ($u > p$) and a weight of zero to all other scenarios. The ED risk measure places a positive weight to all scenarios, strictly increasing in u, thus reflecting the whole of the output distribution. The risk measures are ordered from left to right on the table, in terms of a decreasing weight placed on the right tail of the output loss distribution, noting that $\zeta_{TVaR}(1; 0.99) > \zeta_{TVaR}(1; 0.95) = \zeta_{ED}(1; 20) > \zeta_{ED}(1; 5)$.

To enable comparison, the reported sensitivities are in each case scaled by the risk summary $\rho(Y)$, giving effectively a measure of elasticity to changes in model inputs. It is seen that the highest sensitivity is to the multiplicative factor X_3, followed by the LogNormal claims X_1. The difference between the sensitivities to X_1 and X_2 reduces as less tail-sensitive risk measures are

Table II: Breakdown of scaled sensitivities $S_i(\mathbf{X}, g)/\rho(Y)$ to mean and deviation components, for TVaR with $p = 0.95$.

	$S_i^{\mathcal{M}}(\mathbf{X}, g)/\rho(Y)$	$S_i^{\mathcal{D}}(\mathbf{X}, g)/\rho(Y)$	$S_i(\mathbf{X}, g)/\rho(Y)$
X_1	0.340	0.227	0.567
X_2	0.446	0.008	0.454
X_3	1.018	0.001	1.019
X_4	0.004	0.039	0.043

used. The sensitivity to X_4 is not very high and reduces again as less tail-sensitive risk measures are used. These observations are consistent with the previous discussion of Figure 1.

While the sensitivity values assigned to different inputs by the TVaR and ED measures are different, the importance ranking produced is the same. In Example 2, it will be seen that there are circumstances where the exclusive emphasis of TVaR on high outputs can make it "blind" to sensitivities prevalent in less extreme scenarios. Note also that each point on the curves $u \mapsto s_i(u; \mathbf{X}, g)$ corresponds to sensitivity with respect to the VaR_u measure, with the importance ranking produced strongly dependent on the confidence level u. Hence we would not advise the use of a VaR measure at a single confidence level. More broadly, while a risk measure such as ED can be used to produce a ranking between inputs, consideration of the full set of curves $u \mapsto s_i(u; \mathbf{X}, g)$ in Figure 1 is helpful, as it allows the risk analyst to assess comparative importance of inputs at different output confidence levels.

Further insight may be obtained by apportioning each sensitivity $S_i(\mathbf{X}, g)$ to its mean and deviation elements $S_i^{\mathcal{M}}(\mathbf{X}, g)$ and $S_i^{\mathcal{D}}(\mathbf{X}, g)$, as in (2.2). Such a decomposition for TVaR with $p = 0.95$ is reported in Table II. It can be seen that the sensitivity to X_1 is driven to a large extent by its variability, while the sensitivity to X_2 is almost exclusively driven by its mean.

Similarly, the sensitivity to X_3 is also driven by its mean due to the multiplicative effect of inflation. On the other hand, the sensitivity to the indicator X_4 is mainly driven by variability. In particular, focusing on the second column of the table, it is clear that the risk summary $\rho(Y)$ is more sensitive to the deviation element of the default indicator X_4 than either X_2 or X_3. $\square$

4 EXTENSIONS

4.1 Sensitivity to model parameters

The model function g will generally be characterised by a number r of parameters, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)$, a dependence that has up to now been suppressed in the notation. If there is uncertainty around these parameters, they can be treated as model inputs. Here we consider the case where model parameters are known constants (such as λ in Example 1). Thus we now consider the model as a function of $d + r$ arguments, such that $Y = g(\mathbf{X}, \boldsymbol{\lambda})$, where $\mathbf{X}$ is a random vector and $\boldsymbol{\lambda}$ is a vector of constants.

A simple way of defining the sensitivity of $\rho(Y)$ to the model parameter λ_k is to consider it as a degenerate random variable with all probability mass concentrated at one point. Thus the sensitivity to λ_k may be obtained as

$$S_{d+k}((\mathbf{X}, \boldsymbol{\lambda}), g) = \lambda_k \mathbb{E}\big(g_{d+k}(\mathbf{X}, \boldsymbol{\lambda})\zeta(U_Y)\big). \tag{4.1}$$

Hence the sensitivity to λ_k is nothing but λ_k itself, multiplied by the corresponding model function gradient, averaged over scenarios of interest as specified by ζ. Note that $S^{\mathcal{D}}_{d+k}((\mathbf{X}, \boldsymbol{\lambda}), g) = 0$.

The partial derivative g_{d+k} is not directly obtainable via the local regression scheme of Section 3, since the simulated sample considered there will not supply any variability in the $(d + k)^{th}$ argument of g. A practical way of dealing with this issue is to randomize the parameter vector $\boldsymbol{\lambda}$, such that quantities $g(\mathbf{X}, \boldsymbol{\Lambda})$ are simulated, where $\boldsymbol{\Lambda}$ is a random vector of elements that are independent of each other and of $\mathbf{X}$, with $\mathbb{E}(\Lambda_k) = \lambda_k$. In the current discussion this

randomization serves only numerical purposes and does not represent uncertainty in relation to the value of λ_k. Once a local linear regression model has been estimated on that sample, predictions of the model function and its gradients for fixed $\boldsymbol{\lambda}$ can be simply obtained from (3.2), by evaluating the non-parametric estimates at the points $\mathbf{z}_j = (\mathbf{x}_j, \boldsymbol{\lambda})$. Although the choice of the distribution for Λ_k is arbitrary, the measured sensitivity is robust to different choices. Indeed, the volatility of Λ_k is only used to estimate the gradient g_{d+k} and for sensitivity calculation the conditioning $\Lambda_k = \lambda_k$ applies.

4.2 Sensitivity to statistical parameter uncertainty

To study the sensitivity of the output to the uncertainty of statistical parameters, we specify the distribution of each model input conditionally on the parameters. In particular, we assume that $X_i|\Theta_i = \theta_i \sim F(\cdot|\theta_i)$. The distribution G_i of the random variable or vector Θ_i reflects the uncertainty pertaining to the value of the parameters of the distribution of X_i.[5]

For evaluation of the risk summary and of sensitivities, the unconditional distribution of X_i, given by $\hat{F}_i(x) = \int F_i(x|\theta) dG_i(\theta)$ is used. In particular, when a Monte-Carlo sample is produced, in every state a different observation from Θ_i is obtained, and subsequently an observation of X_i given that outcome of Θ_i is simulated. Thus Θ_i is allowed to vary across simulated scenarios. As opposed to the case of model parameters discussed in Section 4.1, we do not attempt to "smooth out" the uncertainty in Θ_i, but retain it as an explicit feature of the model.

Our purpose now is to calculate the sensitivity of $\rho(Y)$ to each of $\Theta_1, \ldots, \Theta_d$. To do this, we express each model input X_i as a function of Θ_i, representing *parameter uncertainty*, and another random variable Z_i, representing pure *process (stochastic) uncertainty*. This can be achieved, for a conditional distribution F_i, by setting $U_i = F_i(X_i|\Theta_i)$ and $Z_i = F_i^{-1}(U_i|\hat{\theta}_i)$,

[5]Taking a rather pragmatic view, we do not specify the provenance of G_i – it may be derived by a bootstrap procedure, by asymptotic theory or by Bayesian analysis.

where $\hat{\theta}_i = \mathbb{E}(\Theta_i)$ is an unbiased estimate of the unknown parameter. Then we have that Z_i follows the distribution F_i with parameter $\hat{\theta}_i$ and is independent of Θ_i. The precise statement of this argument is given in Appendix C, Lemma 9.

