# City Research Online

## City St George's, University of London

**Citation:** Pesenti, S. M. (2018). Robustness and sensitivity of risk evaluations. (Unpublished Doctoral thesis, Cass Business School, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** https://openaccess.city.ac.uk/id/eprint/21472/

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

# Robustness and Sensitivity of Risk Evaluations

Silvana Manuela Pesenti

*A thesis submitted in fulfilment of the requirements*

*for the degree of*

*Doctor of Philosophy*

Cass Business School, City, University of London

Department of Actuarial Science and Insurance

Supervisor: Dr. Andreas Tsanakas

Dr. Pietro Millossovich

December 2018

*All men by nature desire to know.*

Aristotle

# Table of contents

**List of tables** ix

**List of figures** xi

**1 Introduction** 1

1.1 Sensitivity analysis . . . 1

1.2 Model setup . . . 3

1.2.1 Risk measures . . . 4

1.3 Structure of the thesis . . . 6

1.3.1 Chapter 2 . . . 7

1.3.2 Chapter 3 . . . 8

1.3.3 Chapter 4 . . . 8

1.4 Direction for future research . . . 9

**2 Robustness Regions for Measures of Risk Aggregation** 11

2.1 Introduction . . . 11

2.2 Preliminaries . . . 16

2.3 Robustness . . . 18

2.3.1 Robustness of convex risk measures . . . 18

2.3.2 Robustness regions of convex risk measures . . . 21

2.3.3 Non-robustness of convex risk measures . . . 27

2.3.4 Generalisation to risk measures defined on $\mathcal{L}^p$ . . . . . 31
2.4 Aggregation . . . . . 32
2.4.1 Robustness of aggregation measures . . . . . 32
2.4.2 Aggregation robustness and linear growth . . . . . 38
2.4.3 Aggregation through compound distributions . . . . . 43
2.5 Comparison to robustness regions of Value-at-Risk . . . . . 45
Appendix 2.A Wasserstein space . . . . . 47

**3 Reverse sensitivity testing: What does it take to break the model? 51**
3.1 Introduction . . . . . 51
3.1.1 Problem framing and contribution . . . . . 51
3.1.2 Relation to the literature . . . . . 53
3.1.3 Structure of the chapter . . . . . 55
3.2 Preliminaries . . . . . 56
3.3 Deriving the stressed model . . . . . 57
3.3.1 Problem statement . . . . . 57
3.3.2 Probability constraints . . . . . 59
3.3.3 VaR constraints . . . . . 60
3.3.4 VaR and ES constraints . . . . . 65
3.3.5 Shortfall risk measure constraints . . . . . 69
3.3.6 Multivariate output . . . . . 70
3.4 Sensitivity measures for importance ranking . . . . . 72
3.4.1 Definition of sensitivity measures . . . . . 72
3.4.2 Comparison to other sensitivity measures . . . . . 78
3.4.3 Controlling for dependence in the sensitivity measure . . . . . 80
3.5 Financial application: a London Insurance Market portfolio . . . . . 82
3.6 Conclusions . . . . . 85

Appendix 3.A Stochastic comparisons . . . . . 86

Appendix 3.B Proofs . . . . . 90

**4 Cascade Sensitivity Measures** **99**

4.1 Introduction . . . . . 99

4.2 Preliminaries . . . . . 102

4.3 Sensitivity measures . . . . . 104

4.3.1 Marginal sensitivity . . . . . 104

4.3.2 Cascade sensitivity . . . . . 105

4.3.3 Comparison of the marginal and cascade sensitivity . . . . . 108

4.4 Calculation of the cascade sensitivity . . . . . 110

4.4.1 Stressing through a perturbation . . . . . 110

4.4.2 Stressing through an additive shock . . . . . 113

4.4.3 Numerical evaluation via importance sampling . . . . . 117

4.5 Application to a non-linear insurance portfolio . . . . . 119

4.5.1 The insurance portfolio . . . . . 119

4.5.2 Cascade sensitivity via importance sampling . . . . . 122

4.5.3 Cascade sensitivity via the inverse Rosenblatt transform . . . . 124

4.5.4 Influence of indirect dependence between input factors . . . . . 127

4.6 Application to a black box model . . . . . 128

4.7 Conclusion . . . . . 130

Appendix 4.A Assumptions and proofs . . . . . 131

**References** **135**

# List of tables


3.1 Optimisation (3.4) . . . . . 66

3.2 Optimisation (3.5) . . . . . 70

3.3 Optimisation (3.8) . . . . . 72

3.4 Reverse sensitivity measure . . . . . 78

3.5 Comparison to variance based sensitivity measures . . . . . 80

3.6 $k^{\text{th}}$-order reverse sensitivity measure . . . . . 82

4.1 Derivative of the inverse Rosenblatt transform . . . . . 127

4.2 Marginal and cascade sensitivity . . . . . 128

# List of figures


3.1 RN-density and output distribution of the solution to (3.4) . . . . . 64

3.2 Input distribution of the solution to (3.4) . . . . . 65

3.3 RN-density and output distribution of the solution to (3.5) . . . . . 68

3.4 Input distribution of the solution to (3.5) . . . . . 69

3.5 Output distribution of the solution to (3.8) . . . . . 71

3.6 Reverse and forward sensitivity measure . . . . . 77

3.7 London Market portfolio: Reverse and forward sensitivity measures . . 83

3.8 London Market portfolio: Reverse sensitivity against exposures . . . . . 85

3.9 RN-densities of (3.4) and (3.5) for different stress levels . . . . . 91

4.1 Marginal and cascade sensitivity of VaR . . . . . 120

4.2 Marginal and cascade sensitivity of ES . . . . . 121

4.3 Densities of the input and its distortion . . . . . 123

4.4 Importance weights . . . . . 124

4.5 Distribution of the output and its distortion . . . . . 125

4.6 London Market portfolio: Cascade sensitivity measures . . . . . 130

# Acknowledgements

This thesis significantly benefited from discussions with and critical comments from academics at Cass Business School and the conferences and workshops I attended during my PhD. In particular, I want to thank Vali Asimit, Markus Gesmann, Michael Gordy, Mike Ludkovski, Alfred Müller, Ludger Rüschendorf, Andrew Smith and Ruodu Wang for inspiring conversations and the referees of Pesenti et al. (2016, 2018*b*) for constructive feedback. A special thank you goes to my supervisors Andreas Tsanakas and Pietro Millossovich whose advice and constant support accompanied me throughout every stage of my PhD.

Last but not least I am indebted to my parents, my brother and my sisters for their endless source of inspiration.

London, October 2018 Silvana M. Pesenti

# Declaration

I hereby grant powers of discretion to the University Librarian of Cass Business School, City, University of London, to allow the thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgements.

Silvana Manuela Pesenti

December 2018

# Abstract

This thesis is a collection of three contributions to sensitivity analysis of financial and insurance risk evaluations. Sensitivity analysis constitutes an important component of model building, interpretation and validation, particularly for models whose output is at the core of a risk management decision process. We study models comprising a (random) vector of input factors, an aggregation function mapping input factors to a random output, and a risk measure applied to the output. In most typical insurance and financial applications, the model's characteristic – a non-analytical and numerically expensive aggregation function evaluated on numerous input factors – renders most sensitivity analysis methodologies unfeasible.

We develop sensitivity analysis procedures applicable specifically for the above model setting. First, we address the estimation of risk measures applied to the model output. The fundamental purpose of a risk measure is to distinguish between different risk profiles. However, strong assumptions on the risk measure's ability to distinguish risk severities lead to non-robust estimators. We provide conditions when risk measures exhibit both, robustness and a consistent ranking of risks. Second, we develop a framework termed *reverse sensitivity testing*, that associates a critical increase in the risk measure to specific input factors. We provide analytical solutions of the stressed distribution of input factors that lead to the required increase in the outputs' risk measure. Third, we introduce a novel sensitivity measure, which quantifies the extent to which the model output is affected by a stress in an individual input factor. Compared to other sensitivity measures in the literature, the proposed measure incorporates the direct impact of the stressed input as well as indirect effects via other input factors that are dependent on the one being stressed. In this way the dependence between inputs is explicitly taken into account.

# Chapter 1

# Introduction

## 1.1 Sensitivity analysis

Since the early adoption of statistical models, there is an inherent necessity to understand the underlying relationship between the model inputs and output – a task assigned to sensitivity analysis. The field of sensitivity analysis is best described as "*the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input*" (Saltelli et al., 2004). Sensitivity analysis is of fundamental importance for risk analysts, in particular, when the model output is a decision variable with far-reaching consequences, such as (indicatively) in probability safety assessment (Saltelli, 2002), reliability analysis (Aven and Nøkland, 2010), nuclear waste disposal (Saltelli and Tarantola, 2002), food safety (Frey and Patil, 2002) and in financial and insurance risk management (Gourieroux et al., 2000; Tsanakas and Millossovich, 2016).

Inspiration for this thesis is the development of a sensitivity analysis framework for specific *risk models*, such as *internal models*, used by insurance companies and financial institutions. Common feature of these risk models is that they are highly involved and sophisticated statistical models, whose output is of crucial importance to a risk

management decision process. An internal model, for example, replicates the overall risk exposure and determines the regulatory economic capital of an insurance company (Swiss Solvency Test, 2006; EIOPA, 2009). The mathematical and statistical design of these risk models, however, prevents a straightforward application of commonly used sensitivity analysis methodologies. Indeed, these models exhibit the following characteristics:

- They are evaluated on numerous input factors, which results in the importance ranking of input factors becoming a computationally non-trivial task.
- The input factors are prone to uncertainty, thus modelled using random variables. To capture statistical uncertainty in the input factors, *global sensitivity analysis*, which takes into account the whole range of the input factors, has to be employed (Saltelli, 2002).
- The input factors are mapped to the model output through an *aggregation function*, which is usually not analytically available. Thus, sensitivity analysis techniques that do not rely on restrictive properties of the aggregation functions are of central importance.
- Evaluations of the model's aggregation function, so-called *model runs*, are typically numerically time consuming. Most sensitivity analysis tools, however, are based on algorithms requiring multiple model runs. Therefore, alternative sensitivity analysis techniques without the need of extensive reevaluations of the aggregation function are called for.
- The (random) model output is summarised through a *risk measure* that assesses the level of risk severity of the output by assigning to it a real number. Risk measures are extensively used in financial risk management, however, most

sensitivity analysis techniques have not been extended to study the sensitivity of the output risk measure to random inputs.

- Since the input factors are modelled by random variables, particular interest lies in the sensitivity of the risk measure to the *random* input factors, subsuming sensitivity to input parameters.

This thesis is a collection of three contributions to sensitivity analysis, applicable to the above setting and addressing the challenges outlined above, in the areas of: *robustness of risk measurement procedures* (Hampel et al., 2011; Huber and Ronchetti, 2009), *regional sensitivity analysis* (Spear et al., 1994; Osidele and Beck, 2004) and *factor prioritisation* (Saltelli et al., 2008; Borgonovo and Plischke, 2016). Robustness of risk measurement procedures addresses the numerical calculation of a risk measure estimated from historical and / or simulated data. Regional sensitivity analysis is concerned with the sensitivity of the model output, when the output lays in a specific region. Factor prioritisation, on the other hand, is a methodology that quantifies the extend to which the model output is affected by individual input factors.

## 1.2 Model setup

We consider the standard setting of sensitivity analysis, involving a (typically complicated) function, mapping model inputs to an output that is used in a decision making process. Mathematically, we work on a measurable space $(\Omega, \mathcal{A})$ and denote by $\boldsymbol{X} = (X_1, \ldots, X_n)$ the vector of random *input factors* on $(\Omega, \mathcal{A})$. The (measurable) function $g\colon \mathbb{R}^n \to \mathbb{R}$, is called the *aggregation function*, which gives, when applied to the input factors $\boldsymbol{X}$, the one-dimensional random *output* of interest $Y = g(\boldsymbol{X})$. Throughout, we adopt the convention that large values of the output correspond to adverse realisations.

As is typical in financial risk management, the output is summarised though a *risk measure*, leading to a classification of different levels of risk severities of the output distribution (Artzner et al., 1997, 1999; Föllmer and Schied, 2011). Moments, such as the mean and standard deviation, can be seen as risk measures. In recent years, percentile-based risk measures (Acerbi, 2002) have become prominent, with the most commonly used risk measures being Value-at-Risk (VaR) and Expected Shortfall (ES). Risk measures are used extensively in financial regulation for the calculation of capital requirements, specifically VaR for European insurance companies (EIOPA, 2009) and ES for banks (BCBS, 2012, 2013). The literature on risk measures is developing fast and, subsequently, a discussion of only those risk measures of importance in this thesis is provided.

### 1.2.1 Risk measures

#### Distortion risk measures

A risk measure is a mapping $\rho\colon \mathcal{L}^1 \to \mathbb{R}$, where $\mathcal{L}^1$ is the space of integrable random variables on $(\Omega, \mathcal{A})$. That is, a risk measure associates to each integrable random variable a real number. The class of *distortion risk measures*, introduced by Wang (1996); Acerbi and Tasche (2002) and studied in Chapter 4, is defined, for a random variable $Y$ with distribution function $F_Y$, through

$$\rho_\gamma(Y) = \int_0^1 F_Y^{-1}(u)\gamma(u)\mathrm{d}u,$$

where $\gamma\colon [0,1] \to [0,\infty)$ is a normalised weight function such that $\int_0^1 \gamma(u)\mathrm{d}u = 1$. The inverse of the distribution function of $Y$ is defined by $F_Y^{-1}(u) = \inf\{y \in \mathbb{R} \mid F_Y(y) \geq u\}$, for $0 \leq u \leq 1$, and where, as usual, $\inf \varnothing = +\infty$. The widely used distortion risk measure Value-at-Risk ($\mathrm{VaR}_\alpha$) at level $\alpha \in [0,1]$ is given by the weight function

$\gamma(u) = \delta_\alpha(u)$, for the Dirac measure $\delta_\alpha$, defined by $\int_0^1 l(x)\mathrm{d}\delta_\alpha(x) = l(\alpha)$ for any function $l$ and $0 \leq \alpha \leq 1$. Alternatively, the $\mathrm{VaR}_\alpha$ can be defined as the left $\alpha$-quantile, $\mathrm{VaR}_\alpha(Y) = F_Y^{-1}(\alpha)$. In particular, the essential supremum of $Y$ is $\operatorname{ess\,sup} Y = F_Y^{-1}(1)$.

The Expected Shortfall ($\mathrm{ES}_\alpha$), also called Conditional Value-at-Risk, at level $\alpha \in [0, 1)$ is a distortion risk measure with $\gamma(u) = \frac{1}{1-\alpha}\mathbb{1}_{\{u>\alpha\}}$ and has representation

$$\mathrm{ES}_\alpha(Y) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(Y)\mathrm{d}u = \frac{1}{1-\alpha}E\left((Y - \mathrm{VaR}_\alpha(Y))_+\right) + \mathrm{VaR}_\alpha(Y).$$

Unlike VaR, the ES takes into account the whole tail of the distribution of $Y$, that is all realisations larger than $\mathrm{VaR}_\alpha(Y)$. We refer to Chapter 2 and Föllmer and Schied (2011) for a comprehensive comparison of the two risk measures.

Distortion risk measures with non-decreasing weight function $\gamma$ are called *spectral risk measures* (Acerbi and Tasche, 2002). Clearly, the ES belongs to the class of spectral risk measures while the VaR does not.

### Coherent and convex risk measures

Another approach to classify risk measures is through their properties, which reflect the ways a risk measure distinguishes different risk profiles. The literature on desirable properties of a risk measure is extensive and a complete review is outside of the scope and aim of this thesis. We only recall the definitions needed for the exposition and refer to Artzner et al. (1999); Föllmer and Schied (2011) and references therein for an exhaustive overview. Common properties of a risk measure include, for random variables $Y, Z$:

i) Law-invariance: $\rho(Y) = \rho(Z)$ if $F_Y = F_Z$.

ii) Translation invariance: $\rho(Y + m) = \rho(Y) + m$ for $m \in \mathbb{R}$.

iii) Monotonicity: $\rho(Y) \leq \rho(Z)$ if $Y \leq Z$, $\mathbb{P}$-a.s.

iv) Convexity: $\rho\Big((1-\lambda)Y + \lambda Z\Big) \leq (1-\lambda)\rho(Y) + \lambda\rho(Z)$ for $\lambda \in [0,1]$.

v) Positive homogeneity: $\rho(cY) = c\rho(Y)$ for $c \geq 0$.

vi) Subadditivity: $\rho(Y+Z) \leq \rho(Y) + \rho(Z)$.

A convex risk measure, the focus of Chapter 2, is a risk measure fulfilling ii), iii) and iv), see Föllmer and Schied (2011); Frittelli and Rosazza Gianin (2002) and references therein. The subclass of convex risk measures, called *coherent* risk measures are risk measures that additionally satisfy v) and therefore vi) (Artzner et al., 1999). It is known (Kusuoka, 2001), that spectral risk measures are law-invariant coherent risk measures, thus ES is coherent and convex, while the VaR violates subadditivity and hence convexity.

**Shortfall risk measures**

Shortfall risk measures, associated with utility-type arguments, are defined through $\rho(Y) = \inf\{y \in \mathbb{R} \,|\, E(\ell(Y-y)) \leq y_0\}$, where $\ell$ is a non-decreasing, non-constant and convex loss function and $y_0$ a point in the interior of the range of $\ell$ (Föllmer and Schied, 2002). Examples of shortfall risk measures include entropic risk measures (Gerber, 1974), and the class of generalised quantiles called expectiles (Newey and Powell, 1987; Bellini et al., 2014).

## 1.3 Structure of the thesis

This thesis is based on three (working) papers: Chapter 2 includes Pesenti et al. (2016), Chapter 3 is based on Pesenti et al. (2018*b*) and Chapter 4 is a working paper termed *Cascade sensitivity measures* (Pesenti et al., 2018*a*). All effort has been made to homogenise the notation, however, due to the different approaches of

studying sensitivity analyses, different levels of mathematical notation may be required, rendering it necessary to refine specific notations.

### 1.3.1 Chapter 2

This chapter addresses the interplay between properties of a risk measure to rank different risk profiles and its statistical robustness. The fundamental purpose of a risk measure, particularly in financial risk management applications, is to distinguish consistently between different levels of risk severities. From a practical perspective, a risk measure should be robust, that is, insensitive to small perturbations in input assumptions. It is known in the literature (Cont et al., 2010; Krätschmer et al., 2014), that strong assumptions on the risk measure's ability to distinguish between risks may lead to a lack of robustness. This trade-off between robustness and consistent risk ranking is addressed by specifying the regions in the space of distribution functions, where law-invariant convex risk measures are indeed robust.

We consider the case where a risk measure is evaluated on the output of an aggregation function. Extending the definition of robustness to this setting, we find that law-invariant convex risk measures are robust for any aggregation function that satisfies a linear growth condition in the tail, provided that the set of possible marginals is uniformly integrable. Thus, we show that all law-invariant convex risk measures possess the aggregation-robustness property introduced by Embrechts et al. (2015) and further studied by Krätschmer et al. (2017). This is in contrast to the widely-used, non-convex, risk measure VaR, whose robustness in a risk aggregation context requires restricting the possible dependence structures of the input vectors.

### 1.3.2 Chapter 3

The focus of this chapter is on associating a critical increase in the risk measure applied to the model output, to specific input factors. We propose a global and model-independent framework, termed *reverse sensitivity testing*, comprising three steps: (a) an output stress is specified corresponding to an increase in the risk measure(s); (b) a (stressed) probability measure is derived, minimising the Kullback-Leibler divergence with respect to the baseline probability, under constraints generated by the output stress; (c) changes in the distributions of input factors are evaluated. We argue that a substantial change in the distribution of an input factor corresponds to high sensitivity to that input and introduce a novel sensitivity measure to formalise this insight.

The proposed framework differs from *reverse stress testing*, which is widely used in practical applications, in that it provides the whole distribution of the output and of the input factors under the stressed model, allowing for an in-dept inspection and analysis of the change from the baseline to the stressed model.

Implementation of the reverse sensitivity testing framework in a Monte Carlo setting can be performed on a single set of input / output scenarios, simulated under the baseline model. Thus the approach circumvents the need for additional computationally expensive evaluations of the aggregation function. The proposed approach is illustrated through numerical examples with a simple insurance portfolio and a model of a London Insurance Market portfolio used in industry.

### 1.3.3 Chapter 4

Sensitivity measures are the tools used in factor prioritisation, to quantify the extent to which the distribution of a model output is affected by small changes (stresses) in an individual random input factor. It is known that sensitivity measures defined as

partial derivatives do not fully account for interactions among (or dependence between) input factors (Borgonovo and Plischke, 2016). However, for input factors that are dependent, a stress on one input should also precipitate stresses in other input factors. We introduce a novel sensitivity measure termed *cascade sensitivity*, which captures the direct impact of the stressed input factor on the output, as well as indirect effects via other risk factors that are dependent on the one being stressed. In this way, the dependence between inputs is explicitly taken into account.

Representations of the cascade sensitivity measure, which can be calculated from a single Monte Carlo sample, are provided for two types of stress: (a) a perturbation of the distribution of an input factor, such that the stressed input follows a mixture distribution, and (b) an additive random shock applied to the input factor. These representations do not require simulations under different model specifications or the explicit study of the properties of the model's aggregation function, making the proposed method attractive for practical applications, as we illustrate through numerical examples.

## 1.4 Direction for future research

To increase the impact and applicability of the contribution of this thesis an **R** package incorporating the reverse sensitivity testing framework in Chapter 3 and the cascade sensitivity measure of Chapter 4 is in preparation. The **R** package will facilitate the practicability and usage of our proposed sensitivity methodologies in academia and practice.

Further to the **R** package, the introduced sensitivity concepts open multiple doors for future research projects. Specifically, the generalisation of the reverse sensitivity testing framework to different risk measure constraints, such as to the class of distortion risk measures or to the ES only, as well as utilising other divergences than the

Kullback-Leibler divergence. Furthermore, the reverse sensitivity testing might provide a mathematical foundation to establish a *reverse stress testing* methodology. Stress testing (and reverse stress testing) is intensely used by financial institutions for risk assessment, and also prescribed by regulation Swiss Solvency Test (2006); BCBS (2013); EIOPA (2009). However, a rigorous mathematical theory for reverse stress testing is still lacking and the subject of current research (Glasserman et al., 2015; Ludkovski and Risk, 2017; McNeil and Smith, 2012).

The proposed cascade sensitivity measure, whose characteristic is to explicitly take into account the dependence between input factors, has a natural application in *systemic risk*. Systemic risk is concerned with contagion effects that spread through the financial market and distress the whole financial system. *Systemic risk measures* have been introduced in the literature, see Feinstein et al. (2017) for an overview, however, a sensitivity measure that distinguishes between the direct effect of an input factor and the indirect effects due to dependencies on other input factors, might provide novel insight.

# Chapter 2

# Robustness Regions for Measures of Risk Aggregation

This chapter is based on the publication Pesenti et al. (2016) and has been presented at the XVIII workshop on *Quantitative Finance* (Milan, 2017), at the conference on *Model Uncertainty and Robust Finance* (Milan, 2016) and at the *3rd European Actuarial Journal Conference* (Lyon, 2016).

## 2.1 Introduction

Since the wide-spread adoption of VaR frameworks in the 1990s, risk measures have constituted an integral part of financial risk management. The use of risk measures is prescribed by banking (BCBS, 2012, 2013) and insurance regulation (EIOPA, 2009) for calculating the capital requirements of portfolios of future losses. Furthermore, the use of risk measures, evaluated using internally developed statistical models, is increasingly embedded in the operations of insurance companies (SCOR, 2008; Sandström, 2016).

As a consequence, the discussion of desirable properties of risk measures has been the focus of much academic and industry debate. A first set of considerations relates to risk measures' ability to reflect diversification appropriately, by the properties of

subadditivity (Artzner et al., 1999) and convexity (Föllmer and Schied, 2002; Frittelli and Rosazza Gianin, 2002), and to order risk consistently (Bäuerle and Müller, 2006; Denuit et al., 2006). These issues are interrelated: law-invariant convex risk measures, introduced by Föllmer and Schied (2002); Frittelli and Rosazza Gianin (2002) and subsuming coherent risk measures (Artzner et al., 1999), rank risks in a way that preserves first-order and second-order stochastic dominance (Bäuerle and Müller, 2006). The risk measure ES is the convex risk measure used most widely in the practice of risk management.

A second set of considerations acknowledges that risk measures need to be estimated from historical and / or simulated data and thus require reliable estimators. A fundamental concept is the question of robustness, that is, whether risk measure estimates remain relatively insensitive to small perturbations in the underlying distribution from which data are generated (Hampel, 1971; Huber and Ronchetti, 2009). A growing academic literature is concerned with robustness in the context of risk measurement (Cont et al., 2010; Kiesel et al., 2016; Krätschmer et al., 2014; Bellini et al., 2014; Embrechts et al., 2015; Krätschmer et al., 2017). A key finding of this literature is that robustness is to an extent contradictory to a consistent ordering of risks. In particular, there does not exist a law-invariant convex risk measure that is robust (following the definition of Hampel (1971); Huber and Ronchetti (2009)) on the whole space of integrable random variables. This fact has been used as an argument against the use of convex risk measures such as ES and in favour of the non-convex risk measure VaR (Cont et al., 2010). Such arguments have coloured much of the policy discussion surrounding the relative merits of ES and VaR for use in capital regulation (BCBS, 2012, 2013; IAIS, 2014).

One way to address the apparent conflict between consistency of risk ranking and robustness, is to consider alternative, less restrictive, definitions of robustness (Kiesel

et al., 2016; Krätschmer et al., 2014). Another approach also taken in Krätschmer et al. (2017), which we follow in this chapter, is to relax the requirement that risk measures be robust on the whole space of integrable random variables, given that "... this case is not generally interesting in econometric or financial applications since requiring robustness against all perturbations of the model is quite restrictive..." (Cont et al., 2010). This approach suggests an analysis of regions on which risk measures are robust. Consequently, since in different applications different regions of distributions may form plausible input spaces, selection of a risk measure for a particular application should reflect the extent to which the risk measure is robust on the region of interest.

In this chapter, we study robustness regions for convex risk measures and show that they are characterised by the property of uniform integrability – through examples we demonstrate that this is not an excessively strong requirement on the input space. Furthermore, we consider the realistic case where risk measures are evaluated on (possibly non-linear) functions of random vectors of risk factors, such that the input space consists of multivariate distributions (SCOR, 2008; Tsanakas and Millossovich, 2016). This case, typical in the risk modelling performed by insurance companies, is generally not considered in the literature on robustness, with the exception of Embrechts et al. (2015); Krätschmer et al. (2017) who focus on fixed marginals. However, robustness as defined in Hampel (1971); Huber and Ronchetti (2009), that is, insensitivity to small deviations from the underlying distribution, includes both perturbation in the marginals and the dependence structure of the random vector of input risk factors. Allowing for uncertainty in the marginal distributions, we show that weak restrictions on the marginals (uniform integrability) and the aggregation function (linear growth in the tail) ensure robustness of convex risk measures. Consequently, we argue that in applications where risk aggregation takes place and uncertainty around

the dependence structure is high, convex risk measures such as ES have attractive robustness properties, compared to, say, VaR.

In Section 2.2 notation and mathematical preliminaries are stated. In Section 2.3, the robustness of convex risk measures is studied. First, in Section 2.3.1, robustness is formally defined and its relationship to continuity of risk measures (Hampel's theorem) is presented. A key result (that also follows from Krätschmer et al. (2017)) is then shown: convex risk measures are robust on uniformly integrable sets. Subsequently, in Section 2.3.2, examples of such uniformly integrable sets are given. Uniform integrability is a constraint on the tail behaviour of a set of distributions. Thus convex risk measures are robust on sets including parametric families with bounded second moment; sets of random variables that are less volatile (in convex order) than those in a given uniformly integrable set. Section 2.3.3 presents examples of sets on which convex risk measures are not robust and Section 2.3.4 points at possible extensions to risk measures defined on the set of random variables with finite $p$-th moment.

In Section 2.4, robustness is studied in the context of risk aggregation, where a risk measure is applied on real-valued aggregation function of a random vector of risk factors; we call the composition of the risk measure with the aggregation function an *aggregation measure*. In Section 2.4.1, robustness of aggregation measures is defined with respect to distributions of random vectors. A direct multivariate extension of Hampel's theorem is given, associating robustness with continuity of the aggregation measure. Consequently, if the risk measure is convex and the aggregation function continuous, the aggregation measure is robust as long as the aggregate risk position belongs to a uniformly integrable set. In Section 2.4.2 we show that for robustness of aggregation measures it is sufficient that the marginals of the vector of risk factors belong to uniformly integrable sets and that the aggregation function possesses a linear growth condition in the tail. Significantly, no constraints on the dependence structure

of risk factors is placed. This includes, as a special case, aggregation via the ordinary sum and thus generalises the results on *aggregation robustness* in Embrechts et al. (2015) to the class of law-invariant convex risk measures and the results in Krätschmer et al. (2017) to uncertainty in the marginal distributions. In Section 2.4.3 it is shown that robustness is also satisfied for aggregation via compound distributions, a typical setting in actuarial science, as long as the frequency and severity distributions are dominated (in first-order stochastic dominance) by integrable random variables.

Finally, in Section 2.5, a comparison with the robustness regions of the (non-convex) VaR measure is made. VaR is robust as long as the distribution function is strictly increasing. We argue that in applications, this can be a stronger requirement than the uniform integrability that is required when convex risk measures are used. Non-linear aggregation functions, such as the ones arising in the context of reinsurance, can lead to constant parts of the aggregate distribution function and thus to non-robustness. Furthermore, it is known from the literature on dependence uncertainty that dependence structures can be designed such that the distribution of the sum is not strictly monotonic in the tail, when the marginal distributions satisfy particular ('mixability') conditions (Embrechts, Puccetti and Rüschendorf, 2013; Wang et al., 2013; Bernard et al., 2014; Wang and Wang, 2011, 2016). Thus, robustness of VaR requires restrictions both in the aggregation function and the dependence structure. In applications such as the internal capital modelling performed by insures, we believe that such constraints are unrealistic, compared to those applying to convex risk measures. Thus our findings indicates that in applications where non-linear aggregations and high dependence uncertainty are present, convex risk measures such as ES, may be preferable to VaR.

## 2.2 Preliminaries

Throughout this chapter, we equip the measurable space $(\Omega, \mathcal{A})$ with a probability measure $P$, such that $(\Omega, \mathcal{A}, P)$ is an atomless probability space. We denote the space of real-valued random variables by $\mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{A}, P)$, the subspace of integrable random variables by $\mathcal{L}^1 = \{X \in \mathcal{L}^0 \,|\, \|X\|_1 = E(|X|) < +\infty\}$ and the subset of (essentially) bounded random variables by $\mathcal{L}^\infty$. For $\mathcal{X} \subset \mathcal{L}^0$ we define the corresponding set of distribution functions by $\mathfrak{D}(\mathcal{X}) = \{P \circ X^{-1} \,|\, X \in \mathcal{X}\}$. We denote by $F_X(\cdot) = P(X \leq \cdot)$ the distribution function of $X$ and write $X \sim F_X$, so that $\mathfrak{D}(\mathcal{X}) = \{F_X \,|X \in \mathcal{X}\}$. Note that we identify distribution functions on $\mathbb{R}$ with the corresponding probabilities on the Borel $\sigma$-field $\mathcal{B}(\mathbb{R})$. We write $\mathfrak{M} = \mathfrak{D}(\mathcal{L}^0)$ for the set of all distribution functions on $\mathbb{R}$, and $\mathfrak{M}^1 = \{F \in \mathfrak{M} \,|\, \int_{\mathbb{R}} |x| \mathrm{d}F(x) < +\infty\} = \mathfrak{D}(\mathcal{L}^1)$.

On the space $\mathfrak{M}$ we consider the Prokhorov distance defined for $F, G \in \mathfrak{M}$ through

$$d_P(F, G) = \inf\{\varepsilon > 0 \,|F(B) \leq G(B^\varepsilon) + \varepsilon, \text{ for all Borel sets } B \text{ on } \mathbb{R}\},$$

where $B^\varepsilon = \{x \in \mathbb{R} \,|\, \inf_{y \in B} |x - y| \leq \varepsilon\}$.