As a result, we can express X_i as $X_i = \psi^{(i)}(Z_i, \Theta_i)$, where the first and second arguments of $\psi^{(i)}$ represent the contribution of process and parameter uncertainty to X_i respectively. Therefore, following Lemma 6 in Appendix B, the sensitivities to Z_i and Θ_i can be respectively calculated as

$$\begin{aligned} &\mathbb{E}\big(Z_i g_i(\mathbf{X}) \psi_1^{(i)}(Z_i, \Theta_i) \zeta(U_Y)\big) \\ &\mathbb{E}\big((\Theta_i - \hat{\theta}_i) g_i(\mathbf{X}) \psi_2^{(i)}(Z_i, \Theta_i) \zeta(U_Y)\big). \end{aligned} \tag{4.2}$$

For parameter uncertainty we propose the use of the deviation component of sensitivity only. This is because a marginal change in Θ_i may be hard to interpret in the context of parameter uncertainty; on the other hand it is plausible to reduce the volatility of Θ_i via diligent data collection, expert judgement and further analysis. While it may be possible to evaluate (and differentiate) the functions $\psi^{(i)}$ analytically, when working with a Monte-Carlo sample, they can be estimated directly from the data by local linear regression.

The following example builds on Example 1. By considering the sensitivity to known model parameters, we are able to quantify the importance of the "constants" found in any model, such as the initial inventories of radionuclides in the Level E model[7]. The importance of those constants is assessed in the context of a global sensitivity analysis, since it still depends on the variation of the uncertain inputs. Furthermore, we consider the uncertainty in statistical parameters; this, though simpler, parallels the stochastic volatility model described by Saltelli et al. (Sec. 6.2, Ref.[5]), where the variance of a diffusion process is endowed with its own dynamics.

Example 2. We return to the setting of Example 1 and consider sensitivity to the reinsurance

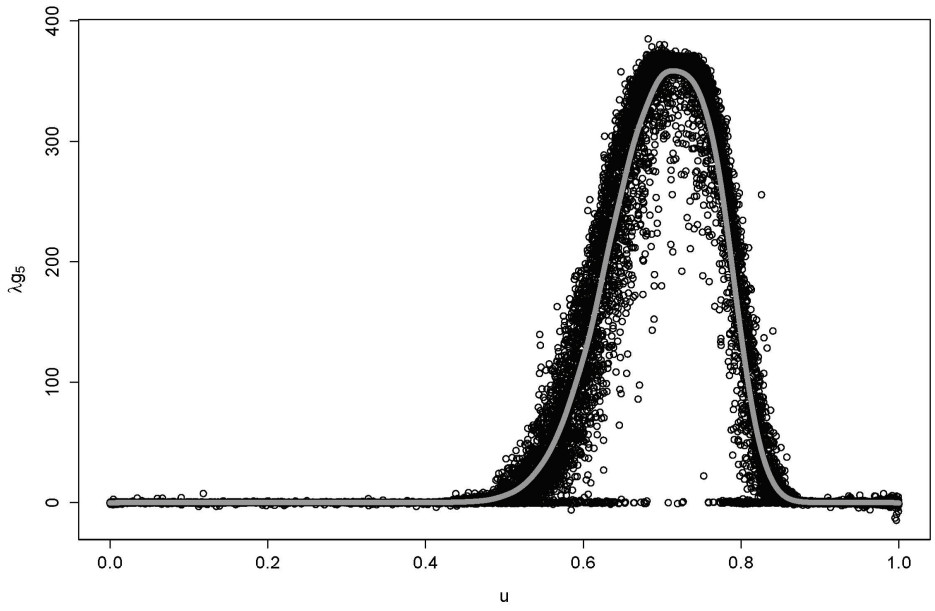

Figure 2: Sensitivity to λ in example 2. Each black circle represents an observation $\lambda\hat{g}_5(\mathbf{x}_{ij})$; the grey line represents a non-parametric estimate of $u \mapsto s_5(u; \mathbf{X}, g) = \lambda\mathbb{E}\big(g_5(\mathbf{X})|U_Y = u\big)$.

deductible parameter λ. For this purpose we run the simulation with a randomized deductible, uniformly distributed around the known value of $\lambda = 380$, such that $\Lambda \sim Unif[375, 385]$. The corresponding plot of the estimated sensitivity function $s_5(u; (\mathbf{X}, \lambda), g)$ is given in Figure 2 and the scaled sensitivities $S_5(u; (\mathbf{X}, \lambda), g)$ for different risk measures are collected in the first row of Table III.

From Figure 2 it is seen that the sensitivity to λ is positive for that range of confidence levels of output where the reinsurance layer is used but not exhausted. Since the parameter l is unchanged, the positive sensitivity represents the fact that for a higher deductible, reinsurance recoveries will be less frequent. The sensitivity is zero for low confidence levels, as in those scenarios the reinsurance layer is not used. It is also zero for high confidence levels, corresponding to situations where the layer is exhausted and the exact loss level at which the layer was activated no longer matters. These considerations are reflected in Table III, where it is seen that the

Table III: Scaled sensitivities to λ, Z_1, Θ_1 for a range of risk measures.

	TVaR($p = 0.99$)	TVaR ($p = 0.95$)	ED (γ=20)	ED (γ=5)
λ	0.000	0.000	0.014	0.182
Z_1	0.490	0.485	0.478	0.389
$\Theta_1 - \hat{\theta}_1$	0.103	0.078	0.073	0.039

sensitivities to the TVaR measures is zero, since these risk measures only consider scenarios at the far tail of F_Y, where the reinsurance layer is completely used up. This shows that a risk measure such as ED, assigning a positive weight at all confidence levels, may be preferable to TVaR. The example also demonstrates the usefulness of plotting the sensitivity function at all confidence levels, in order to detect local effects.

Now we consider the case of parameter uncertainty in the claims from the first line of business. We assume that there is some uncertainty around the first parameter of the distribution of X_1. In particular we let $X_1|\Theta_1 \sim LogNormal\big(\Theta_1, \sigma_1^2\big)$ and $\Theta_1 \sim Normal(\hat{\theta}_1, \sigma^2_{\Theta_1})$. After simple manipulations involving Normal and LogNormal distributions, the decomposition of X_i becomes

$$X_1 = \psi^{(1)}(\Theta_1, Z_1) = \exp(\Theta_1 - \hat{\theta}_1) Z_1,$$

where $Z_1 \sim LogNormal\big(\hat{\theta}_1, \sigma_1^2\big)$. The specific parameters chosen are $\hat{\theta}_1 = 4.99,\ \sigma_1 = 0.230,\ \sigma_{\Theta_1} = 0.163$, such that the unconditional moments of X_1 are (consistently with Example 1) $\mathbb{E}(X_1) = 153,\ \mathbb{V}(X_1) = 44^2$, while conditional moments are $\mathbb{E}(X_1|\Theta_1 = \hat{\theta}_1) = \mathbb{E}(Z_1) = 151,\ \mathbb{V}(X_1|\Theta_1 = \hat{\theta}_1) = \mathbb{V}(Z_1) = 35^2$.

The sensitivities to Z_1 and $\Theta_1 - \hat{\theta}_1$, calculated according to (4.2), are reported in the second and third row of Table III. It is seen that, particularly for the tail-focused TVaR measures, the sensitivity of $\rho(g(\mathbf{X}))$ to parameter uncertainty is significant. $\square$

4.3 Sensitivity to the dependence structure

Finally, we turn our attention to studying the sensitivity of the output risk summary to the dependence structure. We propose an approach that is applicable for dependence structures that display conditional independence, given some observable or latent common model inputs; the latter appear implicitly in standard dependence models, such as the popular class of Archimedean copulas (see Ref.[32], Sections 3.4.1 and 5.4). The idea pursued here is to express each model input X_i as

$$X_i = \psi^{(i)}(Z_i, \mathbf{V}), \tag{4.3}$$

where the random variables $Z_1, \ldots, Z_d$ are independent and $\mathbf{V}$ is the vector of common model inputs.

Henceforth, we call the elements of $\mathbf{V}$ the *common factors* and $Z_1, \ldots, Z_d$ the *idiosyncratic factors*. When the common factors are explicitly modelled and have a direct interpretation (e.g. as natural catastrophe losses, affecting different parts of an insurance portfolio), the representation (4.3) is essentially given. In the case that common factors are implicit in the dependence structure, then the representation (4.3) is again possible by the following argument. For presentational simplicity, let us assume that $F_i(\cdot|\mathbf{v})$, the conditional distribution of X_i given $\mathbf{V} = \mathbf{v}$, is invertible. Then, to obtain the idiosyncratic factor Z_i, we may just set $Z_i = H_i^{-1}(F_i(X_i|\mathbf{V}))$, where H is an arbitrary distribution (a plausible choice may be to let H_i be the unconditional distribution of X_i). Then, Z_i will be independent of $\mathbf{V}$ and we can write $X_i = \psi^{(i)}(Z_i, \mathbf{V}) = F_i^{-1}(H_i(Z_i)|\mathbf{V})$. A precise statement of these facts is given in Lemma 10, Appendix C.[6]

[6]By its construction, $\psi^{(i)}$ is increasing in its first argument. A particular case of interest emerges if $\mathbf{V} \equiv V$ is 1-dimensional and $\psi^{(i)}$ is also increasing in the second argument for all i. In that case, X_i is *stochastically increasing* in V and the random vector $\mathbf{X}$ is *associated*; see Property 7.2.16 in Ref.[30]. Thus, each element of the vector of dependent model inputs $\mathbf{X}$ can be written as a monotonic function of two independent random variables,

The sensitivity to the dependence structure may now be studied by calculating the sensitivity to the common and idiosyncratic factors. For simplicity consider a univariate common factor V. Then the sensitivity of output risk summary to Z_i and V can be respectively calculated as

$$\mathbb{E}\big((Z_i - \mathbb{E}(Z_i))g_i(\mathbf{X})\psi_1^{(i)}(Z_i, V)\zeta(U_Y)\big) \tag{4.4}$$

$$\sum_{j=1}^{d} \mathbb{E}\big((V - \mathbb{E}(V))g_j(\mathbf{X})\psi_2^{(j)}(Z_j, V)\zeta(U_Y)\big). \tag{4.5}$$

We emphasize the deviation component of the sensitivities, since the prime effect of the dependence of $\mathbf{X}$ is on the variability of the output.

The proposed process is demonstrated via the following example, where sensitivity to the dependence between the output loss of Example 1 and an investment position is studied. The importance of dependence between inputs is recognised widely in sensitivity analysis, e.g.[5,7], but, to our knowledge, the sensitivity to factors driving such dependence has not been studied in the literature.

Example 3. To allow the focus on dependence, here we consider a simple extension to Example 1. Assume that assets are invested that produce income A, with expected value equal to that of the total loss Y. The random variable A is not independent of Y. After taking into account such investment, the loss becomes $\tilde{Y} = Y - A$. For consistency with the notation of this section, we re-label the variables as $\tilde{X}_1 \equiv Y,\ \tilde{X}_2 \equiv -A$, such that

$$\tilde{Y} = \tilde{g}(\tilde{X}_1, \tilde{X}_2) = \tilde{X}_1 + \tilde{X}_2.$$

Therefore $\tilde{Y}$ is now the model output, arising through simple addition of the inputs $\tilde{X}_1, \tilde{X}_2$.

The random variable $\tilde{X}_1 \equiv Y$ has the same distribution as in Example 1, denoted here by H_1. The asset value $A = -\tilde{X}_2$ follows a $LogNormal(5.894, 0.1^2)$ distribution. The distribution of $\tilde{X}_2$ is denoted by H_2. For those parameters it is $\mathbb{E}(\tilde{X}_1 + \tilde{X}_2) = 0$, such that assets match liabilities on average.

Z_i, V. While V appears in all model inputs, $Z_1, \ldots, Z_d$ are independent of each other.

The dependence between $\tilde{X}_1$ and $\tilde{X}_2$ is modelled via a Clayton survival copula, with Kendall rank correlation coefficient of $\tau = 0.5$. This is the copula of a bivariate Pareto distribution with parameter $\alpha = 0.5(1/\tau - 1) = 0.5$ (Section 5.4 in Ref.[32]). For this model, we can express $\tilde{X}_1, \tilde{X}_2$ as non-decreasing functions of the bivariate Pareto random variables X_1', X_2', defined by $X_1' = D_1 V',\ X_2' = D_2 V'$, where D_1 and D_2 follow exponential distributions with mean 1, $1/V'$ is Gamma distributed with shape parameter equal to $\alpha = 0.5$ and scale parameter 1, and the random variables D_1, D_2, V' are independent (Section 7.2.4 in Ref.[30]). It is evident that (X_1', X_2'), and therefore $(\tilde{X}_1, \tilde{X}_2)$ are independent conditional on V'. Because of the simple multiplicative structure, there is no need to invoke the construction of Lemma 10.

Consistently with previous arguments, we let the random variables $Z_1 \sim H_1,\ Z_2 \sim H_2$ and $V \sim Unif[0,1]$ be non-decreasing functions of D_1, D_2, V' respectively. Thus the independent random vector (Z_1, Z_2) has the same marginal distributions as $(\tilde{X}_1, \tilde{X}_2)$. A uniform distribution is chosen for the common factor V, since the latent variable V' lacks a natural interpretation in the present application.