The following definition is of central importance: A set of distribution functions $\mathfrak{U} \subset \mathfrak{M}^1$ is *uniformly integrable* if

$$\lim_{K \to +\infty} \sup_{F \in \mathfrak{U}} \int_{|x| > K} |x| \mathrm{d}F(x) = 0.$$

We say a set of random variables $\mathcal{U} \subset \mathcal{L}^1$ is uniformly integrable if $\mathfrak{D}(\mathcal{U})$ is uniformly integrable, equivalently

$$\lim_{K \to +\infty} \sup_{X \in \mathcal{U}} E\Big(|X| \mathbb{1}_{\{|X| > K\}}\Big) = 0.$$

Uniform integrability of a set posits that the contribution of the distributions' far tails can be uniformly controlled across the elements of the set. Thus, it is a stronger condition than requiring that all elements of a set are integrable.

Recall that a risk measure $\rho(\cdot)\colon \mathcal{L}^1 \to \mathbb{R}$ is called convex if it fulfils the properties ii), iii) and iv) in Section 1.2.1. Note, that in this chapter, we consider risk measures that map into the real lines, thus only take finite values. An example of a convex risk measure that is finite on $\mathcal{L}^1$ is the ES, which belongs to the class of spectral risk measures. Spectral risk measures are generally not finite on $\mathcal{L}^1$, however, finiteness is guaranteed if the weight function $\gamma$ is constant on $(1-\varepsilon, 1]$ for $\varepsilon > 0$, as is the case for ES, corresponding to $\gamma(u) = \frac{1}{1-\alpha}\mathbb{1}_{\{u>\alpha\}}$.

A law-invariant risk measure (property i) ) induces a functional on the corresponding set of distribution functions, $\rho[\cdot]\colon \mathfrak{M}^1 \to \mathbb{R}$, through $\rho[F_X] = \rho(X)$ for $F_X \in \mathfrak{M}^1$. For instance, we write $E(X) = E[F_X]$. (Throughout we denote law invariant functionals using round brackets $(\cdot)$ when the argument is a random variable, and square brackets $[\cdot]$ when the argument is a distribution.) The property of law invariance is standard in risk management applications, requiring that risk assessments only depend on the distribution of random losses. Therefore all risk measures in this chapter are tacitly assumed to be law-invariant without this being explicitly stated in the sequel.

We say a risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is continuous on $\mathcal{X} \subset \mathcal{L}^1$ with respect to the Prokhorov distance if the restriction of the induced functional $\rho[\cdot]$ on $\mathfrak{D}(\mathcal{X})$ is continuous with respect to $d_P$. That is, for all $F_0 \in \mathfrak{D}(\mathcal{X})$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all $F \in \mathfrak{D}(\mathcal{X})$ we have $d_P(F_0, F) < \delta$ implies $|\rho[F_0] - \rho[F]| < \varepsilon$.

**Remark 2.2.1.** A substantial part of the early literature considers risk measures, axiomatically introduced in Artzner et al. (1997, 1999), defined on $\mathcal{L}^\infty$; however, insurance and financial portfolios are primarily exposed to unbounded risks. Therefore we choose $\mathcal{L}^1$ as our model space. In fact, the natural model space for law-invariant

convex risk measures is $\mathcal{L}^1$, since outside this space the risk measure can only take value $+\infty$ (Shapiro et al., 2009; Filipović and Svindland, 2012). Selected literature on risk measures defined on a broader space than $\mathcal{L}^\infty$ are Delbaen (2002) for general probability spaces, Shapiro et al. (2009); Kaina and Rüschendorf (2009) on sets of random variables with finite $p$-th moment, Haezendonck and Goovaerts (1982); Cheridito and Li (2009); Krätschmer et al. (2014) on Orlicz spaces and Filipović and Svindland (2012) for extensions of risk measures from $\mathcal{L}^\infty$ to $\mathcal{L}^1$.

## 2.3 Robustness

### 2.3.1 Robustness of convex risk measures

The classical definition of statistical robustness (Hampel, 1971), considers estimators as functionals of empirical distribution functions. For a distribution function $F \in \mathfrak{M}^1$ and sample size $k \geq 1$ the empirical distribution function is defined by the random measure

$$\hat{F}_k(t,\omega) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{\{X_i(\omega)\leq t\}},\ (t,\omega) \in \mathbb{R}\times\Omega,$$

where $X_1,\dots,X_k \in \mathcal{L}^1$ are independent with common distribution function $F$. In the sequel we consider the sequence of estimators $\{\hat{\rho}_k\}_k$ of a risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ by evaluating the risk measure on the empirical distribution functions. That is, for $F \in \mathfrak{M}^1$ and $k \geq 1$, we define

$$\hat{\rho}_k[F](\omega) = \rho[\hat{F}_k(\cdot,\omega)],\ \omega \in \Omega. \tag{2.1}$$

Note that the estimator $\hat{\rho}_k[F]$ is a random variable. Ideally, the estimator $\{\hat{\rho}_k\}_k$ should be consistent and robust. The sequence of estimators is consistent if it converges to the true value, $\hat{\rho}_k[F] \to \rho[F]$ in probability. Robustness, according to Hampel (Hampel,

1971; Huber and Ronchetti, 2009), is understood as insensitivity of estimators to small perturbations in the distribution $F$.

**Definition 2.3.1.** *((Hampel, 1971))*
A risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is *robust* on $\mathcal{X} \subset \mathcal{L}^1$ (equivalently $\rho[\cdot]$ is robust on $\mathfrak{D}(\mathcal{X})$) if for any $F_0 \in \mathfrak{D}(\mathcal{X})$ the sequence of estimators $\{\hat{\rho}_k[F_0]\}_k$, as defined in (2.1), fulfils that for all $\varepsilon > 0$ there exists $\delta > 0$ and $k_0 \in \mathbb{N}$ such that, for all $F \in \mathfrak{D}(\mathcal{X})$ and $k \geq k_0$, we have

$$d_P(F_0, F) < \delta \quad \Rightarrow \quad d_P\Big(\mathfrak{D}\big(\hat{\rho}_k[F_0]\big), \mathfrak{D}\big(\hat{\rho}_k[F]\big)\Big) < \varepsilon.$$

By the celebrated theorem of Hampel (Hampel, 1971), given consistency, robustness of a risk measure is equivalent to continuity with respect to the Prokhorov distance.

**Theorem 2.3.2.** *((Hampel, 1971), Theorem 2.21 in (Huber and Ronchetti, 2009))*
*Let $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ be a risk measure and $\mathcal{X} \subset \mathcal{L}^1$. Assume that the sequence $\{\hat{\rho}_k\}_k$, as defined in (2.1), is consistent in a neighbourhood of $F_0$ for all $F_0 \in \mathfrak{D}(\mathcal{X})$. Then $\rho$ is continuous on $\mathfrak{D}(\mathcal{X})$ with respect to the Prokhorov distance if and only if the risk measure is robust on $\mathfrak{D}(\mathcal{X})$.*

For convex risk measures we obtain a one-to-one correspondence between robustness and continuity, since they are consistent on $\mathfrak{M}^1$.

**Proposition 2.3.3.** Let $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ be a convex risk measure and $\mathcal{X} \subset \mathcal{L}^1$. Then, $\rho$ is continuous with respect to the Prokhorov distance on $\mathfrak{D}(\mathcal{X})$ if and only if it is robust on $\mathfrak{D}(\mathcal{X})$.

*Proof.* We show strong consistency of convex risk measures, that is for $F_0 \in \mathfrak{M}^1$ we have $\hat{\rho}_k[F_0](\omega) \to \rho[F_0]$ for almost every $\omega \in \Omega$. Let $\{\hat{F}_{0k}(\cdot, \omega)\}_k$, $\omega \in \Omega$, be the corresponding sequence of empirical distribution functions. By Glivenko-Cantelli

$\{\hat{F}_{0k}(\cdot,\omega)\}_k$ converges to $F_0(\cdot)$ for almost every $\omega \in \Omega$ in the Prokhorov distance. The strong law of large numbers implies that for $X_{0,i} \sim F_0$, $i = 1, \ldots, k$ and almost every $\omega \in \Omega$

$$\int_{\mathbb{R}} |x| \mathrm{d}\hat{F}_{0k}(x,\omega) = \frac{1}{k}\sum_{i=1}^{k} |X_{0,i}(\omega)| \quad \longrightarrow \quad E(|X_0|) = \int_{\mathbb{R}} |x| \mathrm{d}F_0(x), \text{ as } k \to +\infty.$$

Applying Lemma 2.A.1 $\{\hat{F}_{0k}(\cdot,\omega)\}_k$ converges to $F_0(\cdot)$ in the Wasserstein distance (see Appendix for the definition and properties of such distance) for almost every $\omega \in \Omega$. Since convex risk measures are continuous with respect to the Wasserstein distance, Theorem 2.8 in Krätschmer et al. (2014), $\hat{\rho}_k[F_0](\omega) = \rho[\hat{F}_{0k}(\cdot,\omega)] \to \rho[F_0]$, as $k \to +\infty$, for almost every $\omega \in \Omega$. □

No convex risk measure is robust on the whole of $\mathcal{L}^1$, as shown in Lemma 2.3.4 below.

**Lemma 2.3.4.** There does not exist a convex risk measure that is robust on $\mathcal{L}^1$.

*Proof.* Bäuerle and Müller (2006); Cont et al. (2010); Krätschmer et al. (2014) show that there does not exist a convex risk measure that is continuous with respect to the Prokhorov distance on the whole space of integrable random variables. Applying Proposition 2.3.3 gives the claim. □

Given the importance of both convexity and robustness for risk management, the need emerges to study subsets of $\mathcal{L}^1$ on which convex risk measures become robust. Uniformly integrable sets are at the core of characterising robustness regions for convex risk measures.

**Theorem 2.3.5.** *A convex risk measure is robust on $\mathcal{X} \subset \mathcal{L}^1$ if the set $\mathcal{X}$ is uniformly integrable.*

*Proof.* Convex risk measures are continuous on $\mathfrak{M}^1$ with respect to the Wasserstein distance, Theorem 2.8 in Krätschmer et al. (2014). On a uniformly integrable set the topology induced by the Wasserstein distance is equivalent to the topology induced by the Prokhorov distance, see Lemma 2.A.1 or Theorem 2 in Dobrushin (1970). Hence, on $\mathcal{X}$ the risk measure is continuous with respect to the Prokhorov distance and we can apply Proposition 2.3.3. □

Alternatively, the proof of Theorem 2.3.5 follows from Theorem 2.6 in Krätschmer et al. (2017).

**Remark 2.3.6.** The general concept of robustness is based on continuity with respect to the weak topology on $\mathfrak{M}$ (Huber and Ronchetti, 2009). Due to its tractability, the Lévy distance is frequently used for defining robustness (Cont et al., 2010). Since both the Prokhorov and the Lévy distance generate the weak topology on $\mathfrak{M}$, they give rise to the same notion of robustness (Huber and Ronchetti, 2009). We adopt the Prokhorov distance since it allows for a natural extension to multivariate distribution functions, see Section 2.4.

### 2.3.2 Robustness regions of convex risk measures

In this section, we provide some examples of classes of sets that are uniformly integrable and on which, by Theorem 2.3.5, convex risk measures are robust. It is seen throughout that uniform integrability puts a constraint on the tail behaviour of the risks considered.

First, we note that a convex risk measure is robust when evaluated on a set of empirical distribution functions.

**Lemma 2.3.7.** Let $F \in \mathfrak{M}^1$. A convex risk measure is robust on the sequence of empirical distribution functions $\left\{\hat{F}_k(\cdot,\omega) \,|\, k \geq 1\right\} \subset \mathfrak{M}^1$ for almost every $\omega \in \Omega$.

*Proof.* In the proof of Proposition 2.3.3 it was shown that the sequence $\hat{F}_k(\cdot, \omega)$ converges in the (Prokhorov and) Wasserstein distance to $F$ for almost every $\omega \in \Omega$. By Lemma 2.A.1 this implies that the sequence is, for almost every $\omega$, uniformly integrable and we can apply Theorem 2.3.5. □

More generally, a convex risk measure is robust on sets of uniformly bounded random variables, that is $\{X \in \mathcal{L}^1 \,|\, |X| \leq M,\ P\text{-a.s.}\}$ for $M > 0$, see Durrett (2013, p. 220). Instead of restricting the support of the random variables we could restrict their moments. A convex risk measure is robust on the set of distribution functions $\mathfrak{U} \subset \mathfrak{M}^1$ having uniformly bounded second moments or, more generally, satisfying Billingsley (2008, p. 218)

$$\sup_{F \in \mathfrak{U}} \int_{\mathbb{R}} |x|^{1+\varepsilon} \mathrm{d}F(x) < +\infty, \text{ for some } \varepsilon > 0.$$

Subsequently, a convex risk measure is robust on a family of parametric models, $\{F_\theta \,|\, \theta \in \Theta\}$, if the family fulfils $\int_{\mathbb{R}} |x|^2 \mathrm{d}F_\theta(x) < M$, for all $\theta \in \Theta$. For example, consider the exponential dispersion family, a parametric family of distribution functions with density

$$f(x; \theta, \phi) = \exp\left\{\frac{x\theta - b(\theta)}{\phi/w} + c(x, \phi, w)\right\},\ x \in \mathbb{R}$$

with weight $w > 0$, dispersion parameter $\phi > 0$ and normalising function $c(\cdot, \cdot, \cdot)$. The canonical parameter of the exponential dispersion family is $\theta \in \Theta$, where $\Theta \subset \mathbb{R}$ and $b\colon \Theta \to \mathbb{R}$ is the cumulant function such that the density is well-defined and has identical support for all $\theta \in \Theta$ (Nelder and Wedderburn, 1973). The exponential dispersion family includes the Poisson, Negative-Binomial, Gamma, Gaussian and Inverse Gaussian.

**Lemma 2.3.8.** A convex risk measure is robust on the exponential dispersion family if the parameter space $\Theta$ is compact and the function $b$ twice continuously differentiable on $\Theta$.

*Proof.* Let $X$ follow a distribution that belongs to the exponential dispersion family. Then $E(X) = b'(\theta)$ and $\mathrm{Var}(X) = \frac{\phi}{w} b''(\theta)$ (Wüthrich, 2016). Both the first and second derivative $b', b''$ are continuous and hence bounded on the compact set $\Theta$. □

We refer to Krätschmer et al. (2017) for a broader discussion and examples involving parametric models such as the Normal, Pareto, Gamma and Gumbel distributions.

Now we consider the relationship between uniform integrability and stochastic orderings. A convex risk measure is robust on a set of non-negative random variables that are smaller (in first-order stochastic dominance) than those in a given uniformly integrable set.

**Lemma 2.3.9.** Let $\mathcal{U}$ be a uniformly integrable set of non-negative random variables. A convex risk measure is robust on the set

$$\begin{aligned}\mathcal{N} = \{Y \in \mathcal{L}^1 \mid Y \geq 0 \text{ and there exists } X \in \mathcal{U} \text{ such that} \\ E(f(Y)) \leq E(f(X)) \text{ for all increasing } f\}.\end{aligned}$$

*Proof.* For $K > 0$, the function $f(x) = x\mathbb{1}_{\{x>K\}}$ is increasing. Hence we have, by uniform integrability of $\mathcal{U}$,

$$\lim_{K\to+\infty} \sup_{Y\in\mathcal{N}} E\Big(Y\mathbb{1}_{\{Y>K\}}\Big) \leq \lim_{K\to+\infty} \sup_{X\in\mathcal{U}} E\Big(X\mathbb{1}_{\{X>K\}}\Big) = 0.$$

The conclusion follows by Theorem 2.3.5. □

An example of the application of Lemma 2.3.9 is the *Generalised Pareto Distribution* (GPD) denoted by $G_{\xi;\sigma}$, with shape and scale parameters, $\xi \in \mathbb{R}$ and $\sigma > 0$ respectively,

defined through

$$G_{\xi;\sigma}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left\{ - \frac{x}{\sigma} \right\} & \xi = 0, \end{cases}$$

where $x \geq 0$, if $\xi \geq 0$, and $0 \leq x \leq -\sigma/\xi$, if $\xi < 0$. The GPD is often used in insurance and operational risk management to model portfolios that can produce very large claims, since it is the limit distribution of conditional excesses over high thresholds (Embrechts, Klüppelberg and Mikosch, 2013). The expectation of a GPD is finite if the shape parameter satisfies $\xi < 1$. For a set of GPDs to be uniformly integrable it is necessary that their shape parameters be bounded away from 1; see Proposition 2.3.14 for the necessity of this condition in the more general case of regularly varying distributions. A convex risk measure is robust on the set of distributions $\{G_{\xi;\sigma} \,|\, \sigma \leq \overline{\sigma},\, \xi \leq \overline{\xi}\}$, where $\overline{\xi} < 1$. This follows from Lemma 2.3.9 and the observation that, for fixed $\sigma$ and $0 < \xi < 1$ the family $G_{\xi;\sigma}$ is first-order stochastically ordered in $\xi$ (for fixed $\sigma$) and in $\sigma$ (for fixed $\xi$).

Similarly, a convex risk measure is robust on a set of random variables that are less volatile (in convex order) than those in a given uniformly integrable set. An example is the set of conditional expectations $\{\mathbb{E}[X|\mathcal{G}] \,|\, \mathcal{G}$ sub-$\sigma$-algebra of $\mathcal{A}\}$ for $X \in \mathcal{L}^1$, see Billingsley (2008, p. 469).

**Lemma 2.3.10.** Let $\mathcal{U}$ be a uniformly integrable set. A convex risk measure is robust on the set

$$\mathcal{N} = \{Y \in \mathcal{L}^1 \,|\, \text{there exists } X \in \mathcal{U} \text{ such that } E(f(Y)) \leq E(f(X)) \text{ for all convex } f\}.$$

*Proof.* For $K > 0$, the function $f(x) = (|x| - K)\mathbb{1}_{\{|x|>K\}}$ is convex. Hence we have, for $Y \in \mathcal{N}$ and $X \in \mathcal{U}$ dominating $Y$ in convex order,

$$\begin{aligned} E\Big(|Y|\mathbb{1}_{\{|Y|>K\}}\Big) &= E\Big(\big(|Y| - K\big)\mathbb{1}_{\{|Y|>K\}}\Big) + KP(|Y| > K) \\ &\leq E\Big(\big(|X| - K\big)\mathbb{1}_{\{|X|>K\}}\Big) + KP(|Y| > K) \\ &\leq E\Big(|X|\mathbb{1}_{\{|X|>K\}}\Big) + KP(|Y| > K). \end{aligned}$$

By De la Vallée Poussin's Theorem (Diestel, 1991), there exist a non-decreasing convex function $\psi\colon [0, +\infty) \to [0, +\infty)$ with $\psi(0) = 0$, such that $\lim_{x\to+\infty} \frac{\psi(x)}{x} = +\infty$ and $\sup_{X\in\mathcal{U}} E\Big(\psi(|X|)\Big) < +\infty$. Applying Markov's inequality, we have

$$KP(|Y| > K) \leq \frac{K}{\psi(K)} E\Big(\psi(|Y|)\Big) \leq \frac{K}{\psi(K)} E\Big(\psi(|X|)\Big).$$

By uniform integrability of $\mathcal{U}$,

$$\lim_{K\to+\infty} \sup_{Y\in\mathcal{N}} E\Big(|Y|\mathbb{1}_{\{|Y|>K\}}\Big) \leq \lim_{K\to+\infty} \sup_{X\in\mathcal{U}} \left(E\Big(|X|\mathbb{1}_{\{|X|>K\}}\Big) + \frac{K}{\psi(K)} E\Big(\psi(|X|)\Big)\right) = 0.$$

The conclusion follows by Theorem 2.3.5. □

Note that Lemma 2.3.10, in the special case when $\mathcal{U}$ is a singleton, follows from Proposition 3.3 in Mao and Wang (2015).

We now consider how larger uniformly integrable sets are constructed from other uniformly integrable sets. Finite unions of uniformly integrable sets are uniformly integrable, so that a convex risk measure that is robust on finitely many uniformly integrable sets is also robust on their union. Moreover, to any uniformly integrable set on which a convex risk measure is robust we can add finitely many distribution functions without losing robustness. The next proposition shows that a convex risk measure that is robust on a uniformly integrable set $\mathfrak{U} \subset \mathfrak{M}^1$ is also robust on the larger

set of all possible mixtures of elements of $\mathfrak{U}$. Mixtures are used to model experimental error or contaminations, by assuming that the underlying distribution function $F$ is contaminated with an error, with distribution $G$, that occurs with (small) probability $\lambda \in (0,1)$, so that the contaminated distribution is $(1-\lambda)F + \lambda G$.

**Proposition 2.3.11.** For a uniformly integrable set $\mathfrak{U} \subset \mathfrak{M}^1$, a convex risk measure is robust on the set of mixtures $\Big\{(1-\lambda)F + \lambda G \,|\, F, G \in \mathfrak{U},\ \lambda \in [0,1]\Big\}$.

*Proof.* By Theorem 2.3.5 it is enough to show that $\Big\{(1-\lambda)F+\lambda G \,|\, F, G \in \mathfrak{U},\ \lambda \in [0,1]\Big\}$ is uniformly integrable. For $\lambda \in [0,1]$ and $F, G \in \mathfrak{U}$ we calculate

$$\begin{aligned}
&\sup_{F,G\in\mathfrak{U},\lambda\in[0,1]} \int_{|x|>K} |x| \mathrm{d}[(1-\lambda)F(x) + \lambda G(x)] \\
&\qquad \leq \sup_{F,G\in\mathfrak{U},\lambda\in[0,1]} (1-\lambda) \int_{|x|>K} |x| \mathrm{d}F(x) + \sup_{F,G\in\mathfrak{U},\lambda\in[0,1]} \lambda \int_{|x|>K} |x| \mathrm{d}G(x) \\
&\qquad = \sup_{F\in\mathfrak{U}} \int_{|x|>K} |x| \mathrm{d}F(x),
\end{aligned}$$

which goes to zero, as $K \to +\infty$, by uniform integrability of $\mathfrak{U}$. □

Let $\{F_\theta \,|\, \theta \in \Theta\}$ describe possible model inputs and assume that the set is uniformly integrable, for example a parametric family with bounded second moment. By Theorem 2.3.5, any convex risk measure is robust on $\{F_\theta \,|\, \theta \in \Theta\}$. Assume however, that the data is contaminated, through measurement errors or the parametric family does not fit sufficiently, and the risk measure is evaluated on the mixture

$$(1-\lambda)F_\theta + \lambda G, \text{ for small } \lambda \in (0,1),\ \theta \in \Theta,\ G \in \mathfrak{N},$$

where $\mathfrak{N} \subset \mathfrak{M}^1$ denotes the collection of possible error distributions. If we have additional knowledge on the elements of $\mathfrak{N}$, such as bounded support or (uniformly) bounded mean and variance, then the convex risk measure is robust on the set of all possible mixtures, see Proposition 2.3.11.

### 2.3.3 Non-robustness of convex risk measures

In this section we present examples of sets on which convex risk measures fail to be robust. Such situations can emerge when the set is closed under mixtures and positive shifts. These conditions allow the construction of convergent sequences of distributions with divergent means. Thus situations arise where small changes in distribution can result in huge variations in the value of the risk measure.

**Proposition 2.3.12.** No spectral risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is robust on $\mathcal{X} \subset \mathcal{L}^1$, whenever $\mathfrak{D}(\mathcal{X})$ is closed under mixtures and contains a sequence of distribution functions whose means diverge to $+\infty$. Then, spectral risk measures are not robust at any distribution function $F \in \mathfrak{D}(\mathcal{X})$.

*Proof.* Let $F \in \mathfrak{D}(\mathcal{X})$ and denote by $G_k \in \mathfrak{D}(\mathcal{X})$ the sequence of distribution functions with $\lim_{k\to+\infty} E[G_k] = +\infty$. Choose $C > 0$ and define the mixture

$$F^{(k)} = (1-\lambda_k)F + \lambda_k G_k, \text{ where } \lambda_k = \min\left\{\frac{C}{|\,E[G_k]\,|}, 1\right\}.$$

Note that $\lambda_k \in [0,1]$ converges to 0, as $k \to +\infty$, hence $F^{(k)}$ converges in the Prokhorov distance to $F$. Spectral risk measures are concave with respect to mixtures (Wakker, 1994) and exceed the expectation (Denneberg, 1990), so that

$$\begin{aligned}
\liminf_{n\to+\infty} \rho[F^{(k)}] &\geq \liminf_{k\to+\infty} \Big((1-\lambda_k)\rho[F] + \lambda_k \rho[G_k]\Big) \\
&\geq \liminf_{k\to+\infty} \Big((1-\lambda_k)\rho[F] + \lambda_k E[G_k]\Big) \\
&= \lim_{k\to+\infty} \Big((1-\lambda_k)\rho[F] + \lambda_k E[G_k]\Big) \\
&= \rho[F] + C.
\end{aligned}$$

□

A similar result is now proved for general convex risk measures. For this, we need the additional assumption that the set $\mathfrak{D}(\mathcal{X})$ is closed under positive shifts, that is $F(\cdot - c) \in \mathfrak{D}(\mathcal{X})$ for all $c > 0$, and $F \in \mathfrak{D}(\mathcal{X})$. Note this is stronger than assuming the existence of a sequence of distribution functions with divergent mean. This additional assumption was not needed in the proof of Proposition 2.3.12, where instead the property of concavity with respect to mixtures of spectral risk measures (Wakker, 1994) was used.

**Proposition 2.3.13.** No convex risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is robust on $\mathcal{X} \subset \mathcal{L}^1$, whenever the set of distribution functions $\mathfrak{D}(\mathcal{X})$ is closed under mixtures and positive shifts. In this case, the risk measure is not robust at any distribution function $F \in \mathfrak{D}(\mathcal{X})$.

*Proof.* By Proposition 6.8 in Shapiro et al. (2009) the risk measure is continuous with respect to $\|\cdot\|_1$. Therefore the risk measure admits the Kusuoka representation (Shapiro et al., 2009, Thm. 6.44), that is there exists a set of probability measures $\mathfrak{P}$ on $[0, 1)$ such that the risk measure can be written as

$$\rho[G] = \sup_{\mu \in \mathfrak{P}} \left( \int_0^1 \mathrm{ES}_\alpha[G] \mathrm{d}\mu(\alpha) - \beta(\mu) \right), \text{ for } G \in \mathfrak{M}^1,$$

where $\beta(\cdot)$ is a penalty function on $\mathfrak{P}$, see Shapiro et al. (2009) for the definition. For $C > 0$, define the mixture $F^{(k)} = (1 - \lambda_k)F + \lambda_k G_k$, where $\lambda_k = \min\{C/k, 1\}$ and $G_k(\cdot) = F(\cdot - k)$, $k \geq 1$. Note that the mixture $F^{(k)}$ converges in the Prokhorov distance to $F$. Since $\mathrm{ES}_\alpha$ is concave with respect to mixtures (Wakker, 1994), we

obtain for $k \geq 1$,

$$\begin{aligned}
\liminf_{k\to+\infty} \rho[F^{(k)}] &= \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha\Big[(1-\lambda_k)F + \lambda_k G_k\Big] \mathrm{d}\mu(\alpha) - \beta(\mu) \right\} \\
&\geq \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{ \int_0^1 \Big((1-\lambda_k)\mathrm{ES}_\alpha[F] + \lambda_k \mathrm{ES}_\alpha[G_k]\Big) \mathrm{d}\mu(\alpha) - \beta(\mu) \right\} \\
&= \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha[F] \mathrm{d}\mu(\alpha) - \beta(\mu) + \lambda_k \int_0^1 \Big(\mathrm{ES}_\alpha[G_k] - \mathrm{ES}_\alpha[F]\Big) \mathrm{d}\mu(\alpha) \right\} \\
&= \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha[F] \mathrm{d}\mu(\alpha) - \beta(\mu) \right\} + C \\
&= \rho[F] + C.
\end{aligned}$$

□

In Section 2.3.2 we have seen that for robustness of convex risk measures on the space of heavy tailed distribution functions, in particular GPDs, it is necessary that the shape parameter be bounded away from 1. The following proposition considers the case of regularly varying distribution functions. A distribution function $F \in \mathfrak{M}$ on $(0, +\infty)$ is regularly varying with tail index $\alpha > 0$, if for all $t > 0$ it holds that

$$\lim_{x\to+\infty} \frac{1 - F(xt)}{1 - F(x)} = t^{-\alpha}. \tag{2.2}$$

Note that, for $\xi > 0$, the GPD $G_{\xi;\sigma}$ is regularly varying with tail index $1/\xi$. The next proposition sheds some light on the trade-off between robustness of risk measures and their sensitivity to the tail of distribution functions, see also the discussion in Krätschmer et al. (2014).

**Proposition 2.3.14.** No convex risk measure is robust on the set of regularly varying distribution functions with tail index $\alpha > 1$.

*Proof.* Let $F_{\alpha_1}, F_{\alpha_2} \in \mathfrak{M}^1$ be regularly varying with indexes $\alpha_1 > 1$, respectively $\alpha_2 > 1$. We first show that the set of regularly varying distribution functions is closed

under mixtures, that is

$$F = (1-\lambda)F_{\alpha_1} + \lambda F_{\alpha_2}$$

for $\lambda \in [0,1]$, is regularly varying. Note that $1-F = (1-\lambda)(1-F_{\alpha_1}) + \lambda(1-F_{\alpha_2})$. It is clear that both $(1-\lambda)(1-F_{\alpha_1})$ and $\lambda(1-F_{\alpha_2})$ satisfy the limit in (2.2). Proposition 1.5.7 in Bingham et al. (1989) implies then that the sum $1-F$ of these two functions satisfies again the limit in (2.2) with tail index equal to $\min\{\alpha_1, \alpha_2\}$. Hence, $F$ is a regularly varying distribution function with tail index $\min\{\alpha_1, \alpha_2\} > 1$. Clearly, any shifted regularly varying distribution function is regularly varying with the same tail index. The sequence of Pareto distributions with shape parameter $1 + \frac{1}{k}$ and scale 1, that is $F_k(x) = 1 - x^{-(1+1/k)}$, $x \geq 1$, belongs to the class of regularly varying distribution functions and its mean, $E[F_k] = \frac{1+1/k}{1+1/k-1} = k+1$, diverges to $+\infty$. Applying Proposition 2.3.13 yield the assertion. □

**Remark 2.3.15.** We consider the classical notion of robustness, defined via continuity with respect to the Prokhorov distance. A spectrum of different types of robustness, defined using alternative distances on $\mathfrak{M}$, are introduced by Krätschmer et al. (2014). If a weaker notion of robustness were defined through the Wasserstein distance, see Appendix, the constructed sequence of mixtures appearing in the proof of Proposition 2.3.13, $(1-\lambda_k)F + \lambda_k G_k$, with $F \in \mathfrak{M}^1$ and $G_k(\cdot) = F(\cdot - k)$ would not generate a discontinuity. The mixture converges in the Prokhorov distance to $F$, however, its mean diverges, hence it does not converge in the Wasserstein distance, see Lemma 2.A.1.

### 2.3.4 Generalisation to risk measures defined on $\mathcal{L}^p$

Let $p \in [1, +\infty)$ and define the space of random variables with finite $p$-th moment by $\mathcal{L}^p = \{X \in \mathcal{L}^0 \mid E(|X|^p) < +\infty\}$. Requiring a risk measure to be real-valued on the entire space of integrable random variables excludes interesting examples such as the *mean-deviation risk measures* defined by

$$\rho(X) = E(X) + cE\Big(|X - E(X)|^p\Big)^{1/p}, \quad X \in \mathcal{L}^p,\, c \geq 0.$$

Note that, for every $p \in [1, +\infty)$, the mean-deviation risk measure is convex and finite on $\mathcal{L}^p$ but not on the larger space $\mathcal{L}^r, 1 \leq r < p$ (Shapiro et al., 2009).

The Definition 2.3.1 of robustness can be generalised straightforwardly by replacing the space $\mathcal{L}^1$ with $\mathcal{L}^p$. Then, Theorem 2.3.5 generalises as follows.

**Theorem 2.3.16.** *Let $\rho\colon \mathcal{L}^p \to \mathbb{R}$ be a convex risk measure. Then $\rho$ is robust on $\mathcal{X} \subset \mathcal{L}^p$ if $\mathcal{X}$ is uniformly $p$-integrable, that is*

$$\lim_{K \to +\infty} \sup_{X \in \mathcal{X}} E\Big(|X|^p \mathbb{1}_{\{|X|^p > K\}}\Big) = 0.$$

The proof follows by reasoning similar to that in the proof of Theorem 2.3.5. Alternatively, it follows directly from Krätschmer et al. (2017). We refer to Krätschmer et al. (2017) for a thorough study of robustness of risk measures defined on Orlicz hearts.