For a simulated sample of 20000 observations, the sensitivities to $Z_1,\ Z_2$ and V are calculated. The gradients of the functions $\psi^{(1)}, \psi^{(2)}$ are once more calculated numerically by local linear regression (we treat the gradient of $\tilde{g}$ as given). A pictorial summary of the resulting analysis is provided in Figure 3. In panel a) a scatter plot between $\tilde{X}_1$ and $\tilde{X}_2$ is given showing the very strong dependence for high output values, representing joint adverse events. In panels b), c) and d), estimates of the functions

$$u \mapsto \mathbb{E}\left(\big(V - \mathbb{E}(V)\big)\big(\psi_2^{(1)}(Z_1, V) + \psi_2^{(2)}(Z_2, V)\big) | U_{\tilde{Y}} = u\right);$$
$$u \mapsto \mathbb{E}\left(\big(Z_1 - \mathbb{E}(Z_1)\big)\psi_1^{(1)}(Z_1, V) | U_{\tilde{Y}} = u\right);$$
$$u \mapsto \mathbb{E}\left(\big(Z_2 - \mathbb{E}(Z_2)\big)\psi_1^{(2)}(Z_2, V) | U_{\tilde{Y}} = u\right);$$

are plotted respectively. It is seen how the sensitivity to the common factor V dramatically

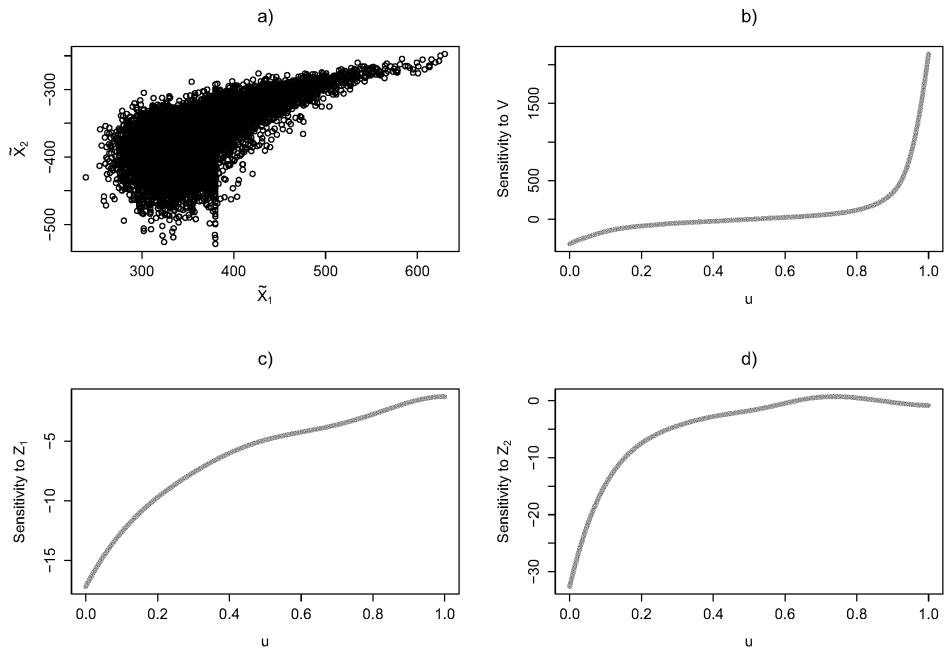

Figure 3: Analysis of the sensitivity to the dependence structure. a) Scatter plot of $\tilde{X}_1, \tilde{X}_2$; b) Sensitivity to common model input V; c) Sensitivity to idiosyncratic model input Z_1; d) Sensitivity to idiosyncratic model input Z_2.

dominates those of the idiosyncratic factors Z_1 and Z_2 for high confidence levels u. In particular, at those levels the sensitivities to Z_1, Z_2 are approaching zero, supporting the intuitive view that scenarios of extreme output are "driven by dependence". □

5 CONCLUSIONS

A framework for global sensitivity measurement using directional derivatives of risk measures has been developed. We focused on the class of distortion risk measures, while our arguments also work for moment-based risk measures, which may be preferred in some applications.

The proposed sensitivity analysis method is applicable to Monte-Carlo samples of non-linear models, augmented with the use of local regression for numerical gradient (i.e. local sensitivity)

estimation, if an analytical expression for the gradient is unavailable or hard to work with. It was shown through examples drawn from insurance loss modelling that informative percentile sensitivity curves can be produced as outputs of the sensitivity analysis exercise, aiding a risk analyst in understanding the ways in which inputs and output interact. In this both the modelled statistical behaviour of model inputs and the non-linearity of the model function are reflected.

Finally, it was demonstrated how the sensitivity measure can be extended to cope with structural parameters, uncertain statistical parameters and common factors driving dependence.

APPENDIX: FORMAL STATEMENTS AND PROOFS

A PROOF OF PROPOSITION 2

The technical conditions required for Proposition 2 include differentiability of g (condition i)), which will not always hold. However, in applications model function g will need to be estimated. In such a context, for reasons of tractability a smooth model is typically used to approximate g – this process was demonstrated in Section 3. Moreover, when the model function is continuous and non-differentiable at only a countable number of points, the proof of Proposition 2 still holds. Condition ii) is relatively weaker; for example any model function that can be written as $g(\mathbf{X}) = g^{(1)}(X_k) + g^{(2)}(\mathbf{X}_{-k})$, where $g^{(1)}$ is an invertible function and $X_k|\mathbf{X}_{-k}$ is continuous, will satisfy this. Condition iii) is implied from i) and ii) plus some integrability conditions that we do not explicitly state, as they are similar to those in[24].

First, in Lemma 3 the derivative of the percentile function with respect to a proportional shock on X_i is worked out. The proof is a direct generalization of an argument given by Tasche[24] in the case of linear models. A related result, under somewhat different assumptions, is obtained by Hong[25].

Lemma 3. *Consider the random vector* $\mathbf{X} = (X_1, \ldots, X_d)$ *taking values in* $\mathcal{X}$*, such that* X_1 *has*

a conditional density given $(X_2, \ldots, X_d)$*, denoted* $\phi(\cdot|x_2, \ldots, x_d)$*. Let the function* $g : \mathcal{X} \to \mathbb{R}$ *be differentiable in the* i^{th} *argument with* $\frac{\partial g(\mathbf{x})}{\partial x_i} = g_i(\mathbf{x})$ *and invertible with respect to* x_1 *for all* $x_2, \ldots, x_d$*. Define* $Y = g(x_1, \ldots, x_d)$ *and* $Y_{i,t} = g(X_1, \ldots, tX_i, \ldots, X_d)$ *and the corresponding quantile function* $\eta_{i,u}(t) = F^{-1}_{Y_{i,t}}(u)$*. If* $\eta_{i,u}(t)$ *is differentiable at* $t = 1$ *for given* $u \in (0, 1)$*, then*

$$\eta'_{i,u}(1) = \mathbb{E}\left(X_i g_i(X) | g(X) = \eta_{i,u}(1)\right).$$

Proof. Drop the subscript in $\eta_{i,u}$, denote by l the inverse of g (assumed increasing) with respect to the first argument and let $\mathbf{X}_{-1} = (X_2, X_3, \ldots, X_d)$.

Let us show first that the density of Y is

$$y \to f_Y(y) = \mathbb{E}\left(\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}) l_1(y, \mathbf{X}_{-1})\right) =$$

$$= \mathbb{E}\left(\frac{\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1})}{g_1(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1})}\right).$$

For a given nonnegative function k, we have

$$\begin{aligned}
\mathbb{E}(k(Y)|\mathbf{X}_{-1}) &= \mathbb{E}(k(g(X_1, \mathbf{X}_{-1}))|\mathbf{X}_{-1}) \\
&= \int_{-\infty}^{+\infty} k(g(x_1, \mathbf{X}_{-1}))\phi(x_1, \mathbf{X}_{-1})\mathrm{d}x_1 \\
&= \int_{-\infty}^{+\infty} k(y)\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}) \frac{1}{g_1(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1})} \mathrm{d}x_1,
\end{aligned}$$

after making the change of variable $g(x_1, \mathbf{X}_{-1}) = y$ and observing that $\mathrm{d}x_1 = \mathrm{d}l(y, \mathbf{X}_{-1}) = \frac{1}{g_1(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1})}$. It follows that the conditional density of Y given $\mathbf{X}_{-1}$ is

$$y \to \phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}) \frac{1}{g_1(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1}}.$$

From this we immediately deduce the unconditional density of Y.

Consider the case $i = 1$. Using the definition of quantile we have

$$u = \mathbb{P}(g(Y_{1,t}) \leq \eta(t))$$

so that, using l and conditioning on $\mathbf{X}_{-1}$, this becomes

$$u = \mathbb{E}\left(\mathbb{P}\left(X_1 \leq \frac{1}{t} l(\eta(t), \mathbf{X}_{-1}) | \mathbf{X}_{-1}\right)\right) =$$
$$= \mathbb{E}\left(\int_{-\infty}^{l(\eta(t),\mathbf{X}_{-1})/t} \phi(x_1|\mathbf{X}_{-1})\mathrm{d}x_1\right).$$

Taking derivatives with respect to t, exchanging the derivative with the expectation and letting then $t = 1$, gives

$$0 = \mathbb{E}\left(\phi(l(\eta(1), \mathbf{X}_{-1})|\mathbf{X}_{-1})\left\{l_1(\eta(1), \mathbf{X}_{-1})\eta'(1) - l(\eta(1), \mathbf{X}_{-1})\right\}\right).$$

Solving with respect to $\eta'(1)$ we get

$$\eta'(1) = \frac{\mathbb{E}\left(l(\eta(1), \mathbf{X}_{-1})\phi(l(\eta(1), \mathbf{X}_{-1})|\mathbf{X}_{-1})\right)}{\mathbb{E}\left(l_1(\eta(1), \mathbf{X}_{-1})\phi(l(\eta(1), \mathbf{X}_{-1})|\mathbf{X}_{-1})\right)}.$$

Now, let us compute the conditional expectation $\mathbb{E}(X_1 g_1(X)|Y)$. For any nonnegative function k, we have

$$\begin{aligned}
\mathbb{E}(X_1 g_1(X)k(Y)) &= \mathbb{E}(\mathbb{E}(X_1 g_1(X_1, \mathbf{X}_{-1})k(Y))|\mathbf{X}_{-1}) \\
&= \mathbb{E}\left(\int_{-\infty}^{+\infty} x_1 g_1(x_1, \mathbf{X}_{-1})k(g(x_1, \mathbf{X}_{-1}))\phi(x_1|\mathbf{X}_{-1})\mathrm{d}x_1\right) \\
&= \mathbb{E}\left(\int_{-\infty}^{+\infty} l(y, \mathbf{X}_{-1})k(y)\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1})\mathrm{d}x_1\right) \\
&= \int_{-\infty}^{+\infty} k(y)\frac{\mathbb{E}(l(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}{f_Y(y)} f_Y(y)\mathrm{d}y \\
&= \mathbb{E}(k(Y)q(Y)),
\end{aligned}$$

with

$$q(y) = \frac{\mathbb{E}(l(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}{f_Y(y)} = \frac{\mathbb{E}(l(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}{\mathbb{E}(l_1(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}$$

so we conclude that

$$\mathbb{E}(X_1 g_1(X_1, \mathbf{X}_{-1})|Y) = q(Y),$$

and the claim follows.

Let us work out the case $i \neq 1$. Using the definition of quantile we have

$$u = \mathbb{P}(g(Y_{i,t}) \leq \eta(t))$$

so that, using l, conditioning on $\mathbf{X}_{-1}$, and denoting $X_{-1,t} = (X_2, \ldots, tX_i, \ldots, X_d)$, this becomes

$$u = \mathbb{E}\left(\mathbb{P}\left(X_1 \leq l(\eta(t), X_{-1,t})|\mathbf{X}_{-1}\right)\right) =$$

$$= \mathbb{E}\left(\int_{-\infty}^{l(\eta(t),X_{-1,t})} \phi(x_1|\mathbf{X}_{-1})\mathrm{d}x_1\right).$$

Taking derivatives with respect to t, exchanging the derivative with the expectation and letting then $t = 1$, gives

$$0 = \mathbb{E}\left(\phi(l(\eta(1), \mathbf{X}_{-1})|\mathbf{X}_{-1})\left\{l_1(\eta(1), \mathbf{X}_{-1})\eta'(1) + X_i l_i(\eta(1), \mathbf{X}_{-1})\right\}\right).$$

Solving with respect to $\eta'(1)$ we get

$$\eta'(1) = -\frac{\mathbb{E}\left(X_i l_i(\eta(1), \mathbf{X}_{-1})\phi(l(\eta(1), \mathbf{X}_{-1})|\mathbf{X}_{-1})\right)}{\mathbb{E}\left(l_1(\eta(1), \mathbf{X}_{-1})\phi(l(\eta(1), \mathbf{X}_{-1})|\mathbf{X}_{-1})\right)}.$$

Now, let us compute the conditional expectation $\mathbb{E}(X_i g_i(X)|Y)$. For any nonnegative function k, we have

$$\begin{aligned}
\mathbb{E}(X_i g_i(X)k(Y)) &= \mathbb{E}(X_i \mathbb{E}(g_i(X_1, \mathbf{X}_{-1})k(Y))|\mathbf{X}_{-1}) \\
&= \mathbb{E}\left(X_i \int_{-\infty}^{+\infty} g_i(x_1, \mathbf{X}_{-1})k(g(x_1, \mathbf{X}_{-1}))\phi(x_1|\mathbf{X}_{-1})\mathrm{d}x_1\right) \\
&= \mathbb{E}\left(X_i \int_{-\infty}^{+\infty} \frac{g_i(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1})k(y)\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1})}{g_1(l(y, \mathbf{X}_{-1}, \mathbf{X}_{-1}))}\mathrm{d}x_1\right) \\
&= -\int_{-\infty}^{+\infty} k(y)\frac{\mathbb{E}(X_i l_i(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}{f_Y(y)} f_Y(y)\mathrm{d}y \\
&= \mathbb{E}(k(Y)q(Y)),
\end{aligned}$$

(we have used the fact that $g_i(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1}) = -g_1(l(y, \mathbf{X}_{-1}), \mathbf{X}_{-1})l_i(y, \mathbf{X}_{-1})$) with

$$q(y) = -\frac{\mathbb{E}(X_i l_i(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}{f_Y(y)} = -\frac{\mathbb{E}(X_i l_i(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}{\mathbb{E}(l_1(y, \mathbf{X}_{-1})\phi(l(y, \mathbf{X}_{-1})|\mathbf{X}_{-1}))}$$

so we conclude that

$$\mathbb{E}(X_i g_i(X_1, \mathbf{X}_{-1})|Y) = q(Y),$$

and the claim follows. □

The proof of Proposition 1 now follows. From noting that

$$\begin{aligned}\frac{d}{dt}\rho(g(X_1,\ldots,tX_i,\ldots,X_d))\Big|_{t=1} &= \int_0^1 \eta'_{i,u}(1)\zeta(u)du \\ &= \int_0^1 \mathbb{E}\big(X_i g_i(X)|Y = F_Y^{-1}(u)\big)\zeta(u)du \\ &= \int_0^1 \mathbb{E}\big(X_i g_i(X)|U_Y = u\big)\zeta(u)du,\end{aligned}$$

the stated formula for $S_i(\mathbf{X}, g)$ follows.