## 2.4 Aggregation

### 2.4.1 Robustness of aggregation measures

In risk management applications, risk measures are often evaluated on the output of a complex model, which generates portfolio losses through a non-linear function of a vector of risk factors. A typical example is the aggregated loss of an insurance portfolio, represented through the insurance company's internal model. Recall, that we describe this setting through a (measurable) aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ that maps an $n$-dimensional vector into a real number. Applying the aggregation function to a random vector of input risk factors, $\boldsymbol{X} = (X_1, \ldots, X_n)$ with (multivariate) cumulative distribution function $\boldsymbol{F_X}$, we evaluate a risk measure at the (one-dimensional random) output $g(\boldsymbol{X})$. We denote the space of $n$-dimensional random vectors by $\boldsymbol{\mathcal{L}^0} = \boldsymbol{\mathcal{L}^0}(\Omega, \mathcal{A}, P)$ and the set of the corresponding (multivariate) distribution functions on $\mathbb{R}^n$ by $\boldsymbol{\mathfrak{M}} = \mathfrak{D}(\boldsymbol{\mathcal{L}^0})$. By equipping $\boldsymbol{\mathcal{L}^0}$ with the norm $\|\boldsymbol{X}\|_1 = \sum_{i=1}^n E(|X_i|)$ we write $\boldsymbol{\mathcal{L}^1} = \{\boldsymbol{X} \in \boldsymbol{\mathcal{L}^0} \mid \|\boldsymbol{X}\|_1 < +\infty\}$ and $\boldsymbol{\mathfrak{M}^1} = \mathfrak{D}(\boldsymbol{\mathcal{L}^1})$.

Throughout this section, we restrict to aggregation functions $g$ that satisfy $g(\boldsymbol{X}) \in \mathcal{L}^1$ whenever $\boldsymbol{X} \in \boldsymbol{\mathcal{L}^1}$. This is guaranteed by, for example, the linear growth condition of Definition 2.4.7; see also the discussion following Theorem 2.4.8. Weaker conditions on $g$ could be required if more restrictions were placed on $\boldsymbol{X}$, consistently with the discussion of Section 2.3.4.

**Definition 2.4.1.** For an aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ and a risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ we define the *aggregation measure* $\rho_g(\cdot)\colon \boldsymbol{\mathcal{L}^1} \to \mathbb{R}$ by $\rho_g(\boldsymbol{X}) = \rho(g(\boldsymbol{X}))$.

Thus, an aggregation measure is a functional of the input vector of risk factors. An aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ induces a functional $T_g[\cdot]\colon \boldsymbol{\mathfrak{M}} \to \mathfrak{M}$ through $T_g[\boldsymbol{F_X}] = \mathfrak{D}\Big(g(\boldsymbol{X})\Big)$. The functional $T_g$ takes the (multivariate) distribution function

$\boldsymbol{F_X} \in \mathfrak{M}$ of the vector $\boldsymbol{X}$ and returns the (univariate) distribution function $T_g[\boldsymbol{F_X}] \in \mathfrak{M}$ of $g(\boldsymbol{X})$. Since risk measures are assumed to be law-invariant, all considered aggregation measures are law-invariant and can be described by a functional on the space of distribution functions $\rho_g[\cdot]\colon \mathfrak{M}^{\mathbf{1}} \to \mathbb{R}$ through

$$\rho_g[\boldsymbol{F_X}] = \rho_g(\boldsymbol{X}) = \rho\Big[T_g[\boldsymbol{F_X}]\Big], \text{ for } \boldsymbol{F_X} \in \mathfrak{M}^{\mathbf{1}}.$$

Note that a continuous aggregation function $g$ induces, by the continuous mapping theorem, an aggregation functional $T_g\colon \mathfrak{M} \to \mathfrak{M}$ that is continuous with respect to the Prokhorov distance, $\mathfrak{M}$, $\mathfrak{M}$ both endowed with the Prokhorov distance. The Prokhorov distance on $\mathfrak{M}$ is defined for $\boldsymbol{F}, \boldsymbol{G} \in \mathfrak{M}$ through

$$d_P(\boldsymbol{F}, \boldsymbol{G}) = \inf\{\varepsilon > 0 \,|\boldsymbol{F}(\boldsymbol{B}) \leq \boldsymbol{G}(\boldsymbol{B}^{\varepsilon}) + \varepsilon, \text{ for all Borel sets } \boldsymbol{B} \text{ on } \mathbb{R}^n\},$$

where $\boldsymbol{B}^{\varepsilon} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \inf_{\boldsymbol{y}\in\boldsymbol{B}} |\boldsymbol{x} - \boldsymbol{y}| \leq \varepsilon\}$ and, for a vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we denote $|\boldsymbol{x}| = \sum_{i=1}^n |x_i|$. We say an aggregation measure $\rho_g\colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$ is continuous on $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}}^{\mathbf{1}}$ with respect to the Prokhorov distance if the restriction of the induced functional $\rho_g[\cdot]$ on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$ is continuous with respect to $d_P$. That is, for all $\boldsymbol{F_0} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ we have $d_P(\boldsymbol{F_0}, \boldsymbol{F}) < \delta$ implies $|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]| < \varepsilon$.

We extend Hampel's definition of robustness to aggregation measures, in order to reflect the sensitivity of the risk assessment to small perturbations in the distribution of the vector of risk factors. Clearly, for an aggregation measure $\rho_g\colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$ a small deviation in the $n$-dimensional input vector includes both perturbations in the marginals and the dependence structure (copula). Analogously to the one-dimensional case, we consider estimators of risk measures evaluated at the multivariate empirical distribution function. For a distribution function $\boldsymbol{F} \in \mathfrak{M}^{\mathbf{1}}$, sample size $k \geq 1$ and

independent random variables $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ with common distribution function $\boldsymbol{F}$, the multivariate empirical distribution function is given by the random measure

$$\hat{\boldsymbol{F}}_k(\boldsymbol{t}, \omega) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{1}_{\{\boldsymbol{X}_i(\omega) \leq \boldsymbol{t}\}}, \ (\boldsymbol{t}, \omega) \in \mathbb{R}^n \times \Omega.$$

For an aggregation measure $\rho_g \colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$ and a distribution function $\boldsymbol{F} \in \boldsymbol{\mathfrak{M}}^{\mathbf{1}}$ we define the sequence of estimators $\{\hat{\rho}_{g,k}\}_{k \geq 1}$ through its evaluation at the multivariate empirical distribution function. That is, for $k \geq 1$ we define

$$\hat{\rho}_{g,k}[\boldsymbol{F}](\omega) = \rho_g[\hat{\boldsymbol{F}}_k(\cdot, \omega)], \ \omega \in \Omega. \tag{2.3}$$

Note that for fixed $\boldsymbol{t} \in \mathbb{R}^n$ the multivariate empirical distribution function, $\hat{\boldsymbol{F}}_k(\boldsymbol{t}, \cdot)$, is a random variable and for fixed $\omega \in \Omega$ a distribution function. Hence, the estimator $\hat{\rho}_{g,k}[\boldsymbol{F}]$ is a random variable.

**Definition 2.4.2.** Let $\rho_g \colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$ be an aggregation measure and $\{\hat{\rho}_{g,k}\}_k$ the sequence of estimators defined in (2.3). We say that the aggregation measure $\rho_g$ is *robust* on $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}}^{\mathbf{1}}$ (equivalently $\rho_g[\cdot]$ is robust on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$) if for any $\boldsymbol{F_0} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ it holds that for all $\varepsilon > 0$, there exists $\delta > 0$ and $k_0 \in \mathbb{N}$ such that for all $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ and $k \geq k_0$ we have

$$d_P(\boldsymbol{F_0}, \boldsymbol{F}) < \delta \quad \Rightarrow \quad d_P\Big(\mathfrak{D}\Big(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\Big), \mathfrak{D}\Big(\hat{\rho}_{g,k}[\boldsymbol{F}]\Big)\Big) < \varepsilon.$$

We obtain a generalisation of Hampel's theorem, Theorem 2.3.2, to the multivariate case. The proof follows mostly the steps of the proof of Hampel's theorem, Theorem 2.3.2, for distribution function on the real line (Huber and Ronchetti, 2009).

**Theorem 2.4.3.** *Let $\rho_g \colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$ be an aggregation measure and $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}}^{\mathbf{1}}$. Assume that the sequence of estimators $\{\hat{\rho}_{g,k}\}_k$, defined in (2.3), is consistent for all $\boldsymbol{F_0} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$.*

*Then, the aggregation measure $\rho_g$ is continuous on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$ with respect to the Prokhorov distance if and only if it is robust on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$.*

*Proof.* Assume the aggregation measure $\rho_g$ is continuous with respect to $d_P$ on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$ and let $\boldsymbol{F_0} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$. Let $\varepsilon > 0$ and $k \in \mathbb{N}$ then for all $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ it holds that

$$\begin{aligned} d_P\Big(\mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\big),\mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F}]\big)\Big) &= d_P\Big(\mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}]\big),\mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_k]\big)\Big) \\ &\leq d_P\Big(\mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}]\big),\mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big)\Big) + d_P\Big(\mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big),\mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_k]\big)\Big). \end{aligned} \tag{2.4}$$

Note that $\rho_g[\boldsymbol{F_0}]$ is a degenerate random variable. For all $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$, the multivariate version of Glivenko-Cantelli states that the empirical distribution function $\hat{\boldsymbol{F}}_k(\cdot,\omega)$ converges for almost every $\omega$ to $\boldsymbol{F}$, as $k \to +\infty$, see Dudley (2002); Shorack and Wellner (2009). The first term on the right hand side in (2.4) can be made arbitrarily small (say $\varepsilon/2$) by choosing $k$ large enough since the aggregation measure is consistent at $\boldsymbol{F_0}$, that is $\hat{\rho}_{g,k}[\boldsymbol{F_0}] = \rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}] \to \rho_g[\boldsymbol{F_0}]$ in probability. Next we show that the second term in (2.4) is smaller than $\varepsilon/2$. By continuity of the aggregation function at $\boldsymbol{F_0}$ there exists $\delta > 0$ such that, for any $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$, $d_P(\boldsymbol{F_0}, \boldsymbol{F}) < \delta$ implies $|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]| < \frac{\varepsilon}{2}$. Thus, we obtain

$$\begin{aligned} P\Big(\Big|\rho_g[\boldsymbol{F_0}] - \rho_g[\hat{\boldsymbol{F}}_k]\Big| \leq \frac{\varepsilon}{2}\Big) &\geq P\Big(\Big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\Big| + \Big|\rho_g[\boldsymbol{F}] - \rho_g[\hat{\boldsymbol{F}}_k]\Big| \leq \frac{\varepsilon}{2}\Big) \\ &= P\Big(\Big|\rho_g[\boldsymbol{F}] - \rho_g[\hat{\boldsymbol{F}}_k]\Big| \leq \frac{\varepsilon}{2} - \Big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\Big|\Big), \end{aligned}$$

where $\frac{\varepsilon}{2} - \Big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\Big| > 0$. As the aggregation measure is consistent, for all $\gamma > 0$ we have $P\Big(\Big|\rho_g[\boldsymbol{F}] - \rho_g[\hat{\boldsymbol{F}}_k]\Big| \leq \gamma\Big) \to 1$ as $k \to +\infty$. Hence, choosing $k$ large enough, we obtain

$$P\Big(\Big|\rho_g[\boldsymbol{F_0}] - \rho_g[\hat{\boldsymbol{F}}_k]\Big| \leq \frac{\varepsilon}{2}\Big) \geq 1 - \frac{\varepsilon}{2},$$

which, by Strassen's theorem (Strassen, 1965), is equivalent to $d_P\Big(\mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_k]\big)\Big) < \frac{\varepsilon}{2}$.

For the converse assume that the aggregation measure is robust on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$. Note that for degenerate distribution functions on $\mathbb{R}$ the Prokhorov distance reduces to the absolute value. Let $\boldsymbol{F_0}, \boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ and interpreting $\rho_g[\boldsymbol{F}]$, $\rho_g[\boldsymbol{F_0}]$ as degenerate random variables we obtain for $k \in \mathbb{N}$

$$\begin{aligned}\Big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\Big| &= d_P\Big(\mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\rho_g[\boldsymbol{F}]\big)\Big) \\ &\leq d_P\Big(\mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\big)\Big) + d_P\Big(\mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F}]\big)\Big) \\ &\quad + d_P\Big(\mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F}]\big), \mathfrak{D}\big(\rho_g[\boldsymbol{F}]\big)\Big).\end{aligned}$$

The second term can be made small by robustness of the aggregation measures. The other two distances can be made arbitrarily small since the sequence of estimators is consistent for any $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$. □

An aggregation measure composed by a continuous aggregation function and a convex risk measure is consistent at each $\boldsymbol{F} \in \boldsymbol{\mathfrak{M}^1}$, that is $\hat{\rho}_{g,k}[\boldsymbol{F}] \to \hat{\rho}_g[\boldsymbol{F}]$ in probability (even $P$-a.s.). Hence, as a generalisation of Proposition 2.3.3 we obtain a one-to-one correspondence between robustness and continuity with respect to the Prokhorov distance.

**Proposition 2.4.4.** Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous aggregation function, $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ be a convex risk measure and $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}^1}$. Then, the aggregation measure $\rho_g\colon \boldsymbol{\mathcal{L}^1} \to \mathbb{R}$ is continuous with respect to $d_P$ on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$ if and only if it is robust on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$.

*Proof.* Let $\boldsymbol{F_0} \in \boldsymbol{\mathfrak{M}}^1$. It is enough to show that for a continuous $g$ and a convex risk measure $\rho$ the aggregation measure $\rho_g = \rho \circ T_g$ is consistent. We even show strong consistency, that is $\rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}](\omega) \to \rho_g[\boldsymbol{F_0}]$ for almost every $\omega \in \Omega$. Since convex risk

measures are continuous with respect to the Wasserstein distance (Shapiro et al., 2009, Prop. 6.8), we have to show that $d_W\Big(T_g[\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)], T_g[\boldsymbol{F}_{\mathbf{0}}]\Big) \to 0$ for almost every $\omega$.

The multivariate empirical distribution function $\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)$ converges for almost every $\omega$ to $\boldsymbol{F}_{\mathbf{0}}$, as $k \to +\infty$, see Dudley (2002); Shorack and Wellner (2009). In particular, for almost every $\omega$, $d_P\Big(\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega), \boldsymbol{F}_{\mathbf{0}}\Big) \to 0$, as $k \to +\infty$, and by continuity of the aggregation function, that is $T_g\colon \mathfrak{M} \to \mathfrak{M}$ is continuous w.r.t $d_P$, $d_P\Big(T_g[\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)], T_g[\boldsymbol{F}_{\mathbf{0}}]\Big) \to 0$, as $k \to +\infty$. For $k \in \mathbb{N}$ and $\omega \in \Omega$ denote by $\boldsymbol{X}_{\mathbf{0}k}^{\omega}$ a random variable that has distribution function $\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)$. Note that, by definition, $T_g[\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)] = \mathfrak{D}(g(\boldsymbol{X}_{\mathbf{0}k}^{\omega})) \in \mathfrak{M}^1$ and we have

$$\int_{\mathbb{R}} |y| \,\mathrm{d}T_g\Big[\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)\Big](y) = \int_{\mathbb{R}^n} |g(\boldsymbol{y})| \mathrm{d}\hat{\boldsymbol{F}}_{\mathbf{0}k}(\boldsymbol{y},\omega) = \frac{1}{k}\sum_{i=1}^{k} \Big|g\Big(\boldsymbol{X}_{\mathbf{0}i}(\omega)\Big)\Big|.$$

By the strong law of large numbers $\frac{1}{k}\sum_{i=1}^{k} \Big|g\Big(\boldsymbol{X}_{\mathbf{0}i}\Big)\Big| \to E(\,|\,g(\boldsymbol{X}_{\mathbf{0}})\,|\,) < +\infty$, $P$-a.s. Hence for almost every $\omega \in \Omega$

$$\int_{\mathbb{R}} |y| \,\mathrm{d}T_g\Big[\hat{\boldsymbol{F}}_{\mathbf{0}k}(\cdot,\omega)\Big](y) \quad \to \quad \int_{\mathbb{R}} |y| \,\mathrm{d}T_g[\boldsymbol{F}_{\mathbf{0}}](y), \text{ as } k \to +\infty.$$

The conclusion follows from Lemma 2.A.1. □

Analogously to Theorem 2.3.5, robustness of the aggregation measure $\rho_g$ depends on uniform integrability of the set of losses produced by the aggregation function $g$.

**Theorem 2.4.5.** *Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous aggregation function and $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ a convex risk measure. Then the aggregation measure $\rho_g\colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$ is robust on $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}}^{\mathbf{1}}$ if the set $g\Big(\boldsymbol{\mathcal{X}}\Big)$ is uniformly integrable.*

*Proof.* If $g(\boldsymbol{\mathcal{X}})$ is uniformly integrable the risk measure is continuous with respect to $d_P$, see Theorem 2.3.5. Therefore the composition $\rho_g = \rho \circ T_g$ is continuous with

respect to Prokhorov distance and by Proposition 2.4.4 the aggregation measure $\rho_g$ is robust on $\boldsymbol{\mathcal{X}}$. □

A similar problem is considered in Krätschmer et al. (2017), when the marginal distributions are fixed. Note that our extension of Hampel's classical definition of robustness to aggregation measures, Definition 2.4.2, requires the aggregation measure to be (relatively) insensitive to perturbations in the underlying distribution. Since the input of the aggregation measure is a random vector of risk factors, perturbation in the distribution can arise from changes in the marginals and / or the copula. Given Theorem 2.4.5, in order to characterise robustness of the aggregation measure $\rho_g$, it is necessary to study which properties of $g$ and the set $\boldsymbol{\mathcal{X}}$ produce a set of losses $g(\boldsymbol{\mathcal{X}})$ that is uniformly integrable. The next section investigates this issue.

**Remark 2.4.6.** It is not necessarily the case in practice that the multivariate distribution function of $\boldsymbol{F}$ is estimated by the empirical distribution of historical data; parametric statistical methods are typically used instead. Nonetheless, the definition of robustness used here remains relevant when calculating $\rho_g[\boldsymbol{F}]$ by Monte-Carlo simulation. In that context, $\mathbf{X}$ is simulated from model $\boldsymbol{F}$ and $\hat{\boldsymbol{F}}_k$ is interpreted as the empirical distribution function of the simulated observations. Then $\rho_g[\boldsymbol{F}]$ is calculated via evaluation of $\rho_g[\hat{\boldsymbol{F}}_k]$, as is typically done in insurance internal models (SCOR, 2008). It is desirable that small changes in the assumed distribution $\boldsymbol{F}$ of risk factors does not produce excessive variation in the estimated aggregate risk.

### 2.4.2 Aggregation robustness and linear growth

A typical setup in risk management is linear risk aggregation, for example when aggregating different lines of business or positions in a portfolio, such that

$$\rho\Big(g(\boldsymbol{X})\Big) = \rho(X_1 + \cdots + X_n), \text{ for } \boldsymbol{X} \in \boldsymbol{\mathcal{L}}^{\mathbf{1}}. \tag{2.5}$$

By Sklar's theorem the distribution of vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is specified through its marginals and its dependence structure (copula). Statistically, estimating copulas can be very challenging and often relies on expert judgement. Since diverse dependence structures can lead to substantial differences in aggregate risk, risk management is especially concerned about misspecification in the copula. A substantial literature exists on *dependence uncertainty*, including calculations of upper and lower bounds for (2.5), for fixed marginals $X_i \sim F_i$, $i = 1, \ldots, n$ and an unspecified copula, see Embrechts, Puccetti and Rüschendorf (2013); Wang et al. (2013); Bernard et al. (2014) and references therein.

Furthermore, Embrechts et al. (2015) show that, when $\rho$ is a spectral risk measure, the aggregation measure defined through (2.5) is robust on the set $\{(X_1, \ldots, X_n) \,|\, X_i \sim F_i,\ i = 1 \ldots, n\}$, where $F_1, \ldots F_n \in \mathfrak{M}^1$ are fixed marginal distributions. Taking a step further, Krätschmer et al. (2017) consider robustness of convex risk measures composed with non-linear aggregation functions for fixed marginals, see discussion after Theorem 2.4.8. Here, we build on Embrechts et al. (2015); Krätschmer et al. (2017) by considering robustness in the more general case of uncertainty in both the dependence structure and the marginals of the model input $\boldsymbol{X}$. Theorem 2.4.8 below shows that robustness is guaranteed if the aggregation function satisfies a linear growth condition in the tail, similar to that of Krätschmer et al. (2017), and the marginals belong to uniformly integrable sets.

For sets of univariate distribution functions $\mathfrak{N}_i \subset \mathfrak{M}^1$, $i = 1, \ldots n$, we define the set of all possible random vectors $\boldsymbol{X} = (X_1, \ldots, X_n)$ with marginals $F_{X_i}$ belonging to the corresponding sets $\mathfrak{N}_i$, $i = 1, \ldots, n$, through

$$\boldsymbol{\mathcal{C}}(\mathfrak{N}_1, \ldots \mathfrak{N}_n) = \Big\{(X_1, \ldots, X_n) \,|\, F_{X_i} \in \mathfrak{N}_i \subset \mathfrak{M}^1,\ i = 1, \ldots n\Big\} \subset \boldsymbol{\mathcal{L}^1}.$$

**Definition 2.4.7.** We say an aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ possesses the *linear growth condition in the tail*, if there exist $A, L, M > 0$ such that

$$|g(\boldsymbol{x})| \leq A + L|\boldsymbol{x}|, \text{ for all } |\boldsymbol{x}| > M.$$

Continuity of $g$ combined with linear growth in the tail as in Definition 2.4.7 guarantee that $g(\boldsymbol{X}) \in \mathcal{L}^1$ for $\boldsymbol{X} \in \boldsymbol{\mathcal{L}^1}$.

**Theorem 2.4.8.** *Let the sets $\mathfrak{U}_1, \dots \mathfrak{U}_n \in \mathfrak{M}^1$ be uniformly integrable, the function $g$ be continuous and satisfy the linear growth condition in the tail, and $\rho$ be a convex risk measure. Then the aggregation measure $\rho_g\colon \boldsymbol{\mathcal{L}^1} \to \mathbb{R}$ is robust on $\boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n)$.*

*Proof.* By Theorem 2.4.5 it is enough to show that $g\Big(\boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n)\Big)$ is uniformly integrable. The aggregation function $g$ is continuous on the compact set $\{\boldsymbol{x} \in \mathbb{R}^n \mid |\boldsymbol{x}| \leq M\}$, hence there exists $C > 0$ such that $\sup_{|\boldsymbol{x}| \leq M} |g(\boldsymbol{x})| \leq C$.

Let $\boldsymbol{X} \in \boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n)$. For $K > \max\{A, C\}$ we have that $\{|\boldsymbol{X}| \leq M \cap |g(\boldsymbol{X})| > K\} = \varnothing$, thus

$$\begin{aligned}
\sup_{g(\boldsymbol{X}) \in g(\boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n))} E\Big(|g(\boldsymbol{X})| \mathbb{1}_{\{|g(\boldsymbol{X})| > K\}}\Big) &= \sup_{\boldsymbol{X} \in \boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n)} E\Big(|g(\boldsymbol{X})| \mathbb{1}_{\{|g(\boldsymbol{X})| > K\}} \mathbb{1}_{\{|\boldsymbol{X}| > M\}}\Big) \\
&\leq L \sup_{\boldsymbol{X} \in \boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n)} E\Big(\sum_{i=1}^n |X_i| \mathbb{1}_{\left\{L \sum_{i=1}^n |X_i| > K - A\right\}}\Big) \\
&\quad + A \sup_{\boldsymbol{X} \in \boldsymbol{\mathcal{C}}(\mathfrak{U}_1, \dots \mathfrak{U}_n)} P\Big(L \sum_{i=1}^n |X_i| > K - A\Big).
\end{aligned} \tag{2.6}$$

The first term in (2.6) can be bounded as follows. Note that for $d \geq 0$ and $x_1, \dots, x_n \in \mathbb{R}$, there exists $j$ such that $\max_{i=1,\dots,n} |x_i| \mathbb{1}_{\{\max_{i=1,\dots,n} |x_i| > d\}} = |x_j| \mathbb{1}_{\{|x_j| > d\}} \leq \sum_{i=1}^n |x_i| \mathbb{1}_{\{|x_i| > d\}}$.

Therefore,

$$
\begin{aligned}
L \sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)} & E\Big(\sum_{i=1}^n |X_i|\mathbb{1}_{\left\{L\sum_{i=1}^n |X_i|>K-A\right\}}\Big) \\
&\le L \sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)} E\Big(n \max_{i=1,\dots,n} |X_i|\mathbb{1}_{\left\{nL \max_{i=1,\dots,n} |X_i|>K-A\right\}}\Big) \\
&\le nL \sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)} \sum_{i=1}^n E(|X_i|\mathbb{1}_{\{|X_i|>(K-A)/(nL)\}}) \\
&\le nL \sum_{i=1}^n \sup_{F_{X_i}\in\mathfrak{U}_i} E(|X_i|\mathbb{1}_{\{|X_i|>(K-A)/(nL)\}}) \to 0,
\end{aligned}
$$

as $K \to +\infty$, by uniform integrability of each $\mathfrak{U}_i$. For the second term in (2.6) we use Markov's inequality

$$
A \sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)} P\Big(\sum_{i=1}^n |X_i| > \frac{K-A}{L}\Big) \le \frac{AL}{K-A} \sum_{i=1}^n \sup_{F_{X_i}\in\mathfrak{U}_i} E(|X_i|),
$$

which goes to zero as $K \to +\infty$. □

Note that Theorem 2.4.8 requires assumptions on the marginal distributions of $\boldsymbol{X}$, but not on its dependence structure. Hence robustness of convex risk measures holds even in the presence of complete dependence uncertainty, where no information on the copula exists. Theorem 2.4.8 offers a slight generalisation of Theorem 4.23 in Krätschmer et al. (2017) to the case of uncertainty in the marginal distributions. Also, Krätschmer et al. (2017) require a global linear growth condition, while we use linear growth in the tail combined with continuity of $g$.

An immediate consequence of Theorem 2.4.8 involves linear aggregation.

**Corollary 2.4.9.** Let the function $g$ be given by $g(\boldsymbol{x}) = \sum_{i=1}^n x_i$, $\boldsymbol{x} \in \mathbb{R}^n$. For a convex risk measure $\rho$, the aggregation measure $\rho_g$ is robust on $\mathcal{C}(\mathfrak{U}_1, \dots \mathfrak{U}_n)$, with $\mathfrak{U}_i$, $i = 1, \dots, n$, uniformly integrable.

There also exist many relevant continuous non-linear aggregation functions that satisfy the linear growth condition of Definition 2.4.7. It is easiest to envisage such situations arising in the context of reinsurance, with the elements of the random vector $\boldsymbol{X}$ representing insurance liabilities (losses from lines of business or individual policies). Then, by standard considerations of insurable interest and moral hazard, it is not plausible to have (re)insurance portfolio losses that increase in $\boldsymbol{X}$ faster than linearly. Note that optimal Pareto reinsurance contracts are typically Lipschitz continuous (Dana and Scarsini, 2007) and hence possess the linear growth condition. For example, a reinsurance company faces the aggregate risk of excess-of-loss reinsurance contracts on individual risks $X_1, \ldots, X_n$ with deductibles $d_i$ and limits $c_i > d_i$, $i = 1, \ldots, n$, such that

$$g(\boldsymbol{X}) = \sum_{i=1}^{n} (X_i - d_i)_+ - (X_i - c_i - d_i)_+,$$

where $(x)_+ = \max\{x, 0\}$. Alternatively, a reinsurance company taking the risk that an aggregated portfolio exceeds $c > 0$, faces claim

$$g(\boldsymbol{X}) = \left( \sum_{i=1}^{n} X_i - c \right)_+ .$$

Note that in the first example $g$ is constant for large $\boldsymbol{x}$ and in the second case it is linear in its marginals, hence fulfilling in both cases the linear growth condition in the tail.

Alternatively, one could view $g(\boldsymbol{X})$ as a portfolio of financial derivatives with underlyings $\boldsymbol{X}$, such that $g(\boldsymbol{X}) = \sum_{i=1}^{n} h_i(X_i)$. Standard options, even leveraged ones with pay-offs of the form $h_i(x) = (\lambda x - c)_+$, $\lambda > 1$, satisfy the linear growth condition. However, note that other exotic options, such as powered options of the form $h_i(x) = \big((x - c)_+\big)^p$, with $p > 1$, do not satisfy the linear growth condition. To

achieve robustness for such pay-offs, one would need to restrict $\boldsymbol{X}$ to $\boldsymbol{\mathcal{L}}^p$, see Section 2.3.4. For details on such derivatives see Jarrow and Turnbull (2000, pp. 168-169).

### 2.4.3 Aggregation through compound distributions

A common form of aggregation in insurance (as well as operational and credit risk modelling), takes place via compound distributions that model the future total claim amount as a random sum of individual claims. Within a specific (homogeneous) line of business, individual claims are modelled as independent and identically distributed positive random variables $X_i$ and the (unknown) number of claims through a (discrete and random) count variable $N$ independent of the $X_i$. The total claim amount $X_1 + \cdots + X_N$ cannot be readily expressed via an aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$. However, the distribution function of the random sum can be straightforwardly defined through an aggregation operator $T$ acting on distributions, namely

$$T[\cdot,\cdot]\colon \mathfrak{M}^1 \times \mathfrak{M}^1 \to \mathfrak{M}^1; \quad T[F,G] = \mathfrak{D}\Big(\sum_{i=1}^{N} X_i\Big),\ X_i \sim_{i.i.d.} F \text{ independent of } N \sim G. \tag{2.7}$$

Therefore, $T[F,G] = \int_0^{+\infty} F^{*(n)}(\cdot)\mathrm{d}G(n)$, where $F^{*(n)}$ is the $n$-th convolution of $F$.

**Theorem 2.4.10.** *Let $\mathfrak{U}$ be a uniformly integrable set of distribution functions on $[0,+\infty)$ and $\mathfrak{N}$ a uniformly integrable set of distributions on the non-negative integers, such that*

$$\int_0^{+\infty} x\,\mathrm{d}F^*(x) < +\infty \text{ and } \int_0^{+\infty} x\,\mathrm{d}G^*(x) < +\infty,$$

*where $F^*$ and $G^*$ are distribution functions given by $F^* = \inf_{F\in\mathfrak{U}} F$ and $G^* = \inf_{G\in\mathfrak{N}} G$ respectively. Let the operator $T$ be defined by (2.7) and $\rho$ be a convex risk measure. Then, the aggregation measure defined by $\rho \circ T\colon \mathfrak{M}^1 \times \mathfrak{M}^1 \to \mathbb{R}$ is robust on $\mathfrak{U} \times \mathfrak{N}$.*

*Proof.* By Theorem 2.4.5 it is enough to show that the set $\{T[F,G] \,|\, F \in \mathfrak{U},\, G \in \mathfrak{N}\}$ is uniformly integrable. Note that $F^*$ is a distribution function. Indeed, $F^*$ is non-decreasing, right continuous (the infimum of a family of right continuous non-decreasing, hence upper semi-continuous, functions is right continuous, see Lemma 2.39 in Aliprantis and Border (2006)) and

$$\lim_{x\to+\infty}(1-F^*(x)) \leq \lim_{x\to+\infty}\sup_{F\in\mathfrak{U}} x(1-F(x)) \leq \lim_{x\to+\infty}\sup_{F\in\mathfrak{U}}\int_{y>x} y\mathrm{d}F(y) = 0,$$

by uniform integrability. Analogously, $\inf_{G\in\mathfrak{N}} G$ is a distribution function on the non-negative integers. Choose $F \in \mathfrak{U}$ and $G \in \mathfrak{N}$ and denote $X_i \sim_{i.i.d.} F$ and $N \sim G$ independent of the $X_i$. Similarly, denote $X_i^* \sim_{i.i.d.} F^*$ and $N^* \sim G^*$ independent of the $X_i^*$ and note that $X_i^*$ and $N^*$ first-order stochastically dominate $X_i$ and $N$, respectively. As first-order stochastic dominance is preserved under compounding (Denuit et al., 2006, Prop. 3.3.31), $\sum_{i=1}^{N} X_i$ is lower than $\sum_{i=1}^{N^*} X_i^*$ in first-order stochastic dominance. The result follows from Lemma 2.3.9 and the fact that the compound sum $\sum_{i=1}^{N^*} X_i^*$ is integrable, given the integrability of $X_i^*$ and $N^*$. □

Examples of sets of distribution functions on the non-negative integers fulfilling the assumptions of Theorem 2.4.10 include the Poisson distribution with parameter $0 < \lambda \leq \overline{\lambda}$ and the Geometric with $p \geq \underline{p} > 0$, see Table 3.1 in Denuit et al. (2006). For the claim size distribution, an example is the family of Pareto distributions $F(x) = 1 - x_m^{\alpha} x^{-\alpha}$ with parameters $0 < x_m \leq \overline{x}_m$ and $\alpha \geq \underline{\alpha} > 1$ or, more generally, the set of GPDs, $\{G_{\xi;\sigma} \,|\, \sigma \leq \overline{\sigma},\, 0 < \xi \leq \overline{\xi}\}$, $\overline{\xi} < 1$.

## 2.5 Comparison to robustness regions of Value-at-Risk

In this section we compare the robustness properties of the popular non-convex risk measure VaR to those of the convex risk measures studied in this chapter. Since different risk measures are robust on different sets, the choice of risk measure should also reflect information on the plausible sets of distribution functions expected to be encountered in particular applications.

VaR at level $\alpha \in (0,1)$ is defined as the left-sided $\alpha$-quantile, $\mathrm{VaR}_\alpha[F] = F^{-1}(\alpha) = \inf\{y \in \mathbb{R} \,|\, F(y) \geq \alpha\}$. It is known that $\mathrm{VaR}_\alpha$ is not robust on the whole of $\mathfrak{M}^1$; however, it is robust on the set of distribution functions that are strictly increasing in a neighbourhood of their $\alpha$-quantile (Hampel, 1971; Cont et al., 2010). In particular, VaR is not robust on discrete random variables and hence the set where VaR is not robust is dense in $\mathfrak{M}^1$.

The following insurance example, where strict increasingness is not satisfied, leads to non-robustness of VaR. Consider the risk exposure $Y = \min\{X, d\}$, $X \in \mathcal{L}^1$, that occurs when an insurer with exposure $X$ buys reinsurance protection with deductible $d \geq 0$. The distribution of $Y$, $F_Y(x) = F_X(x)\mathbb{1}_{\{x<d\}} + \mathbb{1}_{\{x\geq d\}}$, is flat for all $x > d$, hence $\mathrm{VaR}_\alpha$ is not robust at $F_Y$ whenever $\alpha \geq F_X(d)$.

Thus, neither convex risk measures such as ES nor VaR, are robust on $\mathcal{L}^1$. VaR requires strictly increasing distribution functions. Convex risk measures like ES place requirements on the tail of the underlying distribution functions via the uniform integrability condition, see Theorem 2.3.5. A comparative assessment of those two risk measures thus relies on whether strict increasingness or uniform integrability is a more realistic constraint on the set of distributions on which the risk measure is to be evaluated. This depends on the context of the application. For example, in reinsurance

problems where distributions with constant parts can occur, uniform integrability may be a more suitable assumption. On the other hand, when dealing with an asset return with an approximately bell-shaped density but arbitrarily heavy tails, strict increasingness of the distribution appears to be a more appropriate condition.

Turning now to the case of risk aggregation, consider the aggregation measure defined by $\mathrm{VaR}_{\alpha,g}\colon \boldsymbol{\mathcal{L}}^{\mathbf{1}} \to \mathbb{R}$, where $g\colon \mathbb{R}^n \to \mathbb{R}$ is an aggregation function. The aggregation measure $\mathrm{VaR}_{\alpha,g}$ will not be robust if the distribution of $g(\boldsymbol{X})$ is constant in a neighbourhood of $F^{-1}_{g(\boldsymbol{X})}(\alpha)$. Such flat regions can emerge due to the nature of the function $g$. For instance, in a slight generalisation of the previous example, for an insurance company that buys an unlimited layer of reinsurance protection for its portfolio, we have $g(\boldsymbol{X}) = \min\{\sum_{i=1}^n X_i, d\}$.

Flat regions in the distribution of $g(\boldsymbol{X})$ can also appear through the effect of the dependence structure of $\boldsymbol{X}$. This is exemplified by the special case of linear portfolio aggregation, $g(\mathbf{x}) = \sum_{i=1}^n x_i$. Then the aggregation measure $\mathrm{VaR}_{\alpha,g}$ is not robust on a set $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}}^{\mathbf{1}}$ if there exists an input vector $\boldsymbol{X} \in \boldsymbol{\mathcal{X}}$ such that $X_1 + \cdots + X_n$ is discrete for large values. Example 2.2 in Embrechts et al. (2015) provides explicit choices of marginals and copulas that lead to non-robustness of the aggregation measure $\mathrm{VaR}_{\alpha,g}$ through the construction of a degenerate aggregate risk. The problem of the existence of a dependence structure of random variables $X_1, \ldots, X_n$, such that the aggregated risk $X_1 + \cdots + X_n$ is almost surely constant, is extensively studied in probability theory and risk management (Rüschendorf, 1982; Makarov, 1982). Examples of distribution functions include $F_1 = \cdots = F_n$ being Gaussian or Cauchy; we refer the reader to Wang et al. (2013) and references therein in the context of risk management.

In quantitative risk management applications, one is often concerned about aggregate risks. Seldom is a risk measure evaluated on a random loss that does not in turn depend on further risk factors. A particular example is the use of internal

models in insurance for calculating capital requirements across the portfolio. Compared to evaluating a risk measure on a real-valued random variable, in risk aggregation, there is the additional complication of the dependence structure of the input vector. Thus, there are two sources of uncertainty, in the marginal distributions and in the dependence structure. Modelling accurately the dependence structure is usually more challenging than modelling marginals, due to a lack of extensive multivariate data sets. Therefore, it is critical that the risk measure is robust to changes in the dependence structure.

We have seen that robustness of aggregation measures derived from convex risk measures, such as ES, depends on weak assumptions on the aggregation function $g$ and the marginals, while no requirements are placed on the dependence structure. On the other hand, robustness of VaR requires restricting both the form of the aggregation function $g$ and the possible dependence structures of the input vector. In applications such as the internal modelling performed by insurers, such constraints are not necessarily realistic. Thus our findings indicates that in applications where (non-linear) aggregations are present and high dependence uncertainty persists, the use of convex risk measures may be preferable to that of VaR.

# Appendix 2.A Wasserstein space

For $F, G \in \mathfrak{M}^1$, the Wasserstein distance (Dobrushin, 1970; Givens and Rae, 1984) is given by

$$d_W(F,G) = \int_{\mathbb{R}} |F(x) - G(x)| \mathrm{d}x = \int_0^1 |F^{-1}(u) - G^{-1}(u)| \mathrm{d}u,$$

where $F^{-1}(u) = \inf\{y \in \mathbb{R} \,|\, F(y) \geq u\}$, $u \in [0,1)$, is the generalised inverse and we identify $\inf \varnothing = -\infty$.

**Lemma 2.A.1.** (Lemma 8.3 in Bickel and Freedman (1981))

For $F, F_k \in \mathfrak{M}^1$, $k \geq 1$ the following are equivalent

i) $d_W(F_k, F) \to 0$, as $k \to +\infty$.

ii) $d_P(F_k, F) \to 0$ and $\int_{\mathbb{R}} |x| \mathrm{d}F_k(x) \to \int_{\mathbb{R}} |x| \mathrm{d}F(x)$, as $k \to +\infty$.

iii) $d_P(F_k, F) \to 0$ and the set $\{F_k \,|\, k \geq 1\}$ is uniformly integrable.

**Lemma 2.A.2.** A risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is continuous with respect to the norm $\|\cdot\|_1$ on $\mathcal{L}^1$ if and only if it is continuous with respect to the Wasserstein distance on $\mathcal{L}^1$.

*Proof.* Assume that the risk measure is continuous with respect to $\|\cdot\|_1$. On $\mathcal{L}^1$ a sequence of random variables $X_n$ converges in the Wasserstein distance to $X$ if and only if there exist random variables $\tilde{X}_n$ on $\mathcal{L}^1$ with the same distribution as $X_n$ and $\tilde{X}$ with the same distribution as $X$ such that $\|\tilde{X}_n - \tilde{X}\|_1 \to 0$, see Theorem 3.5 in Krätschmer et al. (2014). Hence by law-invariance of the risk measure

$$\rho(X_n) = \rho(\tilde{X}_n) \to \rho(\tilde{X}) = \rho(X), \text{ as } n \to +\infty.$$

For $X, Y \in \mathcal{L}^1$ the inequality $d_W(X, Y) \leq \|X - Y\|_1$ implies that a sequence converging in $\|\cdot\|_1$ also converges in the Wasserstein distance. Hence continuity with respect to $d_W$ implies continuity with respect to $\|\cdot\|_1$. □

On the set of integrable distribution functions over $\mathbb{R}^n$, that is $\boldsymbol{\mathfrak{M}^1} = \mathfrak{D}(\boldsymbol{\mathcal{L}^1})$, the Wasserstein distance is defined for $\boldsymbol{F}, \boldsymbol{G} \in \boldsymbol{\mathfrak{M}^1}$ by

$$d_W(\boldsymbol{F}, \boldsymbol{G}) = \inf \left\{ E\Big( \|\boldsymbol{X} - \boldsymbol{Y}\|_1 \Big) \,\Big|\, \boldsymbol{X} \sim \boldsymbol{F},\, \boldsymbol{Y} \sim \boldsymbol{G} \right\},$$

where the infimum is taken over all joint distribution functions of dimension $2n$ with marginals $\boldsymbol{F}$ and $\boldsymbol{G}$ of size $n$. Note that on the real line we have the dual representation

$d_W(F,G) = \inf\{E(|X-Y|) \,|\, X \sim F, Y \sim G\} = \int_{\mathbb{R}} |F(x) - G(x)| \mathrm{d}x, \; F, G \in \mathfrak{M}^1$ (Vallender, 1974).

# Chapter 3

# Reverse sensitivity testing: What does it take to break the model?

This chapter comprises the publication Pesenti et al. (2018*b*) and earlier versions have been presented at the University of Trieste (2018), at the University of Bologna (2018), at the *Insurance Data Science Conference* (London, 2018), at the *Actuarial and Financial Mathematics Conference* (Brussels, 2018), at the *RSS Applied Probability Section on Probability in Actuarial Science* (London, 2017), at the *41st AMASES Conference* (Cagliari, 2017), at the workshop on *Risk Measurement and Regulatory Issues in Business* (Montréal, 2017) and at the *21st International Congress on Insurance: Mathematics and Economics* (Vienna, 2017).

## 3.1 Introduction

### 3.1.1 Problem framing and contribution

In this chapter we develop a sensitivity analysis framework appropriate for contexts where the following considerations, typical in several fields, including probabilistic safety assessment, reliability analysis and financial / insurance risk management (Saltelli

and Tarantola, 2002; Aven and Nøkland, 2010; Gourieroux et al., 2000; Tsanakas and Millossovich, 2016), hold:

- Model inputs are uncertain, hence sensitivity and uncertainty analyses are interlinked and global sensitivity analysis methods are called for.
- A decision criterion is derived by applying a risk measure on the distribution of the output. Risk measures are used in a variety of operations research and risk analysis applications, with VaR and ES particularly popular choices; indicatively see Rockafellar and Uryasev (2002); Tapiero (2005); Gotoh and Takano (2007); Ahmed et al. (2007); Asimit et al. (2017).
- The value of the risk measure gives an indication of criticality for the system whose uncertainty is analysed. For example, in the context of financial risk management, high values of the output's risk measure may indicate that a portfolio is not admissible, e.g. due to regulatory constraints (Artzner et al., 1999). In the context of probabilistic safety assessment, legislation postulates acceptable probabilities of failure, e.g. of fatality numbers exceeding a threshold (Borgonovo and Cillo, 2017).
- The relationship between model inputs and outputs is complex and not necessarily given in analytical form. Furthermore, evaluations of the model are computationally expensive inducing the need to estimate sensitivity measures from a single sample of input and output scenarios (Beckman and McKay, 1987; Plischke et al., 2013).

We propose a sensitivity analysis framework, adapted to the above context, termed *reverse sensitivity testing*. We work in the standard setting of sensitivity analysis, where a number of random input factors are mapped to a random output via an aggregation function, see also Section 1.2. The *baseline probability measure* summarises

the distribution of inputs and output in the current specification of the model. Our reverse sensitivity testing framework comprises the following steps. First, an output stress is defined, corresponding to an increase in the value of the output risk measure. We focus on the widely used risk measures VaR and ES. The increase in the value of the risk measure is specified so as to produce a stress that is problematic to a decision maker. For example, in a capital management context, a stress on VaR may lead to a situation where insufficient assets are available to satisfy regulatory requirements.

Secondly, a *stressed probability measure* is derived. This is a probability (a) under which the risk measure applied to the model output is at its stressed level and (b) that minimises the Kullback-Leibler (KL) divergence subject to appropriate constraints on the output probability distribution. Thus the stressed probability leads to the most plausible alternative model, under which the output distribution is subjected to the required stress. We derive analytical solutions of the stressed probability measure under an increase of VaR and ES. The form of the solutions allows for numerically efficient implementation via a single set of Monte Carlo simulations.

Finally, the distribution of individual input factors is examined under the baseline and stressed models. Substantial changes in the distribution of a particular input indicate a large sensitivity to that input. A new class of *reverse sensitivity measures* is introduced, quantifying these input changes, and extended to control for statistical dependence between inputs. The sensitivity measures are then used to identify the most influential input factors; in a sense, those factors that may be responsible for 'breaking the model'.

### 3.1.2 Relation to the literature

Prominent sensitivity analysis methods use a (Hoeffding) decomposition of the output variance (Sobol, 1993; Wagner, 1995; Saltelli et al., 2000; Saltelli and Tarantola, 2002;

Saltelli et al., 2008), as well as moment independent approaches (Borgonovo, 2007; Borgonovo et al., 2011). Alternative methods consider partial derivatives of statistical functionals of the output distribution in the direction of parameters of interest, see Glasserman and Liu (2010) for expectation-type and Hong (2009); Tsanakas and Millossovich (2016) for percentile-based functionals.

The sensitivity measures we propose in the present chapter reflect the joint distribution of individual input factors and output; hence our proposed method remains formally within the unifying framework discussed by Borgonovo et al. (2016) and thus are (distantly) related to variance-based and moment-independent sensitivity measures. Nonetheless, our proposed sensitivity measures are conceptually different compared to variance-based metrics and other current approaches in the literature. First, our approach involves an assessment of output uncertainty via tail-risk measures rather than the variance. Second, we adopt a reverse approach of stressing the output and then evaluating the impact on the inputs. Our method allows for flexibility in the stress level on the output, giving a nuanced picture of the sensitivity of input factors. Furthermore, the sensitivity measures we propose can take both negative and positive values, indicating the direction in which input factors affect the output. Thus, we view our proposed sensitivity analysis framework as complementary rather than competing with established methods, as it aims to address different questions.

Conceptually, the reverse direction (from output to input) of the proposed method, is related to *regionalised sensitivity analysis* methods (Spear et al., 1994; Osidele and Beck, 2004). However, there is a key difference between regionalised sensitivity analysis and our approach: in the former, states of the output are identified that are 'out of control', while in the latter what is 'out-of-control' are not individual states but specifications of the output distribution. The numerical tractability of our framework in a Monte Carlo setting is akin to Beckman and McKay (1987).

In the practice of financial risk management and regulation, *reverse stress testing*, starting with a stressed output state and studying the corresponding surface of scenarios that provide the adverse outcome, is frequently used (BCBS, 2013; EIOPA, 2009). For example, "reverse stresses that result in a depletion of capital..." (Lloyd's, 2016) are used in the validation of insurance risk models. The academic literature on reverse stress testing is relatively sparse, with a recent focus towards identifying most likely stress scenarios (McNeil and Smith, 2012; Breuer et al., 2012; Glasserman and Xu, 2014). Our approach differs from reverse stress testing, in that we consider most influential factors in relation to changes in the output distribution and not a particular output state.

The KL-divergence has been widely used in financial risk management, in particular in the context of model uncertainty, where several plausible specifications of the probability measure may co-exist. For example, Breuer and Csiszár (2013); Glasserman and Xu (2014); Blanchet et al. (2017) consider the worst-case probability measure with respect to all probabilities lying within a KL-divergence radius of the baseline probability. In contrast, reflecting our focus on sensitivity rather than model uncertainty, we consider the probability measure with minimal KL-divergence that satisfies given constraints. Our approach is closely related to the work of Cambou and Filipović (2017) with probability set constraints and Weber (2007) with risk measure constraints, see Section 3.3 for a detailed comparison.

### 3.1.3 Structure of the chapter

In Section 3.2, some preliminaries on risk measures and the KL-divergence are given. In Section 3.3, the optimisation problem yielding stressed probability measures is stated and solved under constraints arising from different risk measures, with emphasis on VaR and ES. Explicit solutions allow easy implementation and inspection of

the distributional changes arising. Furthermore, we discuss an extension where the distributions of multiple outputs are stressed. The solutions and their properties are illustrated through an example of a non-linear insurance portfolio evaluated using Monte Carlo simulation.

In Section 3.4 we propose metrics tailored to the proposed reverse sensitivity testing approach. A comparison study of the proposed *reverse sensitivity measures* with moment independent and variance-based sensitivity measures is conducted, illustrating differences between the concepts but also demonstrating coherent sensitivity rankings of inputs. In addition, a generalisation of reverse sensitivity measures is proposed, with the aim of controlling for dependence between input factors.

Section 3.5 demonstrates the applicability of the reverse sensitivity testing framework to a commercially used insurance portfolio model.

Appendix 3.A is devoted to a comparison of the stressed and the baseline probability measures through stochastic order relations, in order to establish formal properties of the proposed framework. We find that the distribution of the output under the baseline probability is first-order stochastically dominated by that under the stressed probability. A similar dominance relation is provided for input factors, under the assumption of a non-decreasing aggregation function and positive dependence between input factors. Moreover, stressed probability measures stemming from different stress severities lead to stochastically ordered input factors and output. All proofs are provided in Appendix 3.B.

## 3.2 Preliminaries

As introduced in Section 1.2, we consider the measurable space $(\Omega, \mathcal{A})$ and denote by $\mathcal{P}$ the set of all probability measures on $(\Omega, \mathcal{A})$. For a random variable $Z$ on $(\Omega, \mathcal{A})$ we write $F_Z^Q(\cdot) = Q(Z \leq \cdot)$ for its distribution under $Q \in \mathcal{P}$, and similarly, $E^Q(\cdot)$ for

its expectation. For a random variable $Z$, we denote by $\mathrm{VaR}_\alpha^Q(Z) = F_Z^{Q,-1}(\alpha)$ the VaR at level $\alpha \in [0,1]$ and by $\mathrm{ES}_\alpha^Q(Z) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u^Q(Z)\mathrm{d}u$ the ES at level $\alpha \in [0,1)$. Throughout, we use the Kullback-Leibler divergence (*KL-divergence*, Kullback and Leibler (1951)) as a measure of discrepancy between two probability measures. For $Q^1, Q^2 \in \mathcal{P}$, the KL-divergence, also known as relative entropy, of $Q^1$ with respect to $Q^2$ is defined by

$$D_{\mathrm{KL}}(Q^1 \| Q^2) = \begin{cases} \int \frac{dQ^1}{dQ^2} \log\left(\frac{dQ^1}{dQ^2}\right) \mathrm{d}Q^2 & \text{if } Q^1 \ll Q^2 \\ +\infty & \text{otherwise,} \end{cases}$$

where we use the convention that $0\log(0) = 0$. The KL-divergence is non-negative, vanishes if and only if $Q^1 \equiv Q^2$, and is in general not symmetric (Kullback, 1959; Cover and Thomas, 2012). The KL-divergence is a special case of the class of *$f$-divergences*, first introduced by Ali and Silvey (1966), for the choice $f(x) = x\log(x)$, $x > 0$. For a given convex function $f$, the $f$-divergence of $Q^1$ with respect to $Q^2$, for any $Q^1, Q^2 \in \mathcal{P}$, is defined through $D_f(Q^1 \| Q^2) = \int f\left(\frac{dQ^1}{dQ^2}\right)\mathrm{d}Q^2$ .

## 3.3 Deriving the stressed model

### 3.3.1 Problem statement

We call the triple $(\boldsymbol{X}, g, P)$, the *baseline model* with *baseline probability measure* $P \in \mathcal{P}$. The probability $P$ is seen as encoding current beliefs regarding (or software implementation of) the distribution of $\boldsymbol{X}$. Under the baseline probability $P$ we suppress the superscript and write, for example, $F_Z(\cdot) = F_Z^P(\cdot)$ and $E(\cdot) = E^P(\cdot)$, and analogously for risk measures, $\mathrm{VaR}_\alpha(\cdot) = \mathrm{VaR}_\alpha^P(\cdot)$ and $\mathrm{ES}_\alpha(\cdot) = \mathrm{ES}_\alpha^P(\cdot)$. We call any $Q \in \mathcal{P}$ an *alternative probability measure* and $(\boldsymbol{X}, g, Q)$ an *alternative model*. A

Radon-Nikodym (RN) density is a non-negative random variable $\zeta$ on $(\Omega, \mathcal{A})$ such that $E(\zeta) = 1$. We denote by $Q^\zeta$ the probability measure which is absolutely continuous with respect to $P$ with RN-density $\zeta$, that is, $\zeta = \frac{\mathrm{d}Q^\zeta}{\mathrm{d}P}$.

The starting point of reverse sensitivity analysis is to define a *stress* on the distribution of the output that would be problematic to a decision maker, such as a risk manager or regulator. For example, one may require that the probability of a particular event, representing system failure, increases to an extent that the risk of failure is no longer acceptable. Specific stress definitions using different risk measures are discussed in Sections 3.3.2-3.3.6. Subsequently, we call $(\boldsymbol{X}, g, Q)$ a *stressed model* with *stressed probability measure* $Q \in \mathcal{P}$ if, under $Q$, the output $Y$ fulfils a set of probabilistic constraints (the stress) and $Q$ has minimal KL-divergence with respect to $P$. Thus, a stressed probability measure is defined as a solution to

$$\min_{Q \in \mathcal{P}} D_{\mathrm{KL}}(Q \| P), \quad \text{s.t. constraints on the distribution of } Y \text{ under } Q \text{ hold.} \tag{3.1}$$

The optimisation problem (3.1) is robust in the sense that convergence in the KL-divergence implies weak convergence of the probability measures (Gibbs and Su, 2002). This means that an alternative probability which satisfies the constraints of (3.1) and is close in KL-divergence to the stressed probability, is also close to the stressed probability in the Lévy metric.

Optimisation problem (3.1) under linear (i.e. moment) constraints was first studied in the seminal paper by Csiszár (1975). In the context of financial risk management, in particular when risk measures are used, optimisation problem (3.1) involves non-linear constraints and Csiszár's theory cannot be applied. Relevant research includes Cambou and Filipović (2017) who consider the optimisation problem for general $f$-divergences and probability set constraints. Weber (2007) works with bounded random variables

and considers risk measure constraints such as ES and shortfall risk measures, see Sections 3.3.3 and 3.3.4 for a more detailed comparison.

### 3.3.2 Probability constraints

Before studying problem (3.1) with constraints involving the risk measures of Section 3.2, we consider stresses under which the probabilities of (adverse) outcomes of $Y = g(\boldsymbol{X})$ are altered. These outcomes are captured by disjoint sets $B_1, \dots, B_I \subseteq \mathbb{R}$, each set $B_i$ associated with an event $\{Y \in B_i\}$ where the system being studied is failing or 'out of control'. In a financial context, where $Y$ is interpreted as a loss, one can identify $B_i$ with a region of extreme losses.

The following result is an immediate consequence of Theorem 3.1 in Csiszár (1975); we also refer to Cambou and Filipović (2017).

**Proposition 3.3.1.** Let $B_1, \dots, B_I \subseteq \mathbb{R}$ be disjoint Borel sets with $P(Y \in B_i) > 0$, $i = 1, \dots, I$, and $\alpha_1, \dots, \alpha_I > 0$ such that $\alpha_1 + \cdots + \alpha_I \leq 1$. Then there exists a unique solution to

$$\min_{Q \in \mathcal{P}} D_{\mathrm{KL}}(Q \| P), \qquad \text{s.t. } Q(Y \in B_i) = \alpha_i,\ i = 1, \dots, I, \tag{3.2}$$

with RN-density given by $\zeta = \sum_{i=0}^{I} \frac{\alpha_i}{P(Y \in B_i)} \mathbb{1}_{\{Y \in B_i\}}$, where we write $\alpha_0 = 1 - \sum_{i=1}^{I} \alpha_i$ and $B_0 = (\bigcup_{i=1}^{I} B_i)^c$.

The RN-density $\zeta$ in Proposition 3.3.1 is a piecewise constant function of $Y$. This implies that all outcomes of $Y$ within a set $B_i$ receive the same probability re-weighting by the change to the stressed probability. In particular, if $\alpha_i > P(Y \in B_i)$, under the alternative probability $Q$ the probability of all outcomes in $B_i$ increases. Note that moving from the baseline to the stressed model might induce a new dependence structure in the input factors.

### 3.3.3 VaR constraints

We now consider optimisation problem (3.1) under a constraint on the risk measure VaR, applied to the output $Y$. A VaR constraint is not equivalent to a probability constraint of optimisation problem (3.2), when $F_Y$ is not strictly increasing.

**Proposition 3.3.2.** Let $0 < \alpha < 1$ and $q \in \mathbb{R}$ such that $\mathrm{VaR}_\alpha(Y) < q < \operatorname{ess\,sup} Y$ and consider the optimisation problem

$$\min_{Q \ll \mathcal{P}} D_{\mathrm{KL}}(Q\|P), \quad \text{s.t. } \mathrm{VaR}_\alpha^Q(Y) = q. \tag{3.3}$$

There exists a unique solution to (3.3) if and only if $P(q-\varepsilon < Y < q) > 0$ for all $\varepsilon > 0$. The RN-density of the solution is given by

$$\zeta = \frac{\alpha}{P(Y < q)} \mathbb{1}_{\{Y<q\}} + \frac{1-\alpha}{P(Y \geq q)} \mathbb{1}_{\{Y \geq q\}}.$$

The assumption $P(q-\varepsilon < Y < q) > 0$ for all $\varepsilon > 0$, implies that $q$ cannot be chosen arbitrarily. In particular, problem (3.3) does not have a solution if the distribution of $Y$ is constant to the left of $q$ ($q$ excluded); this includes the (uncommon in practice) case where $Y$ is a discrete random variable. This complication arises from using the constraint $\mathrm{VaR}_\alpha^Q(Y) = q$ rather than $Q(Y \leq q) = \alpha$. If $q$ cannot be chosen to fulfil the assumptions in Proposition 3.3.2, the form of $\zeta$ in Proposition 3.3.2 remains meaningful: by Proposition 3.3.1, it is the solution to an optimisation problem where the constraint $\mathrm{VaR}_\alpha^Q(Y) = q$ is replaced by $Q(Y < q) = \alpha$.

The RN-density $\zeta$ of the solution to (3.3) is a non-decreasing function of $Y$ since $\alpha \leq P(Y \leq \mathrm{VaR}_\alpha(Y)) \leq P(Y < q)$. Hence, under the stressed probability, adverse realisations of the output are given higher probabilities of occurrence.

**Remark 3.3.3.** Propositions 3.3.1 and 3.3.2 hold true for any $f$-divergence with a strictly convex function $f$. In particular, the RN-densities $\zeta$ of the solutions of (3.2) and (3.3) are independent of the choice of $f$-divergence. We do not provide a proof for this statement, however the steps of the proofs of Propositions 3.3.1 and 3.3.2 can be closely retraced if one substitutes the KL-divergence with a general $f$-divergence. We refer to Ben-Tal et al. (2013) for robust linear optimisation with general $f$-divergence constraints.

**Remark 3.3.4.** Let $\mathrm{VaR}_\alpha(Y) < q^* < \operatorname{ess\,sup} Y$ be a stress for which the distribution function of $Y$ under $P$ is increasing and continuous in a neighbourhood of $q^*$, so that a solution of problem (3.3) exists for all $q$ in that neighbourhood. Then, viewed as a function of $q$, the RN-density $\zeta(q)$ is a.s. continuous under $P$. Thus, the corresponding probability measure $Q^{\zeta(q)}$, solution of problem (3.3), converges in total variation distance to $Q^{\zeta(q^*)}$, implying that stressed models are robust with respect to stresses in VaR.

The explicit form of the RN-density in Proposition 3.3.2 (as well as the subsequent Propositions 3.3.5-3.3.7), allows easy implementation of the change of measure in a Monte Carlo simulation context similar to Beckman and McKay (1987). Note that the RN-density is a function of $Y$, in the sense that $\zeta(\omega) = \eta(Y(\omega))$, $\omega \in \Omega$, for a function $\eta$. Then, one can follow the process:

1. Sample $M$ multivariate scenarios $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(M)}$ from $\boldsymbol{X}$ under $P$. Calculate $y^{(k)} = g(\boldsymbol{x}^{(k)})$, $k = 1, \ldots, M$.

2. Set $\zeta^{(k)} = \eta(y^{(k)})$, $k = 1, \ldots, M$.

3. The distributions of the output and inputs under the stressed measure $Q$ are estimated by:

$$F_Y^Q(y) = \frac{1}{M}\sum_{k=1}^{M} \zeta^{(k)} \mathbb{1}_{\{y^{(k)} \leq y\}}, \; y \in \mathbb{R},$$
$$F_{X_i}^Q(x) = \frac{1}{M}\sum_{k=1}^{M} \zeta^{(k)} \mathbb{1}_{\{x_i^{(k)} \leq x\}}, \; x \in \mathbb{R}, \; i = 1, \ldots, n.$$

Thus, the process of working out the distribution of input factors under the stressed measure is akin to importance sampling, with $\zeta^{(k)}$ playing the role of importance weights. Note that this calculation allows stressing the model without the need to re-simulate scenarios under $Q$, which can be of practical importance if evaluation of $g$ is computationally expensive. Straightforward implementation yields a computational cost of $M(n+1)$ for calculating the empirical distribution functions under the stressed model for the output and all input factors. In a simulation environment, convergence can be improved if Quasi Monte Carlo sampling is deployed, which is not a route we pursue here. Note that when simulations are computationally very expensive, meta-modelling techniques are often used in practice.

**Example.** The following insurance portfolio, similar to Example 1 in Tsanakas and Millossovich (2016), will be used as an illustrative example throughout the chapter. An insurance company faces a loss $L$ resulting from two lines of business. The two lines produce losses $X_1, X_2$ respectively, which are subject to the same multiplicative inflation factor $X_3$, such that $L = X_3(X_1 + X_2)$. The insurance company has a reinsurance contract on the loss $L$ with limit $l$ and deductible $d$. The total portfolio loss for the insurance company is

$$Y = L - (1 - X_4)\min\{(L-d)_+, l\},$$

where $X_4$ captures the percentage lost due to a default of the reinsurance company.

In this example, the two lines of business $X_1, X_2$ are truncated Log-Normal$(4.98, 0.23^2)$ and Gamma(100, 2) distributed, with respective means 150, 200 and standard deviations 35, 20. The truncation point for $X_1$ is chosen to be the 99.9% quantile. The multiplicative factor $X_3$ follows a truncated Log-Normal$(0.05, 0.02^2)$ distribution with mean 1.05, standard deviation 0.02 and truncation point equal to the 99.9% quantile. The default loss $X_4$ is modelled through a Beta(0.125, 1.125) distribution, corresponding to mean 0.1 and standard deviation 0.2. To complete the specification of the joint distribution of $\boldsymbol{X}$, we further assume that $X_1, X_2, X_3$ are independent and $X_4$ is independent of $(X_1, X_2, X_3)$ given $L$. Additionally, $X_4$ is taken to be dependent on the aggregated loss $L$ through a Gaussian copula with correlation 0.6. The deductible of the insurance contract is $d = 380$ and the limit $l = 30$.

Consider optimisation problem (3.3) with a 10% increase in $\mathrm{VaR}_{0.9}$, that is

$$\min_{Q \ll \mathcal{P}} D_{\mathrm{KL}}(Q \| P), \quad \text{s.t. } \mathrm{VaR}_{0.9}^{Q}(Y) = 1.1\, \mathrm{VaR}_{0.9}(Y). \tag{3.4}$$

The solution to the problem (3.4) is estimated from a Monte Carlo sample containing $M = 100{,}000$ simulated scenarios from $(\boldsymbol{X}, Y)$. Simulated values of the RN-density $\zeta$ are plotted in the left of Figure 3.1, against samples from $Y$. It is seen that the RN-density is a non-decreasing function of $Y$ and thus gives more weight to adverse outcomes of $Y$.