Now all is left is to derive the slightly more general formula for $S_i(Z; \mathbf{X}, g)$. Define the function $\tilde{g} : \mathbb{R}^{d+1} \mapsto \mathbb{R}$ such that

$$\tilde{g}(v_1,\ldots,v_i,v_{i+1},\ldots,v_{d+1}) = g(v_1,\ldots,v_i+v_{i+1},\ldots,v_{d+1}).$$

Consider the $(d+1)$-dimensional random vector

$$\tilde{\mathbf{X}} = (X_1,\ldots,X_{i-1},X_i - Z, Z, X_{j+1},\ldots,X_d),$$

such that $\tilde{g}(\tilde{\mathbf{X}}) = g(\mathbf{X}) = Y$ and

$$\tilde{g}_{i+1}(\tilde{\mathbf{X}}) = g_i(X_1,\ldots,X_{j-1},X_i - Z + Z, X_{j+1},\ldots,X_d) = g_i(\mathbf{X}).$$

Then it is apparent that $S_i(Z; \mathbf{X}, g) = S_{i+1}(\tilde{\mathbf{X}}, \tilde{g})$. So the problem is reduced to finding the sensitivity of $\rho(\tilde{g}(\tilde{\mathbf{X}}))$ to a proportional stress on the $(i+i)^{th}$ element of $\tilde{\mathbf{X}}$. Consequently,

$$S_{i+1}(\tilde{\mathbf{X}}, \tilde{g}) = \mathbb{E}\big(Z\tilde{g}_{i+1}(\tilde{\mathbf{X}})\zeta(U_{\tilde{g}(\tilde{\mathbf{X}})})\big) = \mathbb{E}\big(Z g_i(\mathbf{X})\zeta(U_{g(\mathbf{X})})\big).$$

B FORMAL PROPERTIES OF THE SENSITIVITY MEASURE

Before proceeding, some fundamental dependence concepts are recalled. A random vector $\mathbf{X}$ is *comonotonic* if

$$\mathbf{X} \stackrel{d}{=} (h_1(U), \ldots, h_d(U)), \tag{B.1}$$

where U is a random variable and $h_1, \ldots, h_d$ are non-decreasing functions[48]. Comonotonicity is an extremal form of positive dependence, whereby all model inputs are non-decreasing functions of one another. Note that for any distortion risk measure, comonotonicity of $\mathbf{X}$ implies $\rho(\sum_{j=1}^{d} X_i) = \sum_{j=1}^{d} \rho(X_i)$. For details see Ref.[30], Property 2.6.5. A weaker positive dependence property is association. The random vector $\mathbf{X}$ is *associated* if, for any non-decreasing functions $h_1, h_2 : \mathcal{X} \mapsto \mathbb{R}$, such that the expectations below exist, it is

$$\mathbb{E}\big(h_1(\mathbf{X})h_2(\mathbf{X})\big) \geq \mathbb{E}\big(h_1(\mathbf{X})\big)\mathbb{E}\big(h_2(\mathbf{X})\big). \tag{B.2}$$

Association of $\mathbf{X}$ holds for several standard dependence structures, including independence, comonotonicity, and dependence by mixture (Ref.[30], Sections 7.2.3-7.2.4).

Lemma 4. *Let $g(\mathbf{x})$ be convex in the i^{th} argument for all $\mathbf{x}_{-i}$ and ζ be non-decreasing. Then*

$$\rho(g(\mathbf{X} + \epsilon \mathbf{e}_i Z)) - \rho(g(\mathbf{X})) \geq \epsilon S_i(Z; \mathbf{X}, g).$$

Proof. The stated result follows from

$$\begin{aligned}
\rho(g(\mathbf{X} + \epsilon Z \mathbf{e}_i)) - \rho(g(\mathbf{X})) &= \mathbb{E}\big(g(\mathbf{X} + \epsilon Z \mathbf{e}_i)\zeta(U_{g(\mathbf{X}+\epsilon Z \mathbf{e}_i)})\big) - \mathbb{E}\big(g(\mathbf{X})\zeta(U_{g(\mathbf{X})})\big) \\
&\geq \mathbb{E}\Big(g\big(\mathbf{X} + \epsilon Z \mathbf{e}_i\big) - g(\mathbf{X})\big)\zeta(U_{g(\mathbf{X})})\Big) \\
&\geq \mathbb{E}\big(\epsilon Z g_i(\mathbf{X})\zeta(U_{g(\mathbf{X})})\big).
\end{aligned}$$

The first inequality above follows from Proposition 6.2.6 in Ref.[30], after noting that the two random vectors $\big(g(\mathbf{X}+\epsilon Z \mathbf{e}_i), \zeta(U_{g(\mathbf{X}+\epsilon Z \mathbf{e}_i)})\big)$ and $\big(g(\mathbf{X}+\epsilon Z \mathbf{e}_i), \zeta(U_{g(\mathbf{X})})\big)$ have the same marginal

distributions but the former is comonotonic (both its elements are non-decreasing functions of $U_{g(\mathbf{X}+\epsilon Z\mathbf{e}_i)}$). The second inequality follows from the convexity of g in the i^{th} argument. □

The proof of Lemma 4 makes no use of differentiability of $\rho(g(\mathbf{X}+\epsilon Z\mathbf{e}_i))$; indeed g_i can stand for any subderivative of $x_i \mapsto g(\mathbf{x})$. It follows that $\mathbb{E}\big(Zg_i(\mathbf{X})\zeta(U_{g(\mathbf{X})})\big)$ is a subderivative of $\epsilon \mapsto \rho(g(\mathbf{X}+\epsilon Z\mathbf{e}_i))$, without reference to the technical conditions of Proposition 2.

The function g is *directionally convex* if, for any $\mathbf{x}, \mathbf{x}', \mathbf{z}$ such that $\mathbf{x} \leq \mathbf{x}'$ and $\mathbf{z} \geq \mathbf{0}$, it is $g(\mathbf{x}+\mathbf{z}) - g(\mathbf{x}) \leq g(\mathbf{x}'+\mathbf{z}) - g(\mathbf{x}')$ (Ref.[30], Def. 3.4.62). If g is twice differentiable, this condition is equivalent to all second derivatives g_{ij} being non-negative.

Lemma 5. *Assume that ζ is non-decreasing, g is non-decreasing in each argument and directionally convex, and $\mathbf{X}$ is associated. Then $S_i^{\mathcal{D}}\big(\mathbf{X}, g\big) \geq 0$. If in addition it is $\mathbb{E}(X_i) \geq 0$, then $S_i\big(\mathbf{X}, g\big) \geq 0$.*

Proof. Let $h_1(\mathbf{x}) = x_i$. As the random variable U_Y is comonotonic to $Y = g(\mathbf{X})$, we can write $U_Y = \xi(Y)$, where ξ is some non-decreasing function. By the stated assumptions on ζ and g, the function $h_2(\mathbf{x}) = g_i(\mathbf{x})\zeta(\xi(g(\mathbf{x})))$ is non-decreasing. Therefore, by the association of $\mathbf{X}$,

$$\mathbb{E}\big(X_i g_i(\mathbf{X})\zeta(U_Y)\big) = \mathbb{E}\big(h_1(\mathbf{X})h_2(\mathbf{X})\big) \geq \mathbb{E}\big(h_1(\mathbf{X})\big)\mathbb{E}\big(h_2(\mathbf{X})\big) = \mathbb{E}\big(X_i\big)\mathbb{E}\big(g_i(\mathbf{X})\zeta(U_Y)\big),$$

from which $S_i^{\mathcal{D}}(\mathbf{X}, g) \geq 0$ follows. For the last statement, it is enough to note that $S_i(\mathbf{X}, g) = S_i^{\mathcal{M}}(\mathbf{X}, g) + S_i^{\mathcal{D}}(\mathbf{X}, g)$ and $\mathbb{E}(X_i) \geq 0 \implies S_i^{\mathcal{M}}(\mathbf{X}, g) \geq 0$. □