The empirical distribution functions of the total loss $Y$ of the insurance company under the baseline probability (dashed black) and the stressed probability (solid grey) are displayed in the right of Figure 3.1. The output distribution under the stressed probability lies beneath, and therefore first-order stochastically dominates, the distribution of $Y$ under the baseline probability. We refer to Section 3.A for a more detailed discussion of stochastic comparisons of stressed and baseline probabilities.

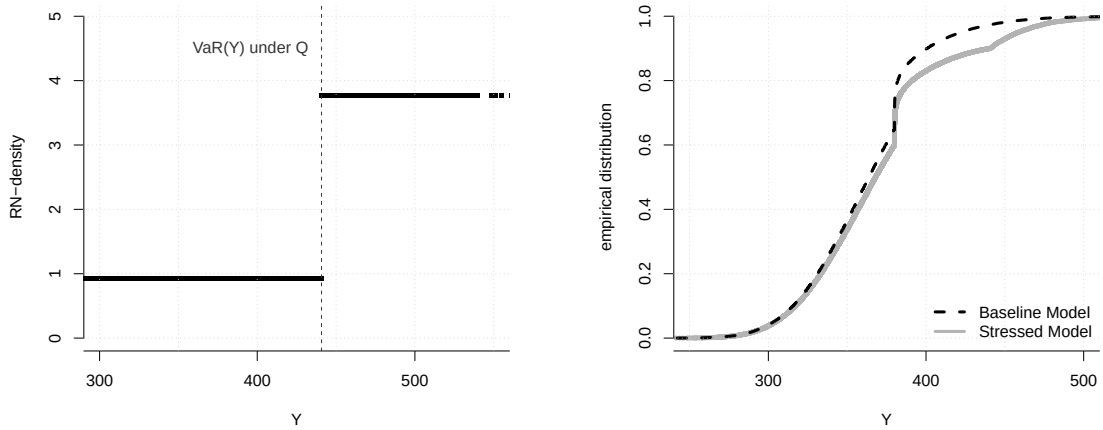


Figure 3.1: Left: simulated RN-density of the solution to (3.4). Right: simulated empirical distribution functions of the output under the baseline (dashed black) and the stressed (solid grey) model under problem (3.4).

Figure 3.2 displays the change in distribution of the input factors when moving from the baseline model to the stressed model. It can be seen that all factors under the stressed probability first-order stochastically dominate the corresponding inputs under the baseline probability. However, not all input factors are impacted the same: the distributions of inputs $X_1$ and $X_4$ are stressed more compared to the baseline model. This indicates a higher sensitivity to $X_1$ and $X_4$, compared to $X_2$ and $X_3$. A specific sensitivity measure reflecting the above observations is introduced in Section 3.4.

Table 3.1 summarises basic characteristics of the change in the output and the input factors under the two models. Consistently with Figure 3.2, it is seen that $X_1$ and $X_4$ are the most affected input factors by the change of probability measure. For example, under the stressed probability, $X_1$, $X_4$ are subject to a relative increase of the standard deviation of 17%, 20%, respectively.

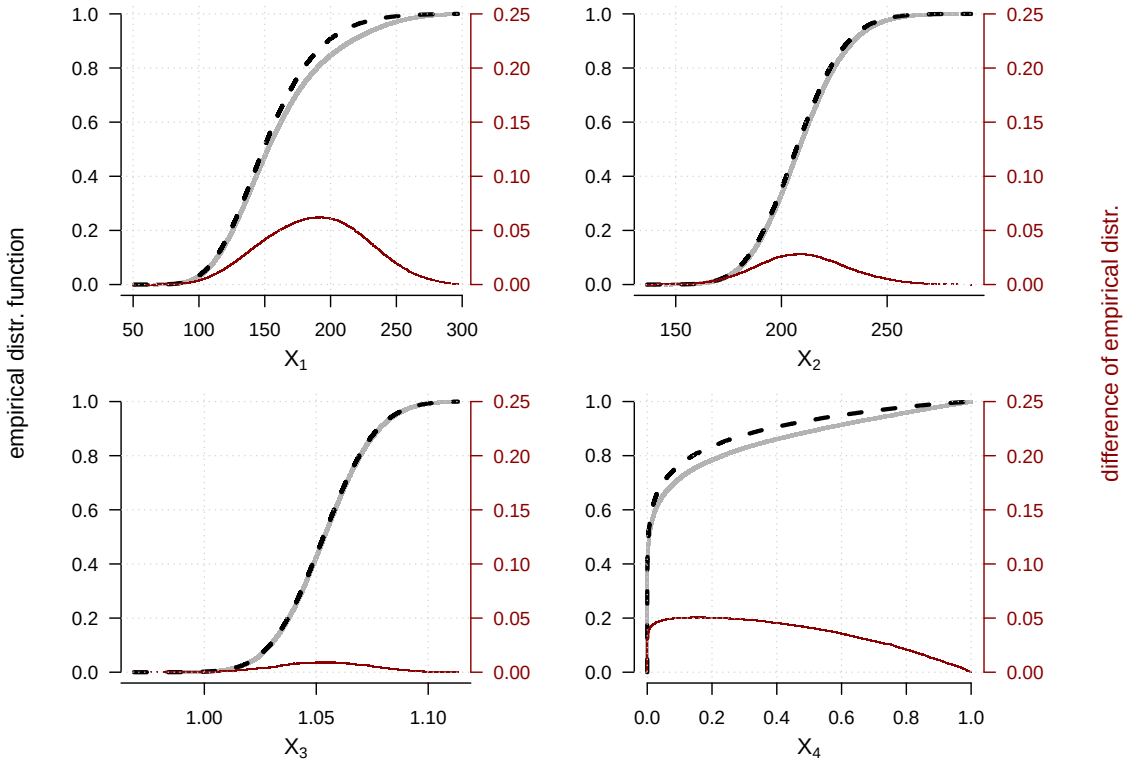


Figure 3.2: Empirical distribution functions of the input factors under the baseline (dashed black) and the stressed model (solid grey) under problem (3.4). The dark red line displays the difference of the distribution functions according to the axis on the right.

### 3.3.4 VaR and ES constraints

This section addresses optimisation problem (3.1) with a constraint on both, VaR and ES. Adding to problem (3.3) a constraint on ES allows to stress the whole tail of the output distribution. Weber (2007) considers optimisation problem (3.1) with an ES constraint only. In that case there does not exist an analytic solution of the stressed probability and Weber (2007) offers a procedure for a numerical solution.

**Proposition 3.3.5.** Let $0 < \alpha < 1$ and $q, s \in \mathbb{R}$ such that $\mathrm{VaR}_\alpha(Y) < q < s < \operatorname{ess\,sup} Y$. Assume the cumulant generating function of $Y|Y > q$ under $P$ exists in a

Table 3.1: Distributional characteristics of inputs and output under the baseline and stressed model under problem (3.4).

| **Sensitivity** | **Input factors** | | | | **Output** |
|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Y$ |
| Mean under $P$ | 150 | 200 | 1.05 | 0.10 | 362 |
| Mean under $Q$ | 156 | 201 | 1.05 | 0.14 | 369 |
| Standard deviation under $P$ | 35 | 20 | 0.02 | 0.20 | 36 |
| Standard deviation under $Q$ | 41 | 21 | 0.02 | 0.24 | 45 |
| Skewness under $P$ | 0.6 | 0.2 | 0.0 | 2.5 | 0.4 |
| Skewness under $Q$ | 1.2 | 0.5 | 0.1 | 2.4 | 1.2 |
| Excess kurtosis under $P$ | 0.5 | 0.1 | -0.1 | 5.6 | 1.3 |
| Excess kurtosis under $Q$ | 0.8 | 0.2 | -0.1 | 3.9 | 1.4 |

neighbourhood of 0 and that $E(Y|Y > q) < s$. Consider the optimisation problem

$$\min_{Q\in\mathcal{P}} D_{\mathrm{KL}}(Q\|P), \qquad \text{s.t. } \mathrm{VaR}_\alpha^Q(Y) = q,\ \mathrm{ES}_\alpha^Q(Y) = s. \tag{3.5}$$

Define the sets $A_1 = \{Y \geq q\}$ and $A_2 = \{Y > q\}$ and, for $i = 1, 2$, denote by $\theta_i^*$ the unique positive solution of the equation

$$E\left((Y - s)e^{\theta(Y-q)} \,\middle|\, A_i\right) = 0. \tag{3.6}$$

There exists a unique solution to problem (3.5) under either

1. $P(q - \varepsilon < Y < q) > 0$ for all $\varepsilon > 0$ and $E\left(e^{\theta_1^*(Y-q)} \,\middle|\, A_1\right) \leq \frac{P(A_1^c)/P(A_1)}{\alpha/(1-\alpha)}$.

2. $P(Y = q) > 0$ and $P(q - \varepsilon < Y < q) = 0$ for some $\varepsilon > 0$, and $E\left(e^{\theta_2^*(Y-q)} \,\middle|\, A_2\right) \geq \frac{P(A_2^c)/P(A_2)}{\alpha/(1-\alpha)}$.

The corresponding RN-density of the solution is

$$\zeta_i = \frac{\alpha}{P(A_i^c)} \mathbb{1}_{A_i^c} + \frac{1-\alpha}{E\left(e^{\theta_i^*(Y-q)} \mathbb{1}_{A_i}\right)} e^{\theta_i^*(Y-q)} \mathbb{1}_{A_i},\ i = 1, 2.$$

Note that, compared to stressing solely the VaR, adding an ES constraint may provide a solution even for an output following a discrete distribution. The condition on the moment generating function in cases 1 and 2 restricts the choice of $s$ and $q$, such that the stressed risk measure values cannot be chosen independently.

The RN-density of Proposition 3.3.5 under case 1, $\zeta_1$, is a non-decreasing function of $Y$. Under Proposition 3.3.5 case 2, the RN-density $\zeta_2$ is not monotone. However, both RN-densities are exponentially increasing for realisations of $Y$ exceeding $q$. Thus, under the stressed model, adverse outcomes of $Y$, such as tail events, admit a higher likelihood compared to the baseline model.

**Remark 3.3.6.** Let $\mathrm{VaR}_\alpha(Y) < q^* < s^* < \operatorname{ess\,sup} Y$ be a stress of VaR and ES for which the cumulant generating function of $Y \,|\, Y > q^*$ under $P$ exists in a neighbourhood of 0, $E(Y \,|\, Y > q^*) < s^*$, the distribution function of $Y$ under $P$ is increasing and continuous in a neighbourhood of $q^*$ and the second inequality in case 1 of Proposition 3.3.5 holds strictly. Then, a solution of problem (3.6) exists in a neighbourhood of $(q^*, s^*)$. Viewed as a function of $(q, s)$, the RN-density $\zeta(q, s)$ is a.s. continuous under $P$. Thus, the corresponding probability measure $Q^{\zeta(q,s)}$, solution of problem (3.5), converges in total variation distance to $Q^{\zeta(q^*,s^*)}$, implying that stressed models are robust with respect to stresses in VaR and ES.

**Example (continued).** We consider optimisation problem (3.5) with a 10% increase in $\mathrm{VaR}_{0.9}$ and a 13% increase in $\mathrm{ES}_{0.9}$. Figure 3.3 displays samples of the RN-density of the stressed probability measure, see Proposition 3.3.5 case 1 For high outcomes of the output $Y$, the RN-density $\zeta$ is exponentially increasing as a function of $Y$, hence

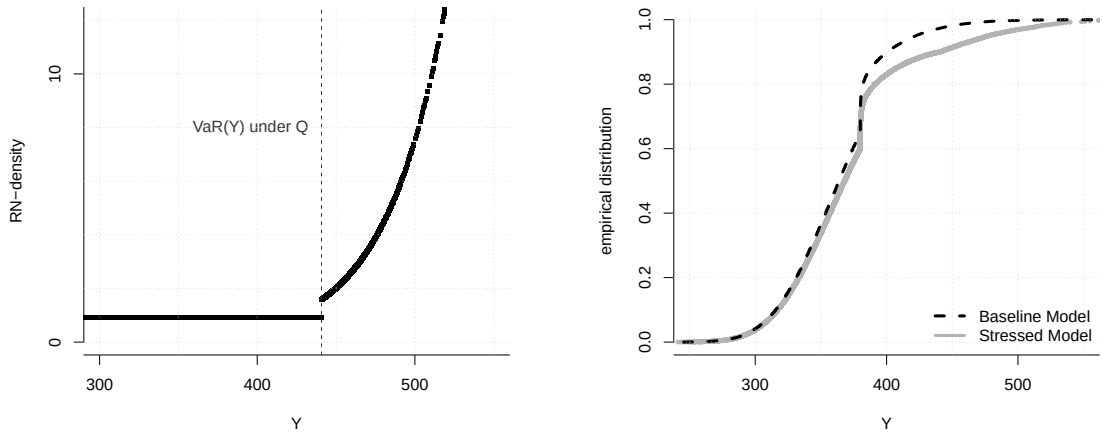


Figure 3.3: Left: simulated RN-density of the solution. Right: simulated empirical distribution functions of the output under the baseline (dashed black) and the stressed (solid grey) model under problem (3.5) with a 10% increase in $\mathrm{VaR}_{0.9}$ and a 13% increase in $\mathrm{ES}_{0.9}$.

inflates stressed tail probabilities. On the right hand side, the empirical distribution functions of the output under the baseline (dashed black) and the stressed model (solid grey) are shown.

Observe that the stressed distribution of the output appears similar to the stressed distribution of optimisation problem (3.4), see Figure 3.1. This is due to the fact that increasing $\mathrm{VaR}_{0.9}$ by 10% in optimisation problem (3.4), already leads to an increase of 8.5% in $\mathrm{ES}_{0.9}$ under the stressed model. However, comparing Tables 3.1 and 3.2 it is seen that the standard deviation, skewness and kurtosis of $Y$ increase more when stressing VaR and ES, compared to stressing VaR alone.

Similar to optimisation problem (3.4), the output and the input factors under the baseline probability are first-order stochastically dominated by the stressed probability, as can be seen in Figure 3.3 and 3.4. We refer to Section 3.A for a formal treatment of stochastic comparison of the stressed and baseline probabilities.

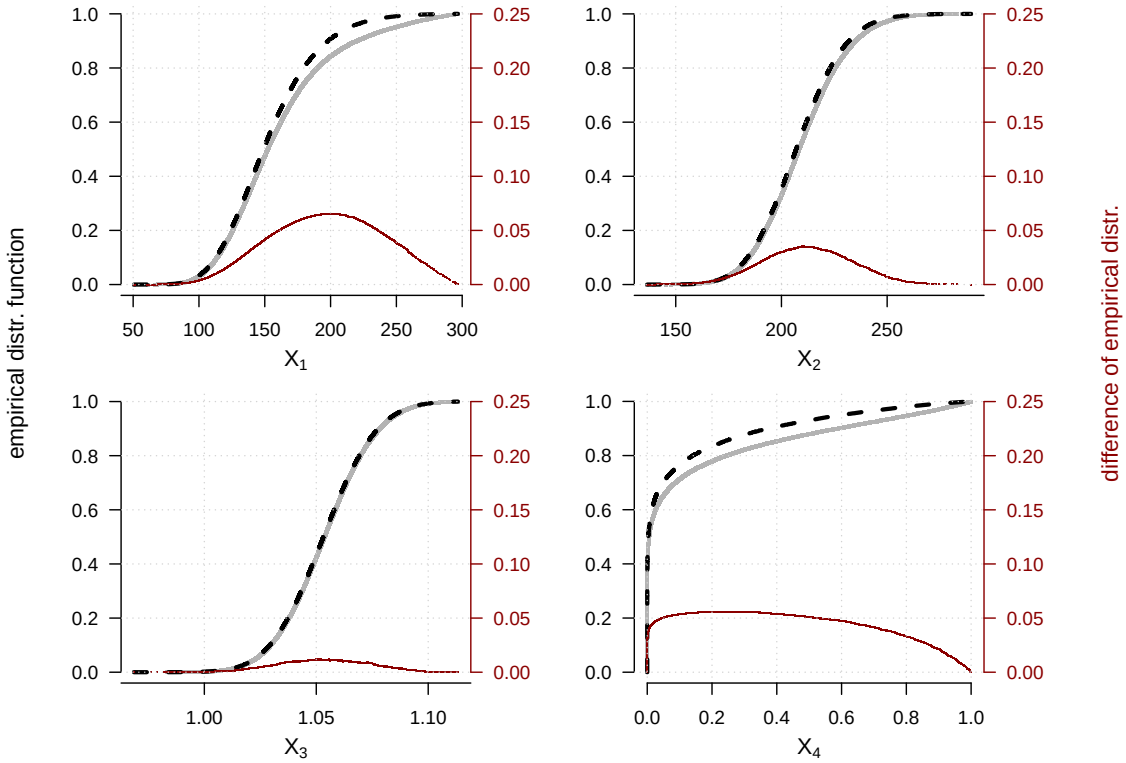


Figure 3.4: Empirical distribution functions of the input factors under the baseline (dashed black) and the stressed model (solid grey) under problem (3.5) with a 10% increase in $\mathrm{VaR}_{0.9}$ and a 13% increase in $\mathrm{ES}_{0.9}$. The dark red line displays the difference of the distribution functions according to the axis on the right.

### 3.3.5 Shortfall risk measure constraints

Optimisation problem (3.1) with shortfall risk measure constraints is studied in Weber (2007) and is a direct application of Theorem 3.1 in Csiszár (1975). Nonetheless, we present the solution for completeness.

**Proposition 3.3.7.** Let $\rho$ be a shortfall risk measure with loss function $\ell$ and $y_0$, and $q \in \mathbb{R}$ in the support of $Y$ such that $E(\ell(Y-q)) < y_0$. If the moment generating

Table 3.2: Distributional characteristics under the baseline and the stressed model under problem (3.5) with a 10% increase in $\mathrm{VaR}_{0.9}$ and a 13% increase in $\mathrm{ES}_{0.9}$.

| **Sensitivity** | **Input factors** | | | | **Output** |
|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Y$ |
| Mean under $P$ | 150 | 200 | 1.05 | 0.10 | 362 |
| Mean under $Q$ | 157 | 202 | 1.05 | 0.14 | 371 |
| Standard deviation under $P$ | 35 | 20 | 0.02 | 0.20 | 36 |
| Standard deviation under $Q$ | 43 | 21 | 0.02 | 0.26 | 50 |
| Skewness under $P$ | 0.6 | 0.2 | 0.0 | 2.5 | 0.4 |
| Skewness under $Q$ | 1.4 | 0.5 | 0.1 | 2.4 | 1.7 |
| Excess kurtosis under $P$ | 0.5 | 0.1 | -0.1 | 5.6 | 1.3 |
| Excess kurtosis under $Q$ | 1.3 | 0.2 | -0.1 | 3.7 | 2.8 |

function of $\ell(Y-q)$ exists in a neighbourhood of 0, then the optimisation problem

$$\min_{Q\in\mathcal{P}} D_{\mathrm{KL}}(Q\|P), \qquad \text{s.t. } \rho^Q(Y)=q, \tag{3.7}$$

has a unique solution whose density is given by $\zeta = \frac{1}{E\left(e^{\theta^*\ell(Y-q)}\right)} e^{\theta^*\ell(Y-q)}$, where $\theta^*$ is the unique positive solution of $E\left(\left(\ell(Y-q)-y_0\right)e^{\theta\ell(Y-q)}\right)=0$.

### 3.3.6 Multivariate output

Problem (3.1) can be extended to constraints on a multivariate output, that is, to $\boldsymbol{Y}=(Y_1,\dots,Y_k)=g(\boldsymbol{X})$, for $k\in\mathbb{N}$ and aggregation function $g\colon\mathbb{R}^n\to\mathbb{R}^k$. The general case under set constraints can be treated along the lines of Cambou and Filipović (2017), Section 7. We provide below an example based on two outputs.

**Example (continued).** We revisit the insurance portfolio example of Section 3.3.3 and view as output both the loss before reinsurance, $L$, and after reinsurance, $Y$. To shorten notation let $v_L=\mathrm{VaR}_{0.9}(L)$ and $v_Y=\mathrm{VaR}_{0.9}(Y)$ and consider the problem of

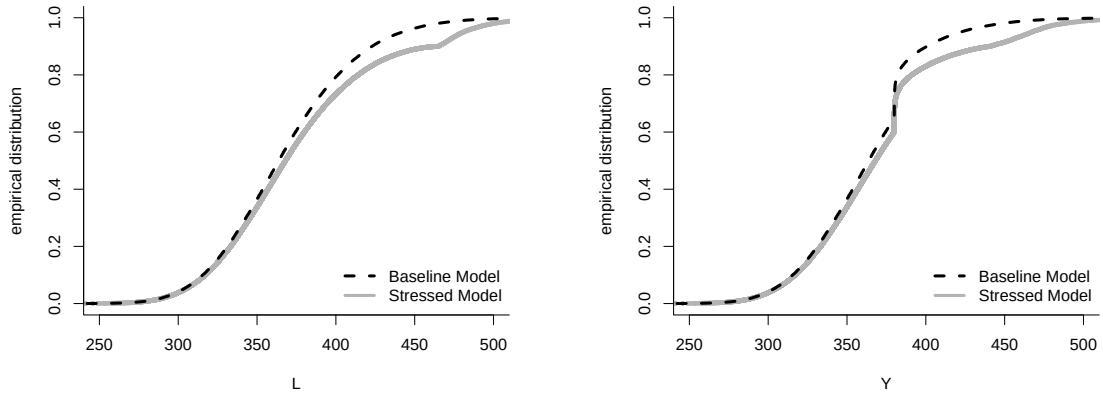


Figure 3.5: The simulated empirical distribution of the output under the baseline model (dashed black) and under the stressed model (solid grey) under problem (3.8). The left graph depicts the output before reinsurance, $L$, whereas the right plot shows the output after reinsurance, $Y$.

stressing the VaR of $L$ and $Y$ by 10%

$$\min_{Q\in\mathcal{P}} D_{\mathrm{KL}}(Q\|P), \qquad \text{s.t. } \mathrm{VaR}^Q_{0.9}(L) = 1.1\, v_L,\ \mathrm{VaR}^Q_{0.9}(Y) = 1.1\, v_Y. \tag{3.8}$$

The constraints can be written as $Q(L \leq v_L) = 0.9$ and $Q(Y \leq v_Y) = 0.9$, since the distribution functions of $L$ and $Y$ are increasing and continuous around those stressed VaR. Thus, problem (3.8) can be solved straightforwardly using the Lagrange multiplier technique. The RN-density of the solution of (3.8) is constant on the four sets $\{L \leq v_L,\, Y \leq v_Y\}$, $\{L > v_L,\, Y \leq v_Y\}$, $\{L \leq v_L,\, Y > v_Y\}$ and $\{L > v_L,\, Y > v_Y\}$.

Figure 3.5 displays the simulated empirical distribution of $L$ and $Y$ under the baseline model and the stressed model. Note that stressing both $L$ and $Y$, in contrast to stressing only $Y$, do not lead to radical different results, as can be seen comparing Tables 3.1 and 3.3. This is because the stressed model, solution of problem (3.4), already induces an increase of 7% in $\mathrm{VaR}_{0.9}(L)$. Moreover, stressing $L$, the portfolio

Table 3.3: Distributional characteristics under the baseline and the stressed model under problem (3.8).

| **Sensitivity** | **Input factors** | | | | **Output** | |
|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $L$ | $Y$ |
| Mean under $P$ | 150 | 200 | 1.05 | 0.10 | 367 | 362 |
| Mean under $Q$ | 157 | 202 | 1.05 | 0.13 | 376 | 370 |
| Standard deviation under $P$ | 35 | 20 | 0.02 | 0.2 | 42 | 36 |
| Standard deviation under $Q$ | 42 | 21 | 0.02 | 0.23 | 52 | 46 |
| Skewness under $P$ | 0.6 | 0.2 | 0.0 | 2.5 | 0.4 | 0.4 |
| Skewness under $Q$ | 1.3 | 0.5 | 0.1 | 2.4 | 1.2 | 1.3 |
| Excess kurtosis under $P$ | 0.5 | 0.1 | -0.1 | 5.6 | 0.3 | 1.3 |
| Excess kurtosis under $Q$ | 0.9 | 0.2 | -0.1 | 4.3 | 0.7 | 1.6 |

loss before reinsurance, in addition to $Y$, reduces the importance of the default of the reinsurance displayed in the change of $X_4$ under $Q$.

## 3.4 Sensitivity measures for importance ranking

### 3.4.1 Definition of sensitivity measures

Plots such as the ones shown in Figures 3.2 and 3.4 provide some insight into the sensitivity of the output risk measure to different input factors. In order to produce a ranking of inputs, it is necessary to introduce a formal sensitivity or importance measure; this is especially the case for models with large numbers of inputs for which succinct sensitivity summaries are needed. Here we develop a sensitivity measure that quantifies changes in input factors under the stressed model, compared to the baseline model.

Before proceeding to the definitions, some preliminaries are due. The random couple $(V, W)$ is comonotonic if it can be written as $(V, W) \stackrel{d}{=} (F_V^{-1}(U), F_W^{-1}(U))$, for a uni-

formly distributed random variable $U$ on $(0,1)$. In contrast, $(V,W)$ is counter-monotonic if $(V,W) \stackrel{d}{=} (F_V^{-1}(U), F_W^{-1}(1-U))$. Comonotonicity and counter-monotonicity correspond to extremal positive and negative dependence structures respectively, for a random couple with fixed marginals (Müller and Stoyan, 2002). For a random variable $V$, we denote by $V_{|W}, V_{|W^\dagger}$ the random variables satisfying $V_{|W} \stackrel{d}{=} V_{|W^\dagger} \stackrel{d}{=} V$, such that $(V_{|W}, W)$ is comonotonic and $(V_{|W^\dagger}, W)$ is counter-monotonic. Then for any $V' \stackrel{d}{=} V$ it holds that (Rüschendorf, 1983),

$$E(WV_{|W^\dagger}) \leq E(WV') \leq E(WV_{|W}).$$

The subsequent definition introduces a sensitivity measure that captures the extent to which a random variable is affected by a stress on the baseline model, that is, a change in probability measure.

**Definition 3.4.1.** Let $Q^\xi$ be an alternative probability with RN-density $\xi = \frac{\mathrm{d}Q^\xi}{\mathrm{d}P}$. The sensitivity of a random variable $Z$ to the change of measure is given by

$$\mathcal{S}(Z,\xi) = \begin{cases} \dfrac{E(Z\xi) - E(Z)}{\max_{\psi \stackrel{d}{=} \xi} E(Z\psi) - E(Z)} & E(Z\xi) \geq E(Z), \\ -\dfrac{E(Z\xi) - E(Z)}{\min_{\psi \stackrel{d}{=} \xi} E(Z\psi) - E(Z)} & \text{otherwise}, \end{cases}$$

where we use the convention $\pm\frac{\infty}{\infty} = \pm 1$ and $\frac{0}{0} = 0$.

In the definition of $\mathcal{S}(Z,\xi)$, the numerator $E(Z\xi) - E(Z)$ reflects the increase in the expectation of $Z$ under the alternative model. The denominator normalises this difference, as it represents the maximal (or minimal) increase of the expectation of $Z$, under all alternative models with density $\psi$ that are equal in distribution to $\xi$. This ensures normalisation of the sensitivity measure to $[-1,1]$. If $\mathcal{S}(Z,\xi) = 1$ or $\mathcal{S}(Z,\xi) = -1$, the alternative model produces a maximal stress on the variable $Z$,

representing a positive or negative impact of the changes in probability measure on $Z$ respectively.

Note that $\arg\max_{\psi \overset{d}{=} \xi} E(Z\psi) = \xi_{|Z}$ and $\arg\min_{\psi \overset{d}{=} \xi} E(Z\psi) = \xi_{|Z^\dagger}$. This allows for a straightforward calculation of the sensitivity measure. If working within a Monte Carlo simulation context, as is common in risk analysis, $\xi_{|Z}$, resp. $\xi_{|Z^\dagger}$, can be simply obtained by re-arranging samples of $\xi$ to be sorted in the same, resp. opposite, way as samples from $Z$. This context gives a different perspective on the constraint $\psi \overset{d}{=} \xi$: if simulated elements of $\xi$ represent a particular scheme for re-weighting simulated scenarios, then $\psi$ are vectors containing the same weights as $\xi$, but re-arranged to potentially prioritise different scenarios.

Next we define two sensitivity measures that are specific to the reverse sensitivity analysis framework of this chapter.

**Definition 3.4.2.** Let $Q^\zeta$ be an alternative model with density $\zeta = \frac{\mathrm{d}Q^\zeta}{\mathrm{d}P} = \eta(Y)$, for a non-decreasing function $\eta$. For input $X_i$ and output $Y$, we define the *reverse* and *forward sensitivity measures* $\Gamma_i$ and $\Delta_i$ by:

$$\Gamma_i = \mathcal{S}(X_i, \zeta),$$
$$\Delta_i = \mathcal{S}(Y, \zeta_{|X_i}).$$

Here, $\zeta = \eta(Y)$ can be arrived at as the solution of optimisation problems (3.3), (3.5) or (3.7). $\Gamma_i$ thus reflects the extent to which the reverse sensitivity test affects the expectation of the input factor $X_i$. Note that for $E(X_i\zeta) \geq E(X_i)$, we can write $\Gamma_i = \frac{\mathrm{Cov}(X_i,\zeta)}{\max_{\psi \overset{d}{=} \zeta} \mathrm{Cov}(X_i,\psi)}$, showing that the reverse sensitivity measure can also be understood as a dependence measure between $X_i$ and $Y$. In this sense it is closely related to the dependence measure introduced by Kachapova and Kachapov (2012). Indeed, sensitivity measures considering the dependence between $X_i$ and $Y$ have a

rich history in sensitivity analysis, for an overview see for example Borgonovo et al. (2016). Thus, in contrast to variance-based sensitivity measures, $\Gamma_i$ and $\Delta_i$ can take both negative and positive values, indicating the direction in which input factors affect the output.

A possible criticism of the measure $\Gamma_i$ and, by extension, the reverse sensitivity testing framework we propose, is as follows. Let $\Gamma_i$ be high. This implies that stressing the model output $Y$ leads to a substantial change in the distribution of the input factor $X_i$. However, this is not equivalent to a perturbation in the distribution of $X_i$ leading to a sizeable stress in the distribution of the output $Y$. Such a discrepancy, though uncommon, is theoretically possible and has been termed *probabilistic dissonance* (Cooke and van Noortwijk, 1999).

This motivates the introduction of the forward sensitivity measure $\Delta_i$, as a companion measure to $\Gamma_i$. The definition of the forward sensitivity measure $\Delta_i$ is analogous to that of $\Gamma_i$, but with a focus on the change in the expectation of $Y$ when perturbing the distribution of the model input $X_i$. Recall that $\zeta_{|X_i} = \arg\max_{\psi \overset{d}{=} \zeta} E(\psi X_i)$. Therefore, $\zeta_{|X_i}$ is a RN-density with the same distribution as $\zeta$ that has the most adverse effect on the input factor $X_i$. Thus $\Delta_i$ captures the impact of a change in the input $X_i$ on the output $Y$. Reporting $\Delta_i$ along with $\Gamma_i$ can thus produce warning signs of probabilistic dissonance.

Properties of the sensitivity measures $\Gamma_i$ and $\Delta_i$, reflecting their nature as dependence measures, are summarised below.

**Proposition 3.4.3.** Using the above introduced notation, the sensitivity measures $\Gamma_i$ and $\Delta_i$ are well-defined and have the following properties:

1. $-1 \leq \Gamma_i, \Delta_i \leq 1$.

2. $\Gamma_i = \Delta_i = 0$, if $X_i, Y$ are independent.

3. $\Gamma_i = \Delta_i = 1$, if $(X_i, Y)$ is comonotonic.

4. $\Gamma_i = \Delta_i = -1$, if $(X_i, Y)$ is counter-monotonic.

5. $\Gamma_i = \Delta_i \geq 0$, if $(X_i, Y)$ are positively quadrant dependent[1].

6. $\Gamma_i = \Delta_i \leq 0$, if $(X_i, Y)$ are negatively quadrant dependent[1].

**Remark 3.4.4.** Let $\zeta$ be the RN-density of the solution of problem (3.3) or (3.5). Then, for an input $X_i$ with continuous distribution function, the corresponding reverse and forward sensitivity measures, $\Gamma_i, \Delta_i$, are robust in the size of the stress as long as the RN-density is a.s. continuous in a neighbourhood of that stress. We refer to the remarks on robustness in Section 3.3.3 and 3.3.4 for details on the conditions required on the RN-density.

The above defined sensitivity measures focus on the difference of expectations under an alternative and the baseline model. If the interest lies in other distributional properties, such as tails, Definition 3.4.2 can be extended to consider monotone transformations of input factors. Specifically, one can calculate $\mathcal{S}(u(X_i), \zeta)$, respectively $\mathcal{S}(u(Y), \zeta_{|X_i})$, for an appropriately chosen non-decreasing function $u$. As the couple $(u(X_i), X_i)$ is comonotonic, the interpretation of the sensitivity measures remains unchanged. One particular example is the choice

$$u_v(X_i) = (X_i - F_{X_i}^{-1}(v))_+ - (F_{X_i}^{-1}(1-v) - X_i)_+, \quad 0.5 \leq v < 1. \tag{3.9}$$

For $v = 0.5$, the function $u_{0.5}$ is the identity and thus $\mathcal{S}(u_{0.5}(X_i), \zeta) = \Gamma_i$, respectively $\mathcal{S}(u_{0.5}(X_i), \zeta_{|X_i}) = \Delta_i$. When $v > 0.5$, the function $u_v$ is zero whenever $X_i \in \left[F_{X_i}^{-1}(1-v), F_{X_i}^{-1}(v)\right]$ and linearly increasing otherwise. Thus, increasing $v$ places higher emphasis on the tail behaviour of $X_i$. The random variable $u_v(Y)$ is defined and interpreted in a similar way.

[1]These concepts are reviewed in Section 3.A.

We denote $\Gamma_{i,v} = \mathcal{S}(u_v(X_i), \zeta)$ and $\Delta_{i,v} = \mathcal{S}(u_v(Y), \zeta_{|X_i})$. It is easily seen that the properties of Proposition 3.4.3 still apply to $\Gamma_{i,v}, \Delta_{i,v}$. In addition, it holds that

$$\mathcal{S}(u_v(aX_i + b), \zeta) = \text{sign}(a)\mathcal{S}(u_v(X_i), \zeta),$$

such that the reverse sensitivity measure is invariant under linear transformations of input factors.

**Example (continued).** Figure 3.6 displays the forward and reverse sensitivity measures $\Gamma_{i,v}, \Delta_{i,v}$ for $v \in [0.5, 0.999)$, for the stressed model arising from optimisation problem (3.5) with a 10% increase in VaR and a 13% increase in ES. Consistently with the example in Section 3.3.4, the highest sensitivity, for both reverse and forward measures, is displayed by $X_1$, followed by $X_4$, $X_2$ and $X_3$. Furthermore, the ranking is not affected by the level $v$ and is thus not sensitive to emphasising the tails of the distributions. In the next section we present a situation where this no longer holds true.

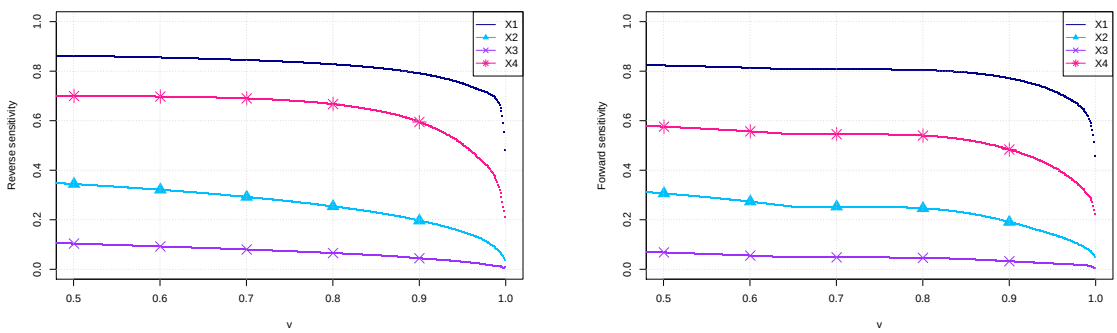


Figure 3.6: Reverse (left plot) and forward (right plot) sensitivity measures $\Gamma_{i,v}, \Delta_{i,v}$ with a 10% increase in VaR and 13% increase in ES.

To illustrate the impact on the sensitivity measure $\Gamma_i$ of the size of the stress applied to the output risk measure, we fix $\alpha = 0.9$ and let $q = \lambda \text{VaR}_\alpha(Y)$, with $\lambda$ ranging from

0.8 to 1.2.[2] Results are given in Table 3.4. For $\lambda < 1$ it is seen that the sensitivity measure $\Gamma_i$ takes negative values. This is a result of applying a negative stress on $Y$, such that the RN-density $\zeta$ becomes a decreasing function of $Y$. The absolute value of the sensitivity measure responds asymmetrically to positive ($\lambda > 1$) or negative ($\lambda < 1$) stresses on $\mathrm{VaR}_\alpha(Y)$, in extremis even leading to a change in the ranking of inputs. This reflects a different sensitivity to input factors with view to increasing or decreasing the VaR of the output.

Table 3.4: Reverse sensitivity measure $\Gamma_i$ for a 10% and 20% increase / decrease of $\mathrm{VaR}_{0.9}$.

| Input | $\lambda = 0.8$ | $\lambda = 0.9$ | $\lambda = 1.1$ | $\lambda = 1.2$ |
|---|---|---|---|---|
| $X_1$ | -0.83 | -0.85 | 0.88 | 0.90 |
| $X_2$ | -0.58 | -0.51 | 0.36 | 0.34 |
| $X_3$ | -0.17 | -0.17 | 0.15 | 0.14 |
| $X_4$ | -0.93 | -0.72 | 0.60 | 0.68 |

### 3.4.2 Comparison to other sensitivity measures

By the proposed reverse sensitivity measure $\Gamma_i$, we aim to quantify the extent to which a stress in the output distribution impacts different inputs. In the present section we compare $\Gamma_i$ to the moment independent sensitivity measure introduced in Borgonovo (2007), and to the variance-based sensitivity measures, see Borgonovo and Plischke (2016) for an overview, which are designed to apportion the output variance to individual input factors. Specifically, the *first order sensitivity index* $S_i$, the *total effects sensitivity index* $T_i$ and the *moment independent sensitivity measure* $\delta_i$ are

[2]Note that in general Proposition 3.3.2 only applies for $\lambda > 1$. However, continuity of $Y$ in this example implies that the RN-density $\zeta$ of Proposition 3.3.2 is a solution to problem (3.3) even for $\lambda < 1$.

respectively defined as

$$
\begin{aligned}
S_i &= \frac{\mathrm{Var}(E(Y|X_i))}{\mathrm{Var}(Y)} \\
T_i &= \frac{E(\mathrm{Var}(Y|\boldsymbol{X}_{-i}))}{\mathrm{Var}(Y)} \\
\delta_i &= \frac{1}{2} E\left( \int \left| f_Y(y) - f_{Y|X_i}(y|X_i) \right| \mathrm{d}y \right),
\end{aligned}
$$

where $\boldsymbol{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, $f_Y$ and $f_{Y|X_i}$ denote the density of the output $Y$ and the density of $Y$ conditional on the input factor $X_i$, respectively. The measure $S_i$ can be understood as the expected reduction in the variance of $Y$ that would be achieved if input $X_i$ could be fixed, whereas $T_i$ is interpreted as the expected variance that would be left if all inputs but $X_i$ could be fixed (Sobol, 1993; Wagner, 1995; Saltelli et al., 2008). The moment independent sensitivity $\delta_i$ can be seen as the expected shift of the output induced by fixing the input factor $X_i$ (Borgonovo, 2007).

Thus, the interpretation of the reverse sensitivity measure $\Gamma_i$ is quite different to that of $S_i$, $T_i$ and $\delta_i$, which are designed to answer different questions. Furthermore, $\Gamma_i$ is designed with reference to a (tail) risk measure like VaR / ES and hence captures distributional impacts differently than the variance-based or the moment independent sensitivities, as is illustrated by the following numerical example.

**Example (continued).** We return to the simple insurance portfolio example of Section 3.3.3 (optimisation problem (3.4)), stressing $\mathrm{VaR}_\alpha(Y)$ by 10%, for $\alpha = 0.5$ and $\alpha = 0.9$. The sensitivity indices $S_i, T_i$ and $\delta_i$ are calculated, in addition to $\Gamma_i$, where the calculation of the variance-based sensitivities is carried out via estimation of the necessary conditional expectations from the existing Monte Carlo sample by local polynomial regression and the estimation of $\delta_i$ utilises kernel smoothing as in Borgonovo et al. (2011).

In Table 3.5, the sensitivity measures are reported for a variation of the model, where the aggregation function $g$ and the marginal distributions of the inputs $X_i$ are unchanged, but the vector $\boldsymbol{X}$ is independent. Assuming that the input factors are independent facilitates comparison of $\Gamma_i$ and $\delta_i$ with the variance based sensitivity measures, $S_i$ and $T_i$. It can be seen that all sensitivity measures produce a consistent ranking, with $X_4$ assigned a very low sensitivity. Note that the first order sensitivities, $S_i$, sum up to 0.98, implying that the interaction terms are nearly as important as $S_3$.

The sensitivity measure $\Gamma_1$ is slightly increasing in $\alpha$, while $\Gamma_2$ is decreasing. This reflects the different tail characteristics of $X_1$ (LogNormal) and $X_2$ (Gamma); for a high $\alpha$, the focus is on the right tail of $Y$, for which the heavier tail of $X_1$ is more important.

Table 3.5: Comparison of the variance based sensitivity measures $S_i$, $T_i$, the moment independent sensitivity measure $\delta_i$ and the reverse sensitivity measure $\Gamma_i$ with a 10% increase in $\mathrm{VaR}_\alpha$, for $\alpha = 0.5$ and $\alpha = 0.9$, for independent input vector $\boldsymbol{X}$.

| **Input** | $S_i$ | $T_i$ | $\delta_i$ | $\Gamma_i$ | |
|---|---|---|---|---|---|
| | | | | $\alpha = 0.5$ | $\alpha = 0.9$ |
| $X_1$ | 0.71 | 0.73 | 0.46 | 0.87 | 0.89 |
| $X_2$ | 0.24 | 0.26 | 0.21 | 0.41 | 0.36 |
| $X_3$ | 0.03 | 0.03 | 0.06 | 0.15 | 0.15 |
| $X_4$ | 0.00 | 0.01 | 0.04 | 0.09 | 0.07 |

### 3.4.3 Controlling for dependence in the sensitivity measure

The literature on sensitivity indices has long been concerned with the implications for sensitivity analyses of statistical dependence between inputs. In particular for variance-based sensitivities, Saltelli and Tarantola (2002); Oakley and O'Hagan (2004), show that correlation between inputs can impact sensitivity measures in ways that do not reflect the functional dependencies in the model's aggregation function and are thus viewed as spurious; see Section 4.3 in Borgonovo and Plischke (2016) for more discussion

of this topic and extensive references. Refinements of variance-based sensitivity indices have been proposed to address dependence between inputs, indicatively see Xu and Gertner (2008*a*,*b*); Mara and Tarantola (2012).

Our sensitivity measure $\Gamma_i$, similarly to many other sensitivity measures (Borgonovo et al., 2016), is evaluated on the joint distribution of $(X_i, Y)$. As a result, it does not control for dependence between inputs and is therefore subject to problems of confounding typical in multivariate analyses. In this section, we put forward a proposal for generalising the sensitivity measure $\Gamma_i$, in order to take into account such effects.

To proceed with the definition, denote by $N = \{1, \ldots, n\}$ and $\mathbf{X}_T = (X_j)_{j \in T}$ for $T \subseteq N$. Consider now, for $T \subseteq N \backslash \{i\}$, the quantity $\mathcal{S}(X_i - E(X_i|\mathbf{X}_T), \zeta)$, measuring the reverse sensitivity to that part of $X_i$ that is not already explained by the inputs $\mathbf{X}_T$. With the above in mind, we define the $k^{th}$*-order reverse sensitivity measure* as

$$\Gamma_i^{(k)} = \frac{1}{c_k} \sum_{T \subseteq N \backslash \{i\},\ |T|=k} \mathcal{S}\left(X_i - E(X_i|\mathbf{X}_T), \zeta\right), \quad k = 0, \ldots, n-1,$$

where $c_k = \binom{n-1}{k}$. Thus $\Gamma_i^{(k)}$ represents the average reverse sensitivity to $X_i$, after controlling for all subsets of inputs of size $k$. Note the special cases $\Gamma_i^{(0)} = \Gamma_i$ and $\Gamma_i^{(n-1)} = \mathcal{S}(X_i - E(X_i|\boldsymbol{X}_{-i}), \zeta)$. If $X_i$ is independent of $\boldsymbol{X}_{-i}$, then $\Gamma_i^{(k)} = \Gamma_i$ for all $k = 0, \ldots, n-1$.

**Example (continued).** Continuing with the insurance portfolio example, we work out $k^{\text{th}}$-order reverse sensitivity measures for $k = 0, \ldots, 3$, with the RN-density $\zeta$ derived from problem (3.4) for an 10% increase of $\text{VaR}_{0.9}(Y)$. The results are summarised in Table 3.6 and show that, as the order of the sensitivity measure increases, the sensitivity of some input factors is impacted more than that of others. This is particularly noticeable for $X_4$: for $k = 3$, where all other input factors are controlled for, the sensitivity drops substantially.

Comparing the first column in Table 3.6, that is $\Gamma_i$, with Table 3.5, the sensitivity for the independent input vector, we see that the rank of $X_2$ and $X_4$ are reversed. The observed impact of dependence between inputs on the importance ranking of $X_4$ can be understood as follows. $X_4$ represents the percentage of reinsurance recovery lost due to default. In our baseline model, $X_4$ is dependent on $L = X_3(X_1 + X_2)$, with Gaussian copula correlation of 0.6. The correlation reflects the notion that such recovery losses are more likely under those scenarios when they are most needed (i.e. $L$ is large), leading to a high sensitivity to $X_4$.

Table 3.6: The $k^{\text{th}}$-order reverse sensitivity measure $\Gamma_i^{(k)}$, for $k = 0, \ldots, 3$, of problem (3.4) for a 10% increase of $\text{VaR}_{0.9}(Y)$.

| **Input** | $k = 0$ | $k = 1$ | $k = 2$ | $k = 3$ |
|---|---|---|---|---|
| $X_1$ | 0.88 | 0.81 | 0.76 | 0.71 |
| $X_2$ | 0.36 | 0.30 | 0.29 | 0.31 |
| $X_3$ | 0.15 | 0.13 | 0.12 | 0.13 |
| $X_4$ | 0.60 | 0.48 | 0.34 | 0.18 |

## 3.5 Financial application: a London Insurance Market portfolio

In this section we demonstrate the use of the sensitivity measures $\Gamma_{i,v}$ and $\Delta_{i,v}$, in a more realistic insurance risk model with a higher number of inputs. This is a proprietary model of a London Insurance Market portfolio, currently in use by a participant in that market. We have been supplied by the model owner with a Monte Carlo sample of size $M = 500{,}000$, containing simulated observations from input factors $\mathbf{X} = (X_1, \ldots, X_{72})$ and output $Y$. Each of the $X_i$'s represents a normalised loss for a particular part of the portfolio and is measured on the same scale. The output $Y$ stands for the portfolio loss.

The aggregation function $g$ is linear, specifically

$$Y = g(\mathbf{X}) = \sum_{j=1}^{72} w_j X_j,$$

for a vector of weights $\mathbf{w} = (w_1, \ldots, w_{72})$. The linearity of $g$ is not used for sensitivity calculations, since the reverse sensitivity testing framework makes no assumptions on the form of $g$. We do not have access to the joint probability distribution that was used to generate samples from the random vector $\mathbf{X}$; in fact the distribution of $\mathbf{X}$ is not given in closed form, as samples from $\mathbf{X}$ are themselves outputs of a different model, which remains a completely black box to us.

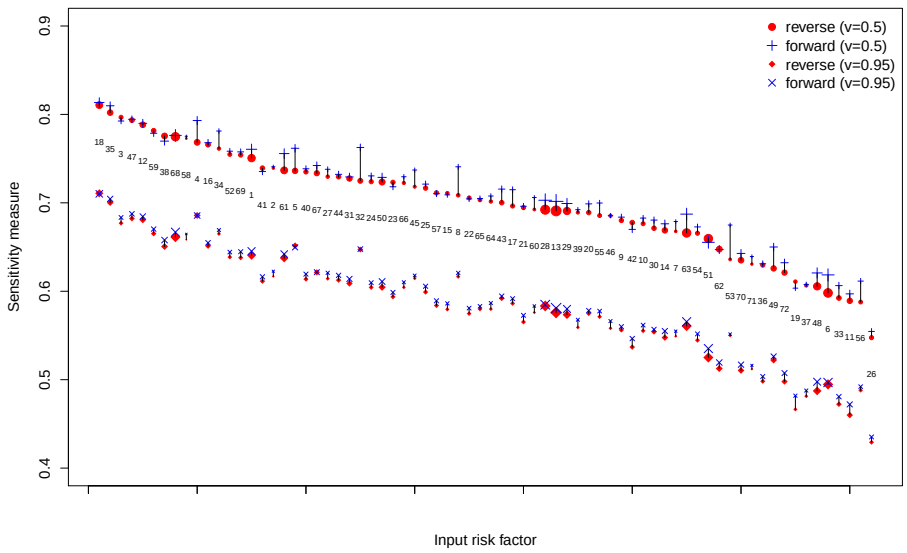


Figure 3.7: Reverse and forward sensitivity measures $\Gamma_{i,v}, \Delta_{i,v}$ for the London Insurance Market portfolio, for $v = 0.5$ and $v = 0.95$.

We consider optimisation problem (3.5) with risk measure constraints on VaR and ES given by $q = \mathrm{VaR}^Q_{0.95}(Y) = 1.08\mathrm{VaR}_{0.95}(Y)$ and $s = \mathrm{ES}^Q_{0.95}(Y) = 1.1\mathrm{ES}_{0.95}(Y)$. In Figure 3.7, the reverse and forward sensitivity measures $\Gamma_{i,v}, \Delta_{i,v}$, for $v = 0.5$ and $v = 0.95$, are presented for all 72 inputs. The input factors are ordered according to

$\Gamma_{i,0.5}$ and the sizes of the markers reflect the weights $w_i$ attached to the individual input factors $X_i$.

Observations on the plot of Figure 3.7:

- The ranking of input factors according to $\Gamma_{i,0.5}$ and $\Gamma_{i,0.95}$ is not fully consistent; moving focus to the tails of input factors changes the order of the sensitivity measures. Hence, under the stressed model, for some input factors the expectation is affected more, while for others the impact is higher in the tail.

- For $v = 0.5$, the ranking produced by the reverse and forward sensitivity metrics is not equivalent. However, once the focus is moved towards the tails of risk factor distributions (e.g. $v = 0.95$), the discrepancy of the two sensitivity measures diminishes.

- There is no clear relation between the sizes of the markers and the ranking of input factors. This means that the sensitivity measure $\Gamma_{i,v}$ does not solely reproduce the size of the weight $w_i$.

To elaborate on the last of those points, in Figure 3.8 (left), the reverse sensitivities $\Gamma_{i,0.95}$ are plotted against the weights $w_i$. There is a broadly increasing relation, which is not unreasonable. Given the linearity of the aggregation function, a higher weight $w_i$ implies a higher *local sensitivity* $\frac{\partial g}{\partial x_i}$ (Borgonovo and Plischke, 2016). But the relation is by no means deterministic: weight is a weak predictor of the reverse sensitivity measure $\Gamma_{i,v}$.

Furthermore, the reverse sensitivity measure does not only reflect the shape of the input risk factor distributions. In Figure 3.8 (right), $\Gamma_{i,0.95}$ is displayed against the scaled percentiles $\frac{\mathrm{VaR}_{0.95}(X_i)}{E(X_i)} - 1$, not showing a clear pattern. Hence the two plots in Figure 3.8 demonstrate that the proposed reverse sensitivity measure does

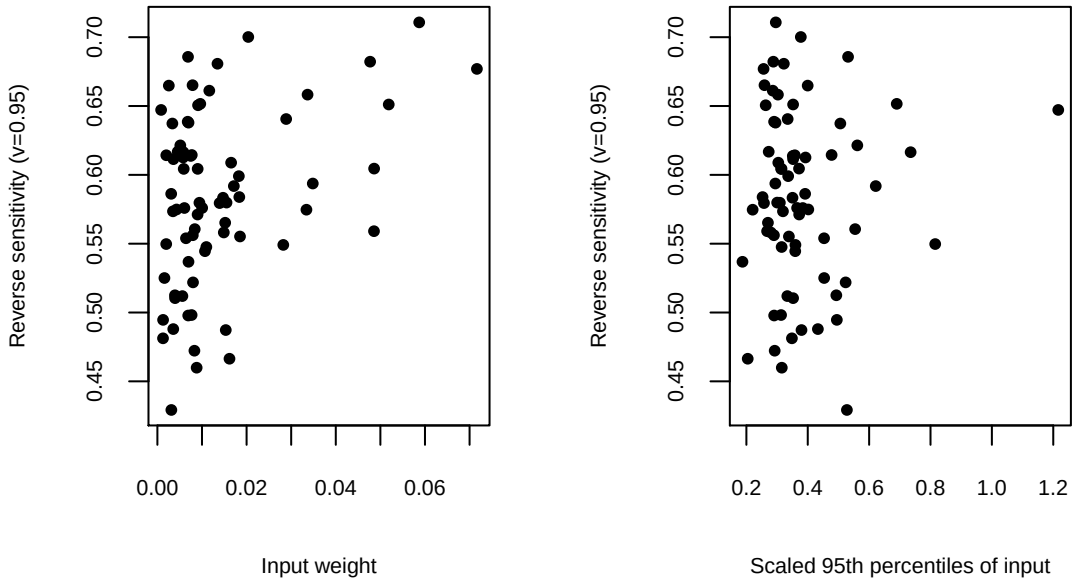


Figure 3.8: Reverse sensitivity measure $\Gamma_{i,0.95}$ for the London Insurance Market portfolio, against weights $w_i$ (left) and scaled input percentiles $\frac{\mathrm{VaR}_{0.95}(X_i)}{E(X_i)} - 1$ (right).

not reproduce easily observed characteristics of the aggregation function $g$ or of the distributions of the inputs $X_i$.

## 3.6 Conclusions

We proposed a reverse sensitivity testing framework that is appropriate for contexts where model inputs are uncertain and the relationship between model inputs and outputs is complex and not necessarily given in analytical form. At the core of the reverse sensitivity framework is a stress on the output distribution, corresponding to an increase in the value of a risk measure applied on the output and representing a plausible but adverse model change. This leads to stressed probabilities under which the distribution of the input factors (marginals and dependence structure) is altered such that the output distribution is subjected to the required stress.

We provided analytical solutions of the stressed probability measure under an increase of the VaR and ES risk measures. These explicit solutions facilitate straightforward implementation in a Monte Carlo simulation context and inspection of changes in the distributions of inputs. A new class of reverse sensitivity measures is introduced, quantifying the extent that the distribution of an input factor is distorted by the transition to a stressed probability. Analysis of stochastic order relations induced by the change of measure provides assurance that the proposed method has desirable properties.

The reverse sensitivity framework can be easily deployed by a risk analyst with access only to a set of input / output scenarios, simulated under the baseline model. Thus there is no need for a detailed consideration of the model structure or of simulating additional scenarios, involving computationally expensive model evaluations. Thus the proposed framework is immediately applicable to industry applications.

# Appendix 3.A Stochastic comparisons

The proposed reverse sensitivity testing framework is based on the change from a baseline probability measure $P$ to a stressed probability $Q$. The optimisation problems of Section 3.3 ensure that under $Q$ the value of particular risk measures applied on $Y$ increases. But the broader changes in the distributions of input factors $\boldsymbol{X}$ and output $Y$ arising from the change of measure are also of interest in a risk management context. For $Q$ to be meaningfully called a 'stressed measure', we argue that three properties should be fulfilled. First, under $Q$ the distribution of the output should dominate (in a suitable stochastic order relation) the output distribution under the baseline model. Second, under the assumptions of a non-decreasing aggregation function and positive dependence between input factors, the distribution of the input vector $\boldsymbol{X}$ under $Q$ should stochastically dominate the distribution of $\boldsymbol{X}$ under $P$. Third, an increase in

the extent to which risk measures are stressed should be reflected in the distributions of output and inputs under the corresponding stressed probabilities. In this section we aim to give precise conditions under which the above properties are fulfilled. Note that most of the discussion is not contingent on $Q$ being a solution of one of the optimisation problems of Section 3.3.

We adopt the standard definitions of stochastic order relations. For distribution functions $F, G$ we write $F \preceq_{st} G$ if $G$ is larger than $F$ in first-order stochastic dominance, that is $F(\boldsymbol{x}) \geq G(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$. For univariate $F, G$, we denote $F \preceq_{icx} G$ if $G$ is larger than $F$ in increasing convex (or stop-loss) order, that is $\int_u^1 F^{-1}(s)\mathrm{d}s \leq \int_u^1 G^{-1}(s)\mathrm{d}s$ for all $u \in (0,1)$. The following dependence concepts are of importance, see Denuit et al. (2006):

- An $m$-dimensional random vector $\boldsymbol{Z}$ is *stochastically increasing* (or *positively regression dependent*) in a random variable $W$, denoted by $\boldsymbol{Z} \uparrow_{si} W$, if $P(\boldsymbol{Z} > \boldsymbol{z} \,|\, W = w)$ is non-decreasing in $w$, for all $\boldsymbol{z} \in \mathbb{R}^m$.

- An $m$-dimensional random vector $\boldsymbol{Z}$ is *associated* if $\mathrm{Cov}(h_1(\boldsymbol{Z}), h_2(\boldsymbol{Z})) \geq 0$, for all component-wise non-decreasing functions $h_1, h_2 : \mathbb{R}^m \to \mathbb{R}$ for which the covariance exists.

- The random couple $(W, Z)$ is *positively quadrant dependent* (PQD) if $P(W \leq w, Z \leq z) \geq P(W \leq w)P(Z \leq z)$ for all $w, z \in \mathbb{R}$.

- The random couple $(W, Z)$ is *negatively quadrant dependent* (NQD) if $P(W \leq w, Z \leq z) \leq P(W \leq w)P(Z \leq z)$ for all $w, z \in \mathbb{R}$.

For a pair of random variables $(W, Z)$ the above definitions are successively weaker: $Z \uparrow_{si} W$ implies that $(Z, W)$ is associated, which implies PQD, see Esary et al. (1967). We write $\boldsymbol{Z}_{-k} = (Z_1, \ldots, Z_{k-1}, Z_{k+1}, \ldots, Z_m)$, $1 \leq k \leq m$ for the $(m-1)$-dimensional sub-vector of $\boldsymbol{Z}$ deprived of its $k$-th component.

The next two propositions characterise the stochastic ordering of inputs and output under two different probabilities $Q^1, Q^2$, making alternative assumptions on distributions under $P$, on $g$ and on the form of the corresponding two RN-densities.

**Proposition 3.A.1.** Let $Q^1, Q^2 \in \mathcal{P}$ be two probability measures with $\frac{\mathrm{d}Q^1}{\mathrm{d}P} = \eta_1(Y)$, $\frac{\mathrm{d}Q^2}{\mathrm{d}P} = \eta_2(Y)$, for some non-negative functions $\eta_j$, $j = 1, 2$. If the RN-densities cross once, such that for some $d \in \mathbb{R}$

$$\eta_2(y) \begin{cases} \leq \eta_1(y) & y < d \\ \geq \eta_1(y) & y \geq d, \end{cases} \tag{3.10}$$

then the following hold:

1. $F_Y^{Q^1} \preceq_{st} F_Y^{Q^2}$
2. For given $i \in \{1, \ldots, n\}$, if $E\left((X_i - t)_+ | Y = y\right)$ is non-decreasing in $y$ for all $t \in \mathbb{R}$, then $F_{X_i}^{Q^1} \preceq_{icx} F_{X_i}^{Q^2}$.
3. For given $i \in \{1, \ldots, n\}$, if $X_i \uparrow_{si} Y$, then $F_{X_i}^{Q^1} \preceq_{st} F_{X_i}^{Q^2}$.

**Proposition 3.A.2.** Let $Q^1, Q^2 \in \mathcal{P}$ be two probability measures with $\frac{\mathrm{d}Q^1}{\mathrm{d}P} = \eta_1(Y)$, $\frac{\mathrm{d}Q^2}{\mathrm{d}P} = \eta_2(Y)$ for some non-negative functions $\eta_j$, $j = 1, 2$. Assume that $\eta_2 - \eta_1$ is non-decreasing. Then the following hold:

1. $F_Y^{Q^1} \preceq_{st} F_Y^{Q^2}$.
2. If the aggregation function $g$ is non-decreasing in coordinate $i$ and $X_i$ is independent of $\boldsymbol{X}_{-i}$, then $F_{X_i}^{Q^1} \preceq_{st} F_{X_i}^{Q^2}$.
3. Assume that the aggregation function $g$ is non-decreasing.
   (a) For given $i \in \{1, \ldots, n\}$, if $(X_i, Y)$ is PQD, then $F_{X_i}^{Q^1} \preceq_{st} F_{X_i}^{Q^2}$.

(b) If $\boldsymbol{X}$ is associated, then $F_{\boldsymbol{X}}^{Q^1} \preceq_{st} F_{\boldsymbol{X}}^{Q^2}$.

Part 1. of both Propositions 3.A.1 and 3.A.2 reflects the comparative impact of the stress on the output $Y$, while parts 2. and 3. characterise the impact of the stress on the inputs. An example where the assumption of Proposition 3.A.1, part 3., is satisfied is the following. Suppose the input vector $\boldsymbol{X}$ is multivariate normal and $Y = h(\sum_{i=1}^n w_i X_i)$ for an increasing function $h$ and $w_i \in \mathbb{R}$ for all $i$. If $\mathrm{Cov}(X_i, h^{-1}(Y)) = \sum_{j=1}^n w_j \mathrm{Cov}(X_i, X_j) \geq 0$, then $X_i \uparrow_{si} Y$ holds. The assumption in Proposition 3.A.2 part 3.(a) holds for example if $\boldsymbol{X}_{-i} \uparrow_{st} X_i$ and $g$ is non-decreasing.

Propositions 3.A.1 and 3.A.2 allow for a stochastic comparison of the output and the input factors under the stressed and the baseline model. In particular, Proposition 3.A.1 applies to the solutions of problems (3.3), (3.5) and (3.7) with $Q^2 = Q$ and $Q^1 = P$. Proposition 3.A.2 applies to optimisation problem (3.1), with $Q^2 = Q$ and $Q^1 = P$, if the RN-density of the solution is a non-decreasing function of $Y$. Recall that the RN-density of the solutions to (3.3), (3.5) case 1, and (3.7) are non-decreasing. Moreover, for a stressed model under which the input $X_i$ stochastically dominates, in first-order or increasing convex order, the distribution of the input under the baseline model, the introduced sensitivity measure $\Gamma_i$ is positive.

Proposition 3.A.1 also enables to contrast stressed probabilities corresponding to different stress levels. For example, when solving optimisation problem (3.3) with two different VaR constraints, the output under the stressed model corresponding to a higher VaR should stochastically dominate the output under the other stressed model. The next lemma associates Proposition 3.A.1 with solutions of the optimisation problems (3.3) and (3.5).

**Lemma 3.A.3.** The crossing condition of Proposition 3.A.1 is satisfied for:

1. Two solutions $Q^1, Q^2$ of optimisation problem (3.3) with constraints $\mathrm{VaR}_\alpha^{Q^1}(Y) = q_1$ respectively $\mathrm{VaR}_\alpha^{Q^2}(Y) = q_2$, and $q_1 < q_2$.

2. Two solutions $Q^1, Q^2$ of optimisation problem (3.5) with constraints $\mathrm{VaR}_\alpha^{Q^1}(Y) = \mathrm{VaR}_\alpha^{Q^2}(Y) = q$ and $\mathrm{ES}_\alpha^{Q^1}(Y) = s_1$, respectively $\mathrm{ES}_\alpha^{Q^2}(Y) = s_2$, and $s_1 < s_2$.

The second part of Lemma 3.A.3 holds true for both types of solutions of (3.5).