For the rest of this section, we express each model input as $X_i = \psi^{(i)}(\tilde{\mathbf{X}}_i)$, where $\tilde{\mathbf{X}}_i = (\tilde{X}_{i1}, \ldots, \tilde{X}_{id_i})$, for differentiable functions $\psi^{(i)} : \mathbb{R}^{d_i} \mapsto \mathbb{R},\ i = 1, \ldots, d$. The $(d_1 + \cdots + d_d)$-dimensional vector of all *granular* model inputs is denoted as $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_d)$. The model function at the lower hierarchical level is then defined by $\tilde{g}(\tilde{\mathbf{x}}) = g\big(\psi^{(1)}(\tilde{\mathbf{x}}_1), \ldots, \psi^{(d)}(\tilde{\mathbf{x}}_d)\big) = g(\mathbf{x})$. The sensitivity to $\tilde{X}_{ij}$ is then $S_{ij}(\tilde{\mathbf{X}}, \tilde{g}) \equiv S_{\sum_{k=1}^{i-1} d_k + j}(\tilde{\mathbf{X}}, \tilde{g})$. For a hierarchical structure with more layers, notation borrowed from graph theory needs to be used[49].

Lemma 6. *The sensitivity of $\rho(\tilde{g}(\tilde{\mathbf{X}}))$ to X_{ij} is given by $S_{ij}(\tilde{\mathbf{X}}, \tilde{g}) = S_i(\tilde{X}_{ij}\psi_j^{(i)}(\tilde{\mathbf{X}}_i); \mathbf{X}, g)$.*

Proof. It is $\tilde{g}_{ij}(\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_d) = \frac{\partial}{\partial \tilde{x}_{ij}} \tilde{g}(\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_d) = g_i\big(\psi^{(1)}(\tilde{\mathbf{x}}_1), \ldots, \psi^{(d)}(\tilde{\mathbf{x}}_d)\big)\psi_j^{(i)}(\tilde{\mathbf{x}}_i)$. By Proposition 2, it therefore is $S_{ij}(\tilde{\mathbf{X}}, \tilde{g}) = \mathbb{E}\big(\tilde{X}_{ij} g_i(\mathbf{X})\psi_j^{(i)}(\tilde{\mathbf{X}}_i)\zeta(U_Y)\big)$, from which the stated formula directly follows. □

Note that the structural information required to apply Lemma 6 is weaker than separability of g. The set-up of Lemma 6 implicity assumes the absence of common factors in the functions $\psi^{(1)}, \ldots, \psi^{(d)}$. However, assessment of the sensitivity to such common factors presents no complications. Let $\tilde{\mathbf{X}}_i = (V, \tilde{X}_{i2}, \ldots, \tilde{X}_{id_i})$, $i = 1, \ldots, d$, where V is a common model input. Define $g^*(v, \tilde{x}_{12}, \ldots, \tilde{x}_{1d_1}, \ldots, \tilde{x}_{d2}, \ldots, \tilde{x}_{dd_d}) = g(\psi^{(1)}(v, \tilde{x}_{12}, \ldots, \tilde{x}_{1d_1}), \ldots, \psi^{(d)}(v, \tilde{x}_{d2}, \ldots, \tilde{x}_{dd_d}))$ and let $\mathbf{X}^* = (V, \tilde{X}_{12}, \ldots, \tilde{X}_{1d_1}, \ldots, \tilde{X}_{d2}, \ldots, \tilde{X}_{dd_d})$, such that $g^*(\mathbf{X}^*) = g(\mathbf{X})$.

Lemma 7. *It is $S_1(\mathbf{X}^*, g^*) = \sum_{i=1}^d S_i(V\psi_1^{(i)}(\tilde{\mathbf{X}}_i); \mathbf{X}, g)$.*

Proof. Follows directly from total differentiation of $g(\psi^{(1)}(v, \tilde{x}_{12}, \ldots, \tilde{x}_{1d_1}), \ldots, \psi^{(d)}(v, \tilde{x}_{d2}, \ldots, \tilde{x}_{dd_d}))$ with respect to v. □

Corollary 8. *Let $d_1 = \cdots = d_d = 1$, such that $X_i = \psi^{(i)}(\tilde{X}_i)$. Then the following hold:*

i) $S_i(\tilde{\mathbf{X}}, \tilde{g}) = S_i(\tilde{X}_i\psi_1^{(i)}(\tilde{X}_i); \mathbf{X}, g)$.

ii) For $\psi^{(i)}(\tilde{x}_i) = \lambda\tilde{x}_i$, $\lambda \neq 0$, $S_i(\tilde{\mathbf{X}}, \tilde{g}) = S_i(\mathbf{X}, g)$.

iii) For $\psi^{(i)}(\tilde{x}_i) = \kappa + \lambda\tilde{x}_i$, $\lambda \neq 0$, $S_i^{\mathcal{D}}(\tilde{\mathbf{X}}, \tilde{g}) = S_i^{\mathcal{D}}(\mathbf{X}, g)$.

C VARIABILITY DECOMPOSITION IN MIXTURES

In the following, continuity of distributions is assumed. This assumption can be relaxed by employing a generalized distributional transform[31].

Lemma 9. *Consider random variables X and Θ, such that the conditional distribution of X given $\Theta = \theta$ is $F(\cdot;\theta)$.*

i) For any uniform random variable U independent of Θ, it is

$$X \stackrel{d}{=} F^{-1}(U;\Theta).$$

ii) Let $F(x;\theta)$ be increasing in x for all θ and define $U = F(X;\Theta)$. Then U is uniform and independent of Θ.

Proof. For part i) write $\mathbb{P}(F^{-1}(U;\Theta) \leq x) = \mathbb{E}(\mathbb{P}(F^{-1}(U;\Theta) \leq x|\Theta))$. The random variable $F^{-1}(U;\theta)$ has distribution $F(\cdot;\theta)$ for all θ. Therefore $\mathbb{P}(F^{-1}(U;\Theta) \leq x|\Theta = \theta) = F(x;\theta)$. Consequently $\mathbb{E}(\mathbb{P}(F^{-1}(U;\Theta) \leq x|\Theta)) = \mathbb{E}(F(x;\Theta))$, which is the (unconditional) distribution of X.