**Example (continued).** Applying Proposition 3.A.1 to the two optimisation problems in this example, we immediately verify that the output under the stressed probabilities first-order stochastically dominates the output under the baseline probability, see Figures 3.1 and 3.3. Moreover, the aggregation function $g$ is non-decreasing and it can be verified that, for instance, $(X_4, Y)$ is PQD. Hence, following Proposition 3.A.2 part 3.(a), the distribution of $X_4$ under the stressed probability first-order stochastically dominates that under the baseline probability. This can be seen in Figures 3.2 and 3.4.

An illustration of Lemma 3.A.3 is given in Figure 3.9. The left plot shows the RN-densities of solutions to (3.3) with two different stress levels. The black line corresponds to an increase of VaR of 10%, the same as in Figure 3.1, and the grey line to an increase of VaR of 15%. The plot to the right displays the RN-densities of solutions to (3.5) for an increase of 10% in VaR and 9% in ES (black) and an increase of 10% in VaR and 13% in ES (grey), see Figure 3.3. It is seen how in both cases, the two RN-densities satisfy the crossing condition of Proposition 3.A.1.

# Appendix 3.B Proofs

*Proposition 3.3.1.* A similar result can be found in Cambou and Filipović (2017), we also refer to Csiszár (1975) for the general form of the solution. It is immediately verified that $\zeta$ is a RN-density for which $Q^\zeta(Y \in B_i) = \alpha_i$, $i = 1, \ldots, I$. Let $\xi$ be any RN-density that satisfies $Q^\xi(Y \in B_i) = \alpha_i$, $i = 1, \ldots, I$. Using Jensen inequality, the

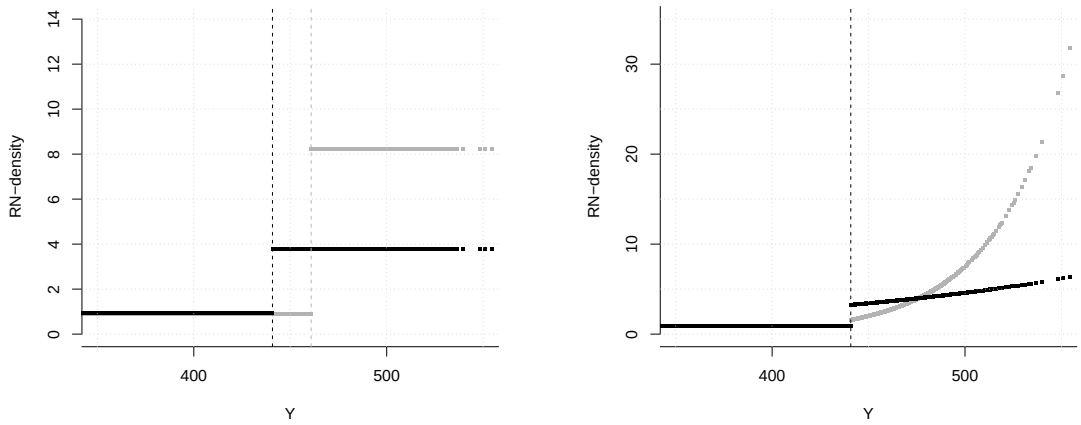


Figure 3.9: Left: simulated RN-densities of the solution to (3.4) with a 10% (black) and 15% (grey) increase in VaR. Right: simulated RN-densities of the solution to (3.5) case 1, with a 10% increase in VaR and 9% (black) and 13% (grey) increase in ES.

KL-divergence of $Q^\xi$ with respect to $P$ fulfils

$$\begin{aligned} D_{\mathrm{KL}}(Q^\xi \| P) &= \sum_{i=0}^{I} E\left(\xi \log(\xi) \,|Y \in B_i\right) P(Y \in B_i) \\ &\geq \sum_{i=0}^{I} E\left(\xi \,|Y \in B_i\right) \log\left(E\left(\xi \,|Y \in B_i\right)\right) P(Y \in B_i) \\ &= \sum_{i=0}^{I} \alpha_i \log\left(\frac{\alpha_i}{P(Y \in B_i)}\right) \\ &= D_{\mathrm{KL}}(Q^\zeta \| P). \end{aligned}$$

Therefore $Q^\zeta$ is a solution of (3.2). Uniqueness follows by strict convexity of the KL-divergence, see Csiszár (1975). □

*Proposition 3.3.2.* Assume that $P(q - \varepsilon < Y < q) > 0$ for all $\varepsilon > 0$. Then, it is immediate to verify that $\zeta$ is a RN-density such that $\mathrm{VaR}_\alpha^{Q^\zeta}(Y) = q$. Let $\xi = \frac{\mathrm{d}Q^\xi}{\mathrm{d}P}$ be a RN-density for which $\mathrm{VaR}_\alpha^{Q^\xi}(Y) = q$. By Jensen inequality, the KL-divergence of $Q^\xi$

with respect to $P$ is

$$\begin{aligned}
D_{\mathrm{KL}}(Q^{\xi}\|P) &= E\left(\xi\log(\xi)\,|\,Y<q\right)P(Y<q) + E\left(\xi\log(\xi)\,|\,Y\geq q\right)P(Y\geq q)\\
&\geq Q^{\xi}(Y<q)\log\left(\frac{Q^{\xi}(Y<q)}{P(Y<q)}\right) + Q^{\xi}(Y\geq q)\log\left(\frac{Q^{\xi}(Y\geq q)}{P(Y\geq q)}\right)\\
&= k\Big(Q^{\xi}(Y<q), P(Y<q)\Big),
\end{aligned}$$

where we define $k(x,y) = x\log(\frac{x}{y}) + (1-x)\log(\frac{1-x}{1-y})$, for $0<x<1$, $0<y<1$. Inspection shows that, for fixed $0<y<1$, $x\to k(x,y)$ is non-increasing on $(0,y]$. Moreover it holds

$$Q^{\xi}(Y<q)\leq\alpha\leq P(Y\leq \mathrm{VaR}_{\alpha}(Y))\leq P(Y<q).$$

The KL-divergence of $Q^{\xi}$ is thus larger than the KL-divergence of $Q^{\zeta}$,

$$\begin{aligned}
D_{\mathrm{KL}}(Q^{\xi}\|P) &\geq k\Big(Q^{\xi}(Y<q), P(Y<q)\Big)\\
&\geq k\Big(\alpha, P(Y<q)\Big)\\
&= \alpha\log\left(\frac{\alpha}{P(Y<q)}\right) + (1-\alpha)\log\left(\frac{1-\alpha}{P(Y\geq q)}\right)\\
&= D_{\mathrm{KL}}(Q^{\zeta}\|P),
\end{aligned}$$

and $Q^{\zeta}$ is a solution of (3.3). Uniqueness follows by strict convexity of the KL-divergence.

Assume now that there exists $\varepsilon>0$ such that $P(q-\varepsilon<Y<q)=0$. If $P(Y=q)=0$, by the absolute continuity of the probability measures, the optimisation problem (3.3) does not admit a solution. Hence, we assume that $P(Y=q)>0$. Let $Q^{\xi}$ be a RN-density for which $\mathrm{VaR}_{\alpha}^{Q^{\xi}}(Y)=q$. Denote $r=Q^{\xi}(Y\leq q)$ and $p=P(Y\leq q)$.

The KL-divergence of $Q^\xi$ with respect to $P$ is

$$\begin{aligned} D_{\mathrm{KL}}(Q^\xi \| P) &= E(\xi \log(\xi) \,|\, Y \leq q)p + E(\xi \log(\xi) \,|\, Y > q)(1-p) \\ &\geq r \log\left(\frac{r}{p}\right) + (1-r)\log\left(\frac{1-r}{1-p}\right) \\ &= D_{\mathrm{KL}}(Q^{\xi^r} \| P), \end{aligned}$$

where we define $\xi^u = \frac{\mathrm{d}Q^{\xi^u}}{\mathrm{d}P} = \frac{u}{p}\mathbb{1}_{\{Y \leq q\}} + \frac{1-u}{1-p}\mathbb{1}_{\{Y>q\}}$, $0 \leq u \leq 1$. The family of RN-densities $\xi^u$ fulfil $\mathrm{VaR}_\alpha^{Q^{\xi^u}}(Y) = q$ if and only if $\alpha \leq u < \alpha\frac{p}{P(Y<q)}$. In particular this holds for the RN-density $\xi^r$. Hence the optimisation problem (3.3) is reduced to minimise $D_{\mathrm{KL}}(Q^{\xi^u}\|P)$ subject to $\alpha \leq u < \alpha\frac{p}{P(Y<q)}$. As a function of $u$ the KL-divergence $D_{\mathrm{KL}}(Q^{\xi^u}\|P)$ is non-increasing on $(0, p]$, hence the optimisation problem does not admit a solution as $\alpha\frac{p}{P(Y<q)} < p$. □

*Proposition 3.3.5.* For $i = 1, 2$, equation (3.6) can be rewritten as

$$\frac{\partial}{\partial\theta} E(e^{\theta(Y-q)}|A_i) = \frac{E\Big((Y-q)e^{\theta(Y-q)}|A_i\Big)}{E(e^{\theta(Y-q)}|A_i)} = s - q.$$

The left hand side is increasing for positive $\theta$, negative for $\theta = 0$ and diverges for $\theta \uparrow \theta_{max}$, where $\theta_{max} = \sup\{\theta > 0 \,|\, E(e^{\theta Y}|A_i) < \infty\}$, by properties of the moment generating function. Thus, for $i = 1, 2$, there exists a unique positive solution $\theta_i^*$ of (3.6).

*Case 1.* The RN-density $\zeta_1$ fulfils the constraints in (3.5) since $Q^{\zeta_1}(Y < q) = \alpha$, $Q^{\zeta_1}(Y \leq q) \geq \alpha$ and the ES constraint is equivalent to $(1-\alpha)(s-q) = E^{Q^{\zeta_1}}((Y-q)_+)$. Let $\xi = \frac{\mathrm{d}Q^\xi}{\mathrm{d}P}$ be a RN-density satisfying the constraints of problem (3.5) and denote $r = Q^\xi(A_1^c)$ and $p = P(A_1^c)$. Using Jensen's inequality, the KL-divergence of $Q^\xi$ with

respect to $P$ fulfils

$$\begin{aligned} D_{\mathrm{KL}}(Q^{\xi}\|P) &= E\left(\xi\log(\xi)\mathbb{1}_{A_1^c}\right) + E\left(\xi\log(\xi)\mathbb{1}_{A_1}\right) + \theta_1^*(1-\alpha)(s-q) \\ &\quad - E\left(\xi\log\left(e^{\theta_1^*(Y-q)}\right)\mathbb{1}_{A_1}\right) \\ &\geq r\log\left(\frac{r}{p}\right) + \theta_1^*(1-\alpha)(s-q) + E\left(\xi\log\left(\frac{\xi}{e^{\theta_1^*(Y-q)}}\right)\Big|A_1\right)(1-p). \end{aligned}$$

Recall that the perspective of a convex function $f$, defined by $h(x,y) = yf(x/y)$ is itself convex, see Boyd and Vandenberghe (2004). Applying then Jensen's inequality to $h(x,y) = y\log(\frac{y}{x})$, the third term becomes

$$\begin{aligned} E\left(\xi\log\left(\frac{\xi}{e^{\theta_1^*(Y-q)}}\right)\Big|A_1\right)(1-p) &\geq E(\xi\,|A_1)\log\left(\frac{E(\xi\,|\,A_1)}{E(e^{\theta_1^*(Y-q)}\,|\,A_1)}\right)(1-p) \\ &= (1-r)\log\left(\frac{(1-r)}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)}\right). \end{aligned}$$

Collecting all terms,

$$\begin{aligned} D_{\mathrm{KL}}(Q^{\xi}\|P) &\geq r\log\left(\frac{r}{p}\right) + \theta_1^*(1-\alpha)(s-q) + (1-r)\log\left(\frac{(1-r)}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)}\right) \\ &= k\Big(r,p,E\Big(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\Big)\Big), \end{aligned}$$

where we define $k(x,y,z) = x\log(\frac{x}{y}) + \theta_1^*(1-\alpha)(s-q) + (1-x)\log(\frac{1-x}{z})$, for $0 < x < 1$ and $y,z > 0$. For fixed $y,z > 0$, the function $x \to k(x,y,z)$ is decreasing on $\left(0, \frac{y}{y+z}\right]$. The condition on $\theta_1^*$ in 1. is equivalent to

$$\alpha \leq \frac{p}{p + E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)}.$$

Therefore, noting that $r \leq \alpha$, we obtain

$$D_{\mathrm{KL}}(Q^{\xi}\|P) \geq k\Big(r, p, E\Big(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\Big)\Big) \geq k\Big(\alpha, p, E\Big(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\Big)\Big) = D_{\mathrm{KL}}(Q^{\zeta_1}\|P).$$

The last equality follows since

$$\begin{aligned}
D_{\mathrm{KL}}(Q^{\zeta_1}\|P) &= \alpha \log\left(\frac{\alpha}{p}\right) \\
&\quad + \frac{1-\alpha}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)} E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1} \log\left(\frac{1-\alpha}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)} e^{\theta_1^*(Y-q)}\right)\right) \\
&= \alpha \log\left(\frac{\alpha}{p}\right) + (1-\alpha)\log\left(\frac{1-\alpha}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)}\right) \\
&\quad + \theta_1^* \frac{1-\alpha}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)} E\Big(e^{\theta_1^*(Y-q)}(Y-q)_+\Big) \\
&= \alpha \log\left(\frac{\alpha}{p}\right) + (1-\alpha)\log\left(\frac{1-\alpha}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)}\right) + \theta_1^*(1-\alpha)(s-q) \\
&= k\Big(\alpha, p, E\Big(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\Big)\Big).
\end{aligned}$$

Therefore $Q^{\zeta_1}$ is a solution of (3.5). Uniqueness follows by strict convexity of the KL-divergence.

*Case 2.* The proof of case 2 is similar to that of case 1, replacing the set $A_1$ with $A_2$ and $\zeta_1$ with $\zeta_2$. The RN-density $\zeta_2$ fulfils the constraints (3.5). Letting $\xi = \frac{\mathrm{d}Q^{\xi}}{\mathrm{d}P}$ be a RN-density satisfying the constraints of problem (3.5), then the KL-divergence of $Q^{\xi}$ with respect to $P$ can be bounded by

$$D_{\mathrm{KL}}(Q^{\xi}\|P) \geq k\Big(Q^{\xi}(A_2^c), P(A_2^c), E\Big(e^{\theta_2^*(Y-q)}\mathbb{1}_{A_2}\Big)\Big),$$

where the function $k(x, y, z)$ has been defined above. For fixed $y, z > 0$, the function $x \to k(x, y, z)$ is increasing on $\left[\frac{y}{y+z}, 1\right)$. Moreover, the condition on $\theta_2^*$ in 2. is equivalent

to

$$\frac{P(A_2^c)}{P(A_2^c) + E\left(e^{\theta_2^*(Y-q)}\mathbb{1}_{A_2}\right)} \leq \alpha.$$

Since $\alpha \leq Q^\xi(A_2^c)$ we obtain

$$D_{\mathrm{KL}}(Q^\xi \| P) \geq k\Big(\alpha, P(A_2^c), E\Big(e^{\theta_2^*(Y-q)}\mathbb{1}_{A_2}\Big)\Big) = D_{\mathrm{KL}}(Q^{\zeta_2} \| P),$$

which is the KL-divergence of $Q^{\zeta_2}$. □

*Proposition 3.4.3.* We also refer to Theorem 6 in Kachapova and Kachapov (2012). The first two properties are immediate. For 3. if $X_i$ and $Y$ are comonotonic, $\zeta$ and $\zeta_{|X_i}$ are also comonotonic since $\zeta$ is a non-decreasing function of $Y$ and $\zeta_{|X_i}$ a non-decreasing function of $X_i$. Part 4. follows by a similar argument. Properties 5. and 6. are consequences of the invariance of PQD (NQD) under non-decreasing (non-increasing) transformations, see Lemma 1 in Lehmann (1966). □

*Proposition 3.A.1.* Let $y \leq d$, then it holds $Q^2(Y \leq y) = E(\eta_2(Y)\mathbb{1}_{\{Y \leq y\}})$ $\leq E(\eta_1(Y)\mathbb{1}_{\{Y \leq y\}}) = Q^1(Y \leq y)$. For $y > d$, it holds $Q^2(Y \leq y) = 1 - Q^2(Y > y) = 1 - E(\eta_2(Y)\mathbb{1}_{\{Y > y\}}) \leq 1 - E(\eta_1(Y)\mathbb{1}_{\{Y > y\}}) = Q^1(Y \leq y)$. For the second part we have, for all $t \in \mathbb{R}$, using the tower property under $P$,

$$E^{Q^2}\left((X_i - t)_+\right) = E^{Q^2}\Big(E\left((X_i - t)_+ | Y\right)\Big) \geq E^{Q^1}\Big(E\left((X_i - t)_+ | Y\right)\Big) = E^{Q_1}\left((X_i - t)_+\right),$$

by first-order stochastic dominance of $Y$ with respect to the measures $Q^1, Q^2$. The last claim follows using a similar argument. □

*Proposition 3.A.2.* The RN-densities have to cross once due to normalisation, therefore part 1. applies. In the rest of the proof, let $h = \eta_2 - \eta_1$.

To prove part 2., let $g$ be non-decreasing in coordinate $i$ and $X_i$ independent of $\boldsymbol{X}_{-i}$. For any $t \in \mathbb{R}$, using the Fortuin-Kasteleyn-Ginibre inequality (Wüthrich and

Merz, 2013), we have

$$Q^2(X_i > t) - Q^1(X_i > t) = E\left(h(Y)\mathbb{1}_{\{X_i>t\}}\right) = E\left(E\Big(h(Y)\mathbb{1}_{\{X_i>t\}} \,|\, \boldsymbol{X}_{-i}\Big)\right)$$
$$\geq E\left(E\Big(h(Y)\,\Big|\,\boldsymbol{X}_{-i}\Big)\right) P(X_i > t) = 0,$$

proving first-order stochastic dominance.

To show part 3.(a), assume that $g$ is non-decreasing and $(X_i, Y)$ are PQD. Hence, for all $t \in \mathbb{R}$,

$$Q^2(X_i > t) - Q^1(X_i > t) = E\left(\mathbb{1}_{\{X_i>t\}}h(Y)\right) \geq 0,$$

where the last inequality follows from Lemma 3 in Lehmann (1966). Part 3.(b) follows by association of the vector $(h(Y), \boldsymbol{X})$, using a similar argument. □

*Lemma 3.A.3.* The first claim follows since $\alpha \leq P(Y < q_1) \leq P(Y < q_2)$. For part 2., consider first the case where $P(q - \varepsilon < Y < q) > 0$ for all $\varepsilon > 0$. Denote by $\theta_1^*, \theta_2^*$ the solutions to (3.6) with $q$ and $s_1$, respectively $s_2$. Hence, $\theta_1^* \leq \theta_2^*$, and there exists a $d > q$ such that for all $\omega \in \Omega$ with $Y(\omega) > d$ we have

$$e^{(\theta_2^* - \theta_1^*)(Y(\omega) - q)} \geq \frac{E\left(e^{\theta_2^*(Y-q)}\mathbb{1}_{A_1}\right)}{E\left(e^{\theta_1^*(Y-q)}\mathbb{1}_{A_1}\right)},$$

which implies $\eta_2 \geq \eta_1$ for all $\omega$ with $Y(\omega) > d$. Since on $A_1^c$, $\eta_1 = \eta_2$ $P$-a.s. the RN-densities admit a (unique) crossing point. The argument also holds if $A_1$ is replace with $A_2$. □

# Chapter 4

# Cascade Sensitivity Measures

This chapter is based on the working paper Pesenti et al. (2018*a*) and has been presented at the University of Trieste (2018), at the University of Milano-Bicocca (2018), at the *Workshop on Recent Developments in Dependence Modelling with Applications in Finance and Insurance* (Aegina, 2018), at the *10$^{th}$ Conference in Actuarial Science & Finance* (Samos, 2018) and at the *4$^{th}$ European Actuarial Journal Conference* (Leuven, 2018).

## 4.1 Introduction

Principal tools in sensitivity analysis are sensitivity measures (also *importance measures*), which assign to each input factor a score, ranking the input factors according to their ability to influence (a probabilistic summary of) the output (Borgonovo and Plischke, 2016). For variance-based sensitivity measures, for example, the input factors are distinguished by their ability to affect the output's variance (Saltelli, 2002). In this chapter, and typical in financial risk management applications, the output distribution is summarised through a risk measure, such as the VaR or the ES.

One way to assess the extent to which an input factor affects the output distribution, is to considers partial (Gâteaux) derivatives of the risk measure applied to the output

in the direction of a stressed version of the input factor (Samuelson, 1941; Helton, 1993; Hong, 2009; Tsanakas and Millossovich, 2016). The choice of stress on the input factor depends on the objective of the sensitivity analysis, and commonly consists of either an additive shock applied to the input factor or a perturbation of its distribution (Borgonovo and Plischke, 2016; Glasserman, 1991; Saltelli et al., 2008).

It is known that sensitivity measures defined as partial derivatives do not fully account for interactions among (or dependence between) input factors, and extensions have so far focused on higher order derivatives, see Borgonovo and Plischke (2016) and references therein. However, dependence structures between the input factors might substantially impact the sensitivity to an input factor. In particular, a stress on an input factor should also precipitate stresses in other input factors that are dependent on it. In financial risk management, for example, a sensitivity measure that accounts for the indirect effects of the dependence between input factors is particularly valuable for assessing *systemic risk*, which is concerned with contagion effects that spread through the financial market and distress the whole financial system.

We propose a novel sensitivity measure, termed *cascade sensitivity*, defined as the partial derivative of the risk measure applied to the output, in the direction of a stressed input factor, which explicitly accounts for the dependence of other input factors on the one being stressed. Underpinning the cascade sensitivity framework is a variation of the inverse Rosenblatt transform (Rosenblatt, 1952), enabling a stress on one input factor to spread through the entire input vector, changing all its components according to the input vectors' structure.

We provide explicit analytical representations of the cascade sensitivity to two types of stresses on input factors; (a) a perturbation of the distribution of an input factor, such that the stressed input factor follows a mixture distribution, and (b) an additive random shock applied to the (tail of the) input factor itself. These representations

allow for a straightforward implementation of the cascade sensitivity on one Monte Carlo sample, requiring only the explicit knowledge of the distribution function and the density of the output and of the input factor of interest. Hence, our proposed cascade sensitivity framework is suitable for various practical applications.

Related literature that focus on sensitivity as a directional derivative of a risk measure applied to the output in direction to an input factors include: Hong (2009); Hong and Liu (2009) for the popular VaR and ES risk measures, Antoniano-Villalobos et al. (2018) consider sensitivity to input parameters, while Wang et al. (2018); Tsanakas (2009) study Haezendonck-Goovaerts and entropic risk measures for linear portfolios, respectively. Cao and Wan (2017) analyse derivatives of expected utility in connection to optimal portfolio selection, while Gourieroux et al. (2000, 2006) consider directional derivatives of distortion risk measures with respect to parameter uncertainty for linear aggregation functions.

A paper close to ours is Tsanakas and Millossovich (2016), focusing on sensitivity analysis when the underlying stress on the input factor is an additive shock. We refer to Cont et al. (2010) for a financial application when the stressed input factor follows a mixture distribution. As far as the authors are aware, while the inverse Rosenblatt transform is used in various contexts, the idea of applying it to measure indirect effects of dependence between input factors for sensitivity purposes, is novel. An notable exception is Mai et al. (2015), who study model robustness through introducing uncertainty via a transformation of the input vector.

This chapter is organised as follows. Section 4.2 introduces the notation and mathematical framework. Section 4.3 defines the cascade sensitivity measure as a partial derivative of the output with respect to a stressed input factor, via a variation of the inverse Rosenblatt transform.

Section 4.4 is devoted to the calculation of the cascade sensitivity for different stresses on the input factor. In particular, we provide representations of the cascade sensitivity for two specific stresses that allow for calculations on a single Monte Carlo sample. The applicability of the cascade sensitivity framework is showcased in Section 4.5 via a non-linear insurance portfolio. We conclude, in Section 4.6, with an illustration of the cascade sensitivity to a commercially used London Insurance Market portfolio. All technical assumptions are provided in the Appendix 4.A.

## 4.2 Preliminaries

As introduced in 1.2, we work with the probability space $(\Omega, \mathcal{A}, P)$ and denote by $\boldsymbol{X} = (X_1, \ldots, X_n)$ the vector of input factors. We call $F_j$ the marginal distribution function of input $X_j,\ j = 1, \ldots, n$, and $\boldsymbol{F}$ the joint distribution function of $\boldsymbol{X}$. It is assumed that the joint density $\boldsymbol{f}$ of $\boldsymbol{X}$ exists and we denote by $f_j$ the density of input factor $X_j,\ j = 1, \ldots, n$. The vector of input factors, $\boldsymbol{X}$, is mapped by an aggregation function, $g\colon \mathbb{R}^n \to \mathbb{R}$, assumed to be almost everywhere differentiable, to the (univariate) output $Y = g(\boldsymbol{X})$. We write $H, h$ for the distribution function and the density of the output $Y$, respectively. For any random variable $W$, we denote by $U_W$ a standard uniform random variable comonotonic to $W$. In the case when $W$ has a continuous distribution function, it holds $U_W = F_W(W)$ a.s., where $F_W$ is the distribution function of $W$. For any $n$-dimensional vector $\boldsymbol{W}$, we denote by $\boldsymbol{W}_{-j} = (W_1, \ldots, W_{j-1}, W_{j+1}, \ldots W_n)$ its sub-vector deprived of the $j^{\text{th}}$ component.

We consider the framework where the output $Y = g(\boldsymbol{X})$ is summarised through a risk measure, and aim to quantify the sensitivity of the risk measure applied to the output with respect to an input factor. In this chapter we work with the class of distortion risk measures, see Chapter 1.2.1, which are defined, for a random variable

$Y$, through

$$\rho_\gamma(Y) = \int_0^1 F_Y^{-1}(u)\gamma(u)\mathrm{d}u = E(F_Y^{-1}(U_Y)\gamma(U_Y)),$$

where $\gamma\colon [0,1] \to [0,\infty)$ is a normalised weight function such that $\int_0^1 \gamma(u)\mathrm{d}u = 1$. Examples include the two most common risk measures in practice, VaR and ES.

The objective of this chapter lies in the study of the sensitivity of $\rho_\gamma(Y)$ to input factor $X_i$, $1 \leq i \leq n$. For simplicity, we fix $i \in \{1, \ldots, n\}$ for the rest of the chapter, such that sensitivity to the same input is considered throughout. We call a *stress* to input factor $X_i$ a family of random variables $X_{i,\varepsilon}(\omega) = K(X_i(\omega), \omega, \varepsilon)$, for $\varepsilon \geq 0$, $\omega \in \Omega$ and some mapping $K$, that is almost everywhere differentiable in $\varepsilon$ in a neighbourhood of 0, uniformly in $x$ and $\omega$. Moreover, $K$ satisfies $K(x, \omega, 0) = x$, for all $x \in \mathbb{R}$ and almost all $\omega \in \Omega$. In particular, for any stress $X_{i,\varepsilon}$, it holds that $(X_1, \ldots, X_{i,\varepsilon}, \ldots, X_n)|_{\varepsilon=0} = \boldsymbol{X}$ a.s. We denote by $F_{i,\varepsilon}$, $\varepsilon \geq 0$, the distribution function of $X_{i,\varepsilon}$.

A typical choice of a stress is to apply a random *shock* $Z$ to the input factor $X_i$, such that $X_{i,\varepsilon} = X_i + \varepsilon Z$ (Tsanakas and Millossovich, 2016). Alternatively, the distribution function of the input factor, $F_i$, can be perturbed. Adding uncertainty via the distribution function of the input factor is conceptually different from adding a shock to the input factor, and a common technique used in Bayesian (Gustafson et al., 1996) and robust (Hampel et al., 2011) statistics. Such a perturbation can be constructed starting from a family of distribution functions $F_{i,\varepsilon}$, $\varepsilon \geq 0$, that is continuously differentiable in $\varepsilon$, admits a density for all $\varepsilon$ in a neighbourhood of 0, and fulfils $F_{i,0} = F_i$. We then define the stress to input factor $X_i$ through a *perturbation* by $X_{i,\varepsilon} = F_{i,\varepsilon}^{-1}(Z)$, for a standard uniform random variable $Z$. Depending on the choice of $Z$, the stress may not only distort the input factor $X_i$ but might also change the dependence structure of the input vector $\boldsymbol{X}$. A natural choice, which we consider

in the sequel, is $Z$ to be comonotonic to $X_i$, generating perturbations of the form $X_{i,\varepsilon} = F_{i,\varepsilon}^{-1}(U_{X_i})$, thus not altering the dependence between input factors.

## 4.3 Sensitivity measures

### 4.3.1 Marginal sensitivity

To assess the sensitivity to input $X_i$, we consider sensitivity measures formed by a directional derivative of the risk measure applied to the output distribution, in the direction of a stress to an input factor.

**Definition 4.3.1.** For a stress $X_{i,\varepsilon}$ and a distortion risk measure $\rho_\gamma$, we define the *marginal sensitivity* to input factor $X_i$ by

$$\mathcal{S}_i(\boldsymbol{X}, g, \rho_\gamma) = \frac{\partial}{\partial \varepsilon} \rho_\gamma \Big( g(X_1, \ldots, X_{i,\varepsilon}, \ldots, X_n) \Big) \Big|_{\varepsilon=0},$$

whenever the derivative exists.

The general form of the marginal sensitivity is known in the literature (Hong and Liu, 2009) and stated in the next proposition for completeness. It consists of an expectation involving the derivative of the stress, the gradient of the aggregation function in the direction of the stressed input factor and a weighting according to the chosen risk measure. Further work, closely related to ours, is Hong (2009); Tsanakas and Millossovich (2016); Antoniano-Villalobos et al. (2018).

**Proposition 4.3.2.** Given a stress $X_{i,\varepsilon}$ and under Assumptions 4.A.1 in the appendix, the marginal sensitivity to input factor $X_i$ is

$$\mathcal{S}_i(\boldsymbol{X}, g, \rho_\gamma) = E\Big( \frac{\partial}{\partial \varepsilon} X_{i,\varepsilon} \Big|_{\varepsilon=0} g_i(\boldsymbol{X}) \gamma(U_Y) \Big),$$

where $g_i(\boldsymbol{x}) = \frac{\partial}{\partial x_i} g(\boldsymbol{x})$ denotes the partial derivative of the aggregation function in the $i^{\text{th}}$ component and $\frac{\partial}{\partial \varepsilon} X_{i,\varepsilon}(\omega) = \frac{\partial}{\partial \varepsilon} K(X_i(\omega), \omega, \varepsilon)$, for almost all $\omega \in \Omega$.

*Proof.* See Hong (2009); Hong and Liu (2009). □

**Remark 4.3.3.** Our framework also includes sensitivity to expected utility, considered in Cao and Wan (2017). Note that for the trivial weight function $\gamma \equiv 1$, the distortion risk measure reduces to the expectation, i.e. $\rho_1(\cdot) \equiv E(\cdot)$. Thus, for an utility function $u\colon \mathbb{R} \to \mathbb{R}$, we can then write

$$E(u(g(\boldsymbol{X})) = \rho_1\Big((u \circ g)(\boldsymbol{X})\Big),$$

implying, that expected utilities are a special case of our framework, with aggregation function $u \circ g\colon \mathbb{R}^n \to \mathbb{R}$ and an expectation risk measure.

### 4.3.2 Cascade sensitivity

The marginal sensitivity, defined as a directional derivative, does not account for interactions among input factors (Borgonovo and Plischke, 2016), since it assumes that, apart from the stressed input factor, all other marginal input distributions are unaltered. We provide a novel sensitivity measure, termed *cascade sensitivity*, that incorporates indirect effects induced by the dependence between the input factors, and is constructed in the following way. Given $i \in \{1, \dots, n\}$, we call the subsequent representation an *inverse Rosenblatt* transform of the input vector $\boldsymbol{X}$ (Rosenblatt, 1952; Rüschendorf and de Valk, 1993)

$$\boldsymbol{X} = \boldsymbol{\psi}(X_i, \boldsymbol{V}) = \Big(\psi^{(1)}(X_i, \boldsymbol{V}), \dots, \psi^{(n)}(X_i, \boldsymbol{V})\Big) \text{ a.s.,} \tag{4.1}$$

for a differentiable function $\boldsymbol{\psi} = (\psi^{(1)}, \dots, \psi^{(n)})^\top\colon \mathbb{R}^n \to \mathbb{R}^n$ and a $(n-1)$-dimensional random vector $\boldsymbol{V} = (V_1, \dots, V_{n-1})$, consisting of independent standard uniform random

variables, independent of $X_i$. Note that for $\psi^{(j)}$, $1 \leq j \leq n$, to be differentiable in the first component, it is sufficient that the joint density $\boldsymbol{f}$ is almost everywhere differentiable.