For part ii), it is enough to note that $\mathbb{P}(U \leq u|\Theta = \theta) = \mathbb{P}(X \leq F^{-1}(u;\Theta)|\Theta = \theta) = u$, which does not depend on θ. □

Lemma 10. *Consider the d-dimensional random vector $\mathbf{X}$ and a random vector $\mathbf{V}$, such that the conditional distribution of X_i given $\mathbf{V} = \mathbf{v}$ is $F_i(\cdot;\mathbf{v})$, and let $\mathbf{X}$ be conditionally independent given $\mathbf{V}$.*

i) For any uniform random variables $U_1, \ldots, U_d$ that are independent of each other and of $\mathbf{V}$, it is

$$\mathbf{X} \stackrel{d}{=} (F_1^{-1}(U_1;\mathbf{V}), \ldots, F_d^{-1}(U_d;\mathbf{V})).$$

ii) Let $F_i(x_i;\mathbf{v})$ be increasing in x_i for all $\mathbf{v}$ and define $U_i = F(X_i;\mathbf{V})$, for $i = 1, \ldots, d$. Then $U_1, \ldots, U_d$ are uniform, independent of each other and of $\mathbf{V}$.

Proof. For part i) write

$$\begin{aligned}\mathbb{P}(F_1^{-1}(U_1;\mathbf{V}) \le x_1, \ldots, F_d^{-1}(U_d;\mathbf{V}) \le x_d) &= \mathbb{E}(\mathbb{P}(F_1^{-1}(U_1;\mathbf{V}) \le x_1, \ldots, F_d^{-1}(U_d;\mathbf{V}) \le x_d|\mathbf{V})) \\ &= \mathbb{E}\left(\prod_{i=1}^{d} \mathbb{P}(F_i^{-1}(U_i;\mathbf{V}) \le x_i|\mathbf{V})\right) \\ &= \mathbb{E}\left(\prod_{i=1}^{d} F_i(x_i;\mathbf{V})\right) \\ &= \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d),\end{aligned}$$

where the third equality follows from Lemma 9 i) and the fourth from the conditional independence of $\mathbf{X}$.

For part ii), we have

$$\begin{aligned}\mathbb{P}(U_1 \le u_1, \ldots, U_d \le u_d|\mathbf{V}) &= \mathbb{P}(X_1 \le F_1^{-1}(u_1;\mathbf{V}), \ldots, X_d \le F_d^{-1}(u_d;\mathbf{V})|\mathbf{V}) \\ &= \prod_{i=1}^{d} \mathbb{P}(X_i \le F_i^{-1}(u_i;\mathbf{V})|\mathbf{V}) = \prod_{i=1}^{d} u_i,\end{aligned}$$

where the second equality follows from the conditional independence of $\mathbf{X}$. □

References

[1] R. L. Iman and J. C. Helton. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Analysis*, 8(1):71–90, 1988.

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

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

[4] J. C. Helton and F. J. Davis. Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis*, 22(3):591–622, 2002.

[5] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis. The Primer*. Wiley, 2008.

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

[7] A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models. *Journal of the American Statistical Association*, 97(459):702–709, 2002.

[8] M. Ellouze, J.-P. Gauchi, and J.-C. Augustin. Global sensitivity analysis applied to a contamination assessment model of listeria monocytogenes in cold smoked salmon at consumption. *Risk Analysis*, 30(5):841–852, 2010.

[9] B. Anderson, E. Borgonovo, M. Galeotti, and R. Roson. Uncertainty in climate change modeling: Can global sensitivity analysis be of help? *Risk Analysis*, 34(2):271–293, 2014.

[10] L. A. Cox. Why risk is not variance: An expository note. *Risk Analysis*, 28(4):925–928, 2008.

[11] E. Borgonovo, W. Castaings, and S. Tarantola. Moment independent importance measures: New results and analytical test cases. *Risk Analysis*, 31(3):404–428, 2011.

[12] P. Wei, Z. Lu, and J. Song. Moment-independent sensitivity analysis using copula. *Risk Analysis*, 34(2):210–222, 2014.

[13] X. Xu, Z. Lu, and X. Luo. A stable approach based on asymptotic space integration for moment-independent uncertainty importance measure. *Risk Analysis*, 34(2):235–251, 2014.

[14] A. Tsanakas and E. Desli. Measurement and pricing of risk in insurance markets. *Risk Analysis*, 25(6):1653–1668, 2005.

[15] C. A. MacKenzie. Summarizing risk using risk measures and risk indices. *Risk Analysis*, 2014. doi: 10.1111/risa.12220.

[16] S. S. Wang, V. R. Young, and H. H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183, 1997.

[17] C. Acerbi. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518, 2002.

[18] J. Belles-Sampera, M. Guillén, and M. Santolino. Beyond Value-at-Risk: GlueVaR distortion risk measures. *Risk Analysis*, 34(1):121–134, 2014.

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

[20] E. Borgonovo and L. Peccati. Finite change comparative statics for risk-coherent inventories. *International Journal of Production Economics*, 131(1):52–62, 2011.

[21] C. Gourieroux, J.P. Laurent, and O. Scaillet. Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7(3-4):225 – 245, 2000.

[22] D. Tasche. Allocating portfolio economic capital to sub-portfolios. In A. Dev (ed.), Economic Capital: A Practitioner's Guide, Risk Books, pp. 275–302, 2004.

[23] A. Tsanakas. To split or not to split: Capital allocation with convex risk measures. *Insurance: Mathematics and Economics*, 44(2):268–277, 2009.

[24] D. Tasche. Conditional expectation as quantile derivative, 2000. URL `http://arxiv.org/pdf/math/0104190v1.pdf`.

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

[26] R. C. Spear and G. M. Hornberger. Eutrophication in peel inlet – II. identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, 14(1):43–49, 1980.

[27] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[28] C. B. Storlie and J. C. Helton. Multiple predictor smoothing methods for sensitivity analysis: Description of techniques. *Reliability Engineering and System Safety*, 93(1):28 – 54, 2008.

[29] Q. Li and J. Racine. Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14:485–512, 2004.

[30] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial Theory for Depedent Risks*. Wiley, 2005.

[31] L. Rüschendorf. On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.

[32] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press, 2005.

[33] European Insurance and Occupational Pensions Authority. Directive 2009/138/EC of the European Parliament and of the Council, 2009.

[34] Federal Office of Private Insurance. Technical document on the Swiss Solvency Test, 2006.

[35] P. Gustafson. Local sensitivity of posterior expectations. *The Annals of Statistics*, 24(1): 174–195, 1996.

[36] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. Wiley, 2011.

[37] I. M. Sobol' and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10): 3009–3017, 2009.

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

[39] R. C. Spear, T. M. Grieb, and N. Shang. Parameter uncertainty and interaction in complex environmental models. *Water Resources Research*, 30(11):3159–3169, 1994.

[40] O. O. Osidele and M. B. Beck. An inverse approach to the analysis of uncertainty in models of environmental systems. *Integrated Assessment*, 4(4):265–282, 2004.

[41] E. Borgonovo and G. E. Apostolakis. A new importance measure for risk-informed decision making. *Reliability Engineering & System Safety*, 72(2):193–212, 2001.

[42] E. Borgonovo. Sensitivity analysis of model output with input constraints: A generalized rationale for local methods. *Risk analysis*, 28(3):667–680, 2008.

[43] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[44] D. Rosen and D. Saunders. Risk factor contributions in portfolio credit risk models. *Journal of Banking and Finance*, 34(2):336–349, 2010.

[45] J. Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004, 1992.

[46] D. Ruppert and M. P. Wand. Multivariate locally weighted least squares regression. *Annals of Statistics*, 22(3):1346–1370, 1994.

[47] T. Hayfield and J. S. Racine. Nonparametric econometrics: The NP package. *Journal of Statistical Software*, 27(5), 2008.

[48] J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31:3–33, 2002.

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