We denote by $\mathcal{R}_i$ the family of inverse Rosenblatt transforms of $\boldsymbol{X}$ with $X_i$ as starting variable as in (4.1), that is $\mathcal{R}_i = \{(\boldsymbol{\psi}, \boldsymbol{V}) \mid \boldsymbol{X} = \boldsymbol{\psi}(X_i, \boldsymbol{V}) \text{ as in (4.1)}\}$. An inverse Rosenblatt transform can be explicitly constructed, for example, through the composition of the Rosenblatt transform with the *standard construction* (Rüschendorf and de Valk, 1993; Rubinstein and Melamed, 1998). For $r = 1, \ldots, n$ and $J \subseteq \{1, \ldots, n\} \backslash \{r\}$, denote by $F_{r|J}(\cdot \mid X_j = x_j,\ j \in J)$ the conditional distribution function of $X_r$ given $X_j = x_j,\ j \in J$. Then, for fixed $i$, it holds a.s. that

$$
\begin{aligned}
X_1 &= F^{-1}_{1|i}(V_1 \,|\, X_i) = \psi^{(1)}(X_i, \boldsymbol{V}), \\
X_2 &= F^{-1}_{2|i,1}(V_2 \,|\, X_i, X_1) = \psi^{(2)}(X_i, \boldsymbol{V}), \\
X_3 &= F^{-1}_{3|i,1,2}(V_3 \,|\, X_i, X_1, X_2) = \psi^{(3)}(X_i, \boldsymbol{V}), \\
&\vdots \\
X_i &= \psi^{(i)}(X_i, \boldsymbol{V}), \\
&\vdots \\
X_n &= F^{-1}_{n|1,\ldots,n-1}(V_{n-1} \,|\, X_1, \ldots X_{n-1}) = \psi^{(n)}(X_i, \boldsymbol{V}),
\end{aligned}
$$

where $\psi^{(i)}$ is the identity function in the first argument. Note that in the above construction, each random variable $X_j$ depends on $X_i$ both directly and indirectly through $X_1, \ldots, X_{j-1}$.

Deploying an inverse Rosenblatt transform of the vector $\boldsymbol{X} = \boldsymbol{\psi}(X_i, \boldsymbol{V}),\ (\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$, we stress input factor $X_i$ and obtain

$$
\boldsymbol{X}_{i,\varepsilon} = \boldsymbol{\psi}(X_{i,\varepsilon}, \boldsymbol{V}) = \Big(\psi^{(1)}(X_{1,\varepsilon}, \boldsymbol{V}), \ldots, \psi^{(n)}(X_{1,\varepsilon}, \boldsymbol{V})\Big), \tag{4.2}
$$

and observe that the stress $X_{i,\varepsilon}$ is carried through the entire input vector, changing all factors according to their dependence on $X_{i,\varepsilon}$, resulting therefore in a cascading effect. Cascade sensitivity measures extend the marginal sensitivity framework and stress the entire input vector via (4.2), taking the dependence of the other input factors on the one being stressed into account.

**Definition 4.3.4.** For a stress $X_{i,\varepsilon}$, $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$ and a distortion risk measure $\rho_\gamma$, we define the *cascade sensitivity* to input factor $X_i$ by

$$\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) = \frac{\partial}{\partial \varepsilon} \rho_\gamma\Big(g\Big(\boldsymbol{\psi}(X_{i,\varepsilon}, \boldsymbol{V})\Big)\Big)\Big|_{\varepsilon=0},$$

whenever the derivative exists.

The set $\mathcal{R}_i$ is not a singleton, in particular the inverse Rosenblatt transform is not invariant under permutation of the order of conditioning. However, as the next result shows, the cascade sensitivity does not depend on the choice of $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$.

**Proposition 4.3.5.** If the cascade sensitivity exists for one $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$, then it exists and admits the same value for all other transforms $(\boldsymbol{\phi}, \boldsymbol{U}) \in \mathcal{R}_i$.

*Proof.* Consider a stress $X_{i,\varepsilon}$ and $(\boldsymbol{\psi}, \boldsymbol{V}), (\boldsymbol{\phi}, \boldsymbol{U}) \in \mathcal{R}_i$. Note that $\boldsymbol{V}$ and $\boldsymbol{U}$ can be chosen to be independent of the stress $X_{i,\varepsilon}$. For a function $l\colon \mathbb{R}^n \to \mathbb{R}$ such that the following expectation exists, it holds that, for all $\varepsilon > 0$,

$$\begin{aligned} E\Big((l \circ \boldsymbol{\psi})(X_{i,\varepsilon}, \boldsymbol{V})\Big) &= E\Big((l \circ \boldsymbol{\psi})(X_i, \boldsymbol{V}) \frac{f_{X_{i,\varepsilon}}(X_i)}{f_i(X_i)}\Big) \\ &= E\Big((l \circ \boldsymbol{\phi})(X_i, \boldsymbol{U}) \frac{f_{X_{i,\varepsilon}}(X_i)}{f_i(X_i)}\Big) \\ &= E\Big((l \circ \boldsymbol{\phi})(X_{i,\varepsilon}, \boldsymbol{U})\Big). \end{aligned}$$

Thus, for all $\varepsilon > 0$, $\boldsymbol{\psi}(X_{i,\varepsilon}, \boldsymbol{V})$ and $\boldsymbol{\phi}(X_{i,\varepsilon}, \boldsymbol{U})$ follow the same distribution and therefore $\frac{\partial}{\partial \varepsilon} \rho_\gamma\Big(g\Big(\boldsymbol{\psi}(X_{i,\varepsilon}, \boldsymbol{V})\Big)\Big)\Big|_{\varepsilon=0} = \frac{\partial}{\partial \varepsilon} \rho_\gamma\Big(g\Big(\boldsymbol{\phi}(X_{i,\varepsilon}, \boldsymbol{U})\Big)\Big)\Big|_{\varepsilon=0}$. □

The cascade sensitivity decomposes into the marginal sensitivity and additional components reflecting, statistical as well as functional, dependence between inputs.

**Proposition 4.3.6.** Given a stress $X_{i,\varepsilon}$, $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$, and under Assumptions 4.A.1 in the appendix, the cascade sensitivity to input factor $X_i$ is

$$\begin{aligned}\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) &= E\Big(\frac{\partial}{\partial \varepsilon} X_{i,\varepsilon}\Big|_{\varepsilon=0} (g \circ \boldsymbol{\psi})_1(X_i, \boldsymbol{V})\gamma(U_Y)\Big) \\ &= \mathcal{S}_i(\boldsymbol{X}, g, \rho_\gamma) + \sum_{j \neq i} E\Big(\frac{\partial}{\partial \varepsilon} X_{i,\varepsilon}\Big|_{\varepsilon=0} g_j(\boldsymbol{X})\psi_1^{(j)}(X_i, \boldsymbol{V})\gamma(U_Y)\Big),\end{aligned}$$

where $(g \circ \boldsymbol{\psi})_1(x_i, \boldsymbol{v}) = \frac{\partial}{\partial x_i} g(\boldsymbol{\psi}(x_i, \boldsymbol{v})) = \sum_{j=1}^n g_j(\boldsymbol{x})\psi_1^{(j)}(x_i, \boldsymbol{v})$.

*Proof.* This is a corollary of Proposition 4.3.2. □

**Remark 4.3.7.** The cascade sensitivity framework also includes sensitivity to uncertain statistical parameters of input factors (Antoniano-Villalobos et al., 2018). Let $F_{i|\Theta_i}(\cdot|\theta_i)$ denote the conditional distribution of $X_i$ given parameter $\Theta_i = \theta_i$. Then it holds almost surely that

$$X_i = F_i^{-1}(U_i \mid \Theta_i) = \eta(\Theta_i, U_i), \tag{4.3}$$

for a function $\eta\colon \mathbb{R}^2 \to \mathbb{R}$ and a standard uniform random variable $U_i$ independent of $\Theta_i$. Hence, instead of stressing the input factor $X_i$, we can perturb the parameter $\Theta_i$ and $U_i$ via representation (4.3), in this way reflecting the sensitivity of $\rho_\gamma(Y)$ to the parameter and process uncertainty of $X_i$, respectively.

### 4.3.3 Comparison of the marginal and cascade sensitivity

For input factors that are independent, the cascade sensitivity reduces to the marginal sensitivity, irrespectively of the aggregation function or the choice of risk measure. The cascade sensitivity dominates the marginal sensitivity, given positive dependence of

the input vector, a non-decreasing aggregation function and a suitable stress.

We recall that a random vector $\boldsymbol{W}$ is said to be *conditionally increasing in sequence* (CIS), if for all $j = 2, \ldots, n$, $E(l(W_j) \mid W_1 = w_1, \ldots, W_{j-1} = w_{j-1})$ is a non-decreasing function of $w_1, \ldots, w_{j-1}$, for all non-decreasing function $l$ for which the expectation exists (Müller and Stoyan, 2002).

**Proposition 4.3.8.** Let $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$ and under Assumptions 4.A.1 in the appendix, the following hold:

1. If $X_i$ is independent of $\boldsymbol{X}_{-i}$, then $\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) = \mathcal{S}_i(\boldsymbol{X}, g, \rho_\gamma)$.

2. If the vector $(X_i, X_{\pi(1)}, \ldots, X_{\pi(n)})$ is CIS for a permutation $\pi$ on $\{1, \ldots, n\}/\{i\}$, the aggregation function is component-wise non-decreasing and $\frac{\partial}{\partial \varepsilon} X_{i,\varepsilon}\Big|_{\varepsilon=0} \geq 0$ a.s., then $\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) \geq \mathcal{S}_i(\boldsymbol{X}, g, \rho_\gamma)$.

*Proof.* Case 1: an inverse Rosenblatt transform is given by $\psi^{(1)}(X_i, \boldsymbol{V}) = F_1^{-1}(V_1)$, $\psi^{(2)}(X_i, \boldsymbol{V}) = F_{2|1}^{-1}(V_2 \mid X_1)$, $\ldots$ , $\psi^{(i)}(X_i, \boldsymbol{V}) = X_i$, $\ldots$ , $\psi^{(n)}(X_i, \boldsymbol{V}) = F_{n|1,\ldots,i-1,i+1,\ldots,n-1,}^{-1}(V_{n-1} \mid X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{n-1})$. Thus, $\psi^{(j)}$, $j \neq i$, do not depend on $X_i$ and we obtain $\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) = \mathcal{S}_i(\boldsymbol{X}, g, \rho_\gamma)$ by uniqueness of the cascade sensitivity, see Proposition 4.3.5.

Case 2: let $\pi$ be a permutation on $\{1, \ldots, n\}/\{i\}$. Then $(X_i, X_{\pi(1)}, \ldots, X_{\pi(n)})$ being CIS implies that the conditional distributions $F_{\pi(j)|i,\pi(1),\ldots,\pi(j-1)}(\cdot \mid X_i = x_i, X_{\pi(1)} = x_{\pi(1)}, \ldots, X_{\pi(j-1)} = x_{\pi(j-1)})$ are non-increasing in $x_i$. Therefore the quantile functions $F^{-1}_{\pi(j)|i,\pi(1),\ldots,\pi(j-1)}(\cdot \mid X_i = x_i, X_{\pi(1)} = x_{\pi(1)}, \ldots, X_{\pi(j-1)} = x_{\pi(j-1)})$ are non-decreasing in $x_i$ and $\psi^{(j)}(X_i, \boldsymbol{V})$, $1 \leq j \leq n$, are non-decreasing functions of $X_i$ and thus $\psi_1^{(j)}(X_i, \boldsymbol{V}) \geq 0$ for $1 \leq j \leq n$. The additional assumptions guarantee that all summands of the formula of the cascade sensitivity in Proposition 4.3.6 are non-negative. $\square$

Examples of stresses with non-negative gradient include additive shocks $X_{i,\varepsilon} = X_i + \varepsilon Z$, for $Z \geq 0$ a.s.. Other examples are perturbations $X_{i,\varepsilon} = F_{i,\varepsilon}^{-1}(U_{X_i})$ with $F_{i,\varepsilon} = (1-\varepsilon)F_i + \varepsilon \hat{F}_i,\ \varepsilon \geq 0$, studied in Section 4.4.1, whenever the distribution $\hat{F}_i$ first order stochastically dominates $F_i$.

Note that by independence of the cascade sensitivity on the choice of $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$, it is enough in Proposition 4.3.8 case 2 that the vector $(X_i, X_{\pi(1)}, \ldots, X_{\pi(n)})$ is CIS for one permutation $\pi$. Examples of vectors that are CIS, which is a dependence concept of the copula alone (Müller and Scarsini, 2001, Prop. 3.5), include the multivariate normal distribution, whose inverse covariance matrix contains non-positive off-diagonal elements, the multivariate logistic, gamma and negative binomial distributions. We also refer to Karlin and Rinott (1980) for further examples of *multivariate totally positive of order 2* distributions, a slightly stronger dependence concept than CIS.

## 4.4 Calculation of the cascade sensitivity

### 4.4.1 Stressing through a perturbation

We consider a stress on the input factor $X_i$ through perturbing its distribution, such that the stressed input follows a *mixture distribution*. Then, the cascade sensitivity can be calculated in a way that circumvents the gradient of the aggregation function composed with the inverse Rosenblatt transform of the input vector, which appears in Proposition 4.3.6.

**Proposition 4.4.1.** Let $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$ and define the perturbation $X_{i,\varepsilon} = F_{i,\varepsilon}^{-1}(U_{X_i})$, where $F_{i,\varepsilon} = (1-\varepsilon)F_i + \varepsilon \hat{F}_i,\ \varepsilon \geq 0$, for a continuous distribution function $\hat{F}_i$. Under

Assumptions 4.A.1 in the appendix, the cascade sensitivity to input factor $X_i$ is

$$\begin{aligned}\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) &= E\Big[\frac{F_i(X_i) - \hat{F}_i(X_i)}{f_i(X_i)}(g \circ \boldsymbol{\psi})_1(X_i, \boldsymbol{V})\gamma(U_Y)\Big] \\ &= E\Big[\frac{H(Y) - \hat{H}(Y)}{h(Y)}\gamma(H(Y))\Big],\end{aligned}$$

where $\hat{H}$ denotes the distribution function of $\hat{Y} = g(\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{V}))$, with $\hat{X}_i = \hat{F}_i^{-1}(U_{X_i})$.

*Proof.* For all $0 < u < 1$ it holds that (Glasserman, 1991, Thm. 1.3)

$$\frac{\partial}{\partial \varepsilon} F_{i,\varepsilon}^{-1}(u)\Big|_{\varepsilon=0} = \frac{u - \hat{F}_i(F_i^{-1}(u))}{f_i(F_i^{-1}(u))}$$

and we have almost surely

$$\frac{\partial}{\partial \varepsilon} X_{i,\varepsilon}\Big|_{\varepsilon=0} = \frac{\partial}{\partial \varepsilon} F_{i,\varepsilon}^{-1}(U_{X_i})\Big|_{\varepsilon=0} = \frac{F_i(X_i) - \hat{F}_i(X_i)}{f_i(X_i)}. \tag{4.4}$$

Thus, applying Proposition 4.3.6 gives the first representation.

To see the second representation, define, for all $\varepsilon \geq 0$, the random variable $\bar{X}_{i,\varepsilon} = X_i \mathbb{1}_A + \hat{X}_i \mathbb{1}_{A^c}$, where $\hat{X}_i = \hat{F}_i^{-1}(U_{X_i})$, $A \in \mathcal{A}$ is independent of $\boldsymbol{X}$ and $\boldsymbol{V}$, with $P(A) = \varepsilon$ and $A^c = \Omega \backslash A$. Then, $\bar{X}_{i,\varepsilon}$ and the stress $X_{i,\varepsilon}$ follow the same distribution function $F_{i,\varepsilon}$. By independence of $\hat{X}_i$ and $\boldsymbol{V}$, the random vectors $\boldsymbol{\psi}(\bar{X}_{i,\varepsilon}, \boldsymbol{V})$ and $\boldsymbol{\psi}(X_{i,\varepsilon}, \boldsymbol{V})$ are equal in distribution for all $\varepsilon > 0$. Thus, the cascade sensitivity to the stress $X_{i,\varepsilon}$ is equal to the cascade sensitivity to the stress $\bar{X}_{i,\varepsilon}$. To calculate the latter, note that the stressed output, $g(\boldsymbol{\psi}(\bar{X}_{i,\varepsilon}, \boldsymbol{V})) = Y \mathbb{1}_A + g(\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{V}))\mathbb{1}_{A^c}$, follows the mixture distribution $(1 - \varepsilon)H + \varepsilon \hat{H}$, where $\hat{H}$ denotes the distribution function of $g(\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{V}))$. The representation of the cascade sensitivity to stress $\bar{X}_{i,\varepsilon}$ follows from a similar argument as in (4.4). □

We provide the representation of the cascade sensitivity for the two most common distortion risk measures in practice, VaR and ES.

**Corollary 4.4.2.** Let $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$ and define the perturbation $X_{i,\varepsilon} = F_{i,\varepsilon}^{-1}(U_{X_i})$, where $F_{i,\varepsilon} = (1-\varepsilon)F_i + \varepsilon \hat{F}_i,\ \varepsilon \geq 0$, for a continuous distribution function $\hat{F}_i$. Denote by $\hat{H}$ the distribution function of $\hat{Y} = g(\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{V}))$, with $\hat{X}_i = \hat{F}_i^{-1}(U_{X_i})$. Then the following hold:

1. Under Assumptions 4.A.1 $i)-iv)$ for $q = \alpha$ in the appendix, the cascade sensitivity to input factor $X_i$ for $\mathrm{VaR}_\alpha,\ 0 < \alpha < 1$, is

$$\mathcal{C}_i(\boldsymbol{X}, g, \mathrm{VaR}_\alpha) = \frac{\alpha - \hat{H}(H^{-1}(\alpha))}{h(H^{-1}(\alpha))}.$$

2. Under Assumptions 4.A.1 $i)-vi)$ for $q = \alpha$ in the appendix, the cascade sensitivity to input factor $X_i$ for $\mathrm{ES}_\alpha, 0 \leq \alpha < 1$, is

$$\mathcal{C}_i(\boldsymbol{X}, g, \mathrm{ES}_\alpha) = \frac{1}{1-\alpha}\Big[E\Big(\big(\hat{Y} - H^{-1}(\alpha)\big)_+\Big) - E\Big(\big(Y - H^{-1}(\alpha)\big)_+\Big)\Big].$$

Note that the second representation of the cascade sensitivity in Proposition 4.4.1 and the formulae in Corollary 4.4.2, do not require the knowledge of the gradient of the aggregation function, but, instead, of the distribution function $\hat{H}$. This distribution function can be interpreted as that of a distorted output, in particular if $\hat{F}_i$ is more disperse than $F_i$.

Rewriting the representation of the cascade sensitivity in Proposition 4.4.1, we obtain

$$\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) = \int \big(H(y) - \hat{H}(y)\big)\gamma\big(H(y)\big)\mathrm{d}y, \tag{4.5}$$

where the integral is over the support of $Y$. Thus, the cascade sensitivity can be seen as a measure of the difference of the distributions of the output $Y$ and the distorted output $\hat{Y} = g(\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{V}))$, weighted according to the choice of the risk measure $\rho_\gamma$. For the ES risk measure, for example, the difference is integrated over the right tail of the

distributions, that is over $\{y > H^{-1}(\alpha)\}$. Formula (4.5) also implies that the cascade sensitivity is robust to small changes in the weight function $\gamma$ of the distortion risk measure. In particular, the cascade sensitivities for the risk measures $\mathrm{VaR}_\alpha$ and $\mathrm{ES}_\alpha$ are robust in $\alpha$.

### 4.4.2 Stressing through an additive shock

The representation of the cascade sensitivity to stressing input factor $X_i$ through a shock $X_{i,\varepsilon} = X_i + \varepsilon Z$ is immediate from Proposition 4.3.6. The choice of the shock $Z$ is, however, not unique and should be aligned with the purpose of the sensitivity analysis. Tsanakas and Millossovich (2016) argue that the only meaningful shocks are functions of $X_i$. We consider shocks comonotonic to input $X_i$, specifically, $X_{i,\varepsilon} = X_i + \varepsilon k(X_i)$, for non-decreasing Lipschitz continuous functions $k\colon \mathbb{R} \to \mathbb{R}$ with Lipschitz constant 1. Moreover, in this section we make further (non-restrictive) regularity assumptions on the input $X_i$ as stated in Lemma 4.A.2. This choice of shock leads to a representation of the cascade sensitivity, similar to that of Proposition 4.4.1, without involving the aggregation function's gradient.

Examples of additive shocks as above include stressing the tails of an input factor: $X_{i,\varepsilon} = X_i + \varepsilon(X_i - t_1)\mathbb{1}_{\{X_i \le t_1\}} + \varepsilon(X_i - t_2)\mathbb{1}_{\{X_i \ge t_2\}}$, for $t_1 \le t_2 \in \mathbb{R}$. Alternatively, for a unimodal input factor $X_i$ with mode $m$, one may define $X_{i,\varepsilon} = X_i + \varepsilon(X_i - m)$. Scaling deviations from the mode preserves, to some extent, the shape of the distribution of $X_i$. For example, if input factor $X_i \sim N(\mu, \sigma^2)$, then $X_{i,\varepsilon} \sim N(\mu, (1+\varepsilon)^2\sigma^2)$, while for an exponentially distributed input factor, $X_i \sim Exp(\lambda)$, the stressed input follows $X_{i,\varepsilon} \sim Exp(\frac{\lambda}{1+\varepsilon})$.

**Proposition 4.4.3.** Let $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$ and define the stress $X_{i,\varepsilon} = X_i + \varepsilon k(X_i)$, for a non-decreasing Lipschitz continuous function $k\colon \mathbb{R} \to \mathbb{R}$ with Lipschitz constant 1, that satisfies $k(x) \le 0$ on the set where $f_i(x)$ is non-decreasing and $k(x) \ge 0$ on the set

where $f_i(x)$ is non-increasing. Under Assumptions 4.A.1 in the appendix, the cascade sensitivity to input factor $X_i$ is

$$\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) = E\Big[\frac{H(Y) - \tilde{H}(Y)}{h(Y)}\gamma(H(Y))\Big],$$

where $\tilde{H}$ denotes the distribution function of $\tilde{Y} = g(\boldsymbol{\psi}(\tilde{X}_i, \boldsymbol{V}))$, with $\tilde{X}_i = \tilde{F}_i^{-1}(U_{X_i})$ and distribution function $\tilde{F}_i(x) = F_i(x) - k(x)f_i(x),\ x \in \mathbb{R}$.

*Proof.* By Lemma 4.A.2, $\tilde{F}_i$ is indeed a distribution function and we denote its density by $\tilde{f}_i$. Define $\xi(y) = \gamma(H(y))$, $y \in \mathbb{R}$. Using independence of $X_i$ and $\boldsymbol{V}$ and the definition of $\tilde{F}_i$, the cascade sensitivity to stress $X_{i,\varepsilon}$ can be written as

$$\begin{aligned}
\mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) &= E(k(X_i)(g \circ \boldsymbol{\psi})_1(X_i, \boldsymbol{V})\xi(Y)) \\
&= E\Big(\int_{\mathbb{R}} k(x)(g \circ \boldsymbol{\psi})_1(x, \boldsymbol{V})\xi(g(\boldsymbol{\psi}(x, \boldsymbol{V})))f_i(x)\mathrm{d}x\Big) \\
&= E\Big(\int_{\mathbb{R}} (F_i(x) - \tilde{F}_i(x))\beta(x)\mathrm{d}x\Big),
\end{aligned}$$

where we denote $\beta(x) = (g \circ \boldsymbol{\psi})_1(x, \boldsymbol{V})\xi(g(\boldsymbol{\psi}(x, \boldsymbol{V})))$ and $B(s) = \int_s^{\operatorname{ess\,sup} X_i} \beta(x)\mathrm{d}x$ hence suppressing the dependence on $\boldsymbol{V}$. Applying Fubini, we obtain

$$\begin{aligned}
E\Big(\int_{\mathbb{R}} (F_i(x) - \tilde{F}_i(x))\beta(x)\mathrm{d}x\Big) &= E\Big(\int_{\mathbb{R}}\int_{\mathbb{R}} \big(f_i(s) - \tilde{f}_i(s)\big)\mathbb{1}_{\{s \le x\}}\mathrm{d}s\beta(x)\mathrm{d}x\Big) \\
&= E\Big(\int_{\mathbb{R}} \big(f_i(s) - \tilde{f}_i(s)\big)B(s)\mathrm{d}s\Big) \\
&= E(B(X_i)) - E(B(\tilde{X}_i)).
\end{aligned}$$

Applying the change of variable $u = g(\boldsymbol{\psi}(t, \boldsymbol{V}))$ we obtain

$$\begin{aligned} B(X_i) &= \int_{X_i}^{\operatorname{ess\,sup} X_i} \beta(t)\mathrm{d}t \\ &= \int_{X_i}^{\operatorname{ess\,sup} X_i} (g \circ \boldsymbol{\psi})_1(t, \boldsymbol{V})\xi(g(\boldsymbol{\psi}(t, \boldsymbol{V})))\mathrm{d}t \\ &= \int_{Y}^{\operatorname{ess\,sup} Y} \xi(u)\mathrm{d}u, \end{aligned}$$

and similarly, $B(\tilde{X}_i) = \int_{\tilde{Y}}^{\operatorname{ess\,sup} \tilde{Y}} \xi(u)\mathrm{d}u$. Thus, using Lemma 4.A.3 the cascade sensitivity becomes

$$\begin{aligned} \mathcal{C}_i(\boldsymbol{X}, g, \rho_\gamma) &= \int_{\mathbb{R}} (1 - \tilde{H}(y))\xi(y)\mathrm{d}y - \int_{\mathbb{R}} (1 - H(y))\xi(y)\mathrm{d}y \\ &= \int_{\mathbb{R}} (H(y) - \tilde{H}(y))\gamma(H(y))\mathrm{d}y \\ &= E\Big(\frac{H(Y) - \tilde{H}(Y)}{h(Y)}\gamma(H(Y))\Big). \end{aligned}$$

□

The cascade sensitivity for the risk measures VaR and ES when stressing an input factor via $X_{i,\varepsilon} = X_i + \varepsilon k(X_i)$, for a function $k$ fulfilling the assumptions in Proposition 4.4.3, are analogous to Corollary 4.4.2, replacing $\hat{H}$ with $\tilde{H}$.

The representation of the cascade sensitivities in Proposition 4.4.3 involves a weighted difference of the distribution functions $\tilde{H}$ and $H$, similarly to Proposition 4.4.1. Thus, calculating the sensitivity to a shock $X_{i,\varepsilon} = X_i + \varepsilon k(X_i)$, corresponds to comparing the output $Y$ with the distorted output $\tilde{Y} = g\Big(\boldsymbol{\psi}(\tilde{F}_i^{-1}(U_{X_i}), \boldsymbol{V})\Big)$, where $\tilde{F}_i(x) = F_i(x) - k(x)f_i(x),\ x \in \mathbb{R}$. The next proposition shows that, provided that $E(k(X_i)) \geq 0$, $\tilde{F}_i^{-1}(U_{X_i})$ dominates $X_i$ in *increasing convex order*, so that the distorted output $\tilde{Y}$ could be seen as more conservative than $Y$, when $g$ is non-decreasing and $\boldsymbol{X}$ positively dependent. Recall that a random variable $W$ dominates $Z$ in increasing

convex order, $Z \preceq_{icx} W$, if $E(l(Z)) \leq E(l(W))$, for all increasing convex functions $l\colon \mathbb{R} \to \mathbb{R}$ such that the expectations exist (Müller and Stoyan, 2002).

**Proposition 4.4.4.** Let $X_i$ have finite expectation and define the random variable $\tilde{X}_i$ with distribution function $\tilde{F}_i(x) = F_i(x) - k(x)f_i(x),\ x \in \mathbb{R}$, as in Proposition 4.4.3. Then the following hold:

1. If $E(k(X_i)) \geq 0$, then $X_i \preceq_{icx} \tilde{X}_i$.

2. If $0 < \operatorname{ess\,sup} k(X_i)$, then $X_i$ does not dominate $\tilde{X}_i$ in increasing convex order.

*Proof.* Applying Lemma 4.A.2 we see that $\tilde{F}_i$ is a distribution function. Note that $\tilde{X}_i$ dominates $X_i$ in increasing convex order, $X_i \preceq_{icx} \tilde{X}_i$, if and only if $E((\tilde{X}_i - t)_+) \geq E((X_i - t)_+)$ for all $t \in \mathbb{R}$. For case 1: let $t \in \mathbb{R}$ and apply Fubini,

$$\begin{aligned}
E\Big((\tilde{X}_i - t)_+\Big) - E\Big((X_i - t)_+\Big) &= \int_t^\infty (x-t)\Big(\tilde{f}_i(x) - f_i(x)\Big)\mathrm{d}x \\
&= \int_t^\infty \int_t^\infty \Big(\tilde{f}_i(x) - f_i(x)\Big)\mathbb{1}_{\{u \leq x\}}\mathrm{d}x\mathrm{d}u \\
&= \int_t^\infty \Big(F_i(u) - \tilde{F}_i(u)\Big)\mathrm{d}u \\
&= \int_t^\infty k(u)f_i(u)\mathrm{d}(u) \\
&= E(k(X_i)\mathbb{1}_{\{X_i > t\}}).
\end{aligned}$$

Recall that $k$ is a non-decreasing function, thus if $k(t) \geq 0$, the above expectation is non-negative. If $k(t) < 0$, we have that $E(k(X_i)\mathbb{1}_{\{X_i > t\}}) \geq E(k(X_i))$, which is non-negative by assumption.

To see case 2: note that $E(k(X_i)\mathbb{1}_{\{X_i > t\}})$ is negative if and only if

$$E(k(X_i) \mid X_i > t) < 0 \text{ for all } t \in \mathbb{R},$$

which is a contradiction to the assumption that $0 < \operatorname{ess\,sup} k(X_i)$. □

Consider an input factor that is symmetric around zero and the stress $X_{i,\varepsilon} = X_i + \varepsilon(X_i - t_1)\mathbb{1}_{\{X_i \le t_1\}} + \varepsilon(X_i - t_2)\mathbb{1}_{\{X_i \ge t_2\}}$ for $t_1 < 0 < t_2$, such that the density of the input is non-decreasing on $\{x \le t_1\}$ and non-increasing on $\{x \ge t_2\}$. Then, Proposition 4.4.4 case 1, is fulfilled if $t_2 < |t_1|$. For a one-sided stress of an input factor, that is $X_{i,\varepsilon} = X_i + \varepsilon(X_i - t)_+,\ t > 0$, Proposition 4.4.4 case 1 is always satisfied. For an unimodal input factor $X_i$ with mode $m$ and stress $X_{i,\varepsilon} = X_i + \varepsilon(X_i - m)$, Proposition 4.4.4, case 1 is satisfied if $E(X_i) \ge m$, and case 2 holds if $m < \operatorname{ess\,sup} X_i$.

### 4.4.3 Numerical evaluation via importance sampling

In practical applications, when the marginals and the copula of the input vector are separately specified and estimated, the calculation of an inverse Rosenblatt transform of $\boldsymbol{X}$ may be available. For example the **R** package *copula* (Hofert et al., 2017) provides the inverse conditional distribution functions for Archimedean and elliptical copulas. An algorithm for the computation of the inverse Rosenblatt transform of canonical and D-vine copulas is presented in Aas et al. (2009, Algorithm 5 and 6), and for R-vine copulas in Schepsmeier (2015). So that $\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{U})$ and $\boldsymbol{\psi}(\tilde{X}_i, \boldsymbol{U})$ in Propositions 4.4.1 and 4.4.3, respectively, can explicitly be calculated.

A computationally expensive aggregation function, however, might render a direct calculation of the cascade sensitivity unfeasible, as Propositions 4.4.1 and 4.4.3 require the evaluation of $\hat{Y} = g\big(\boldsymbol{\psi}(\hat{X}_i, \boldsymbol{V})\big)$ and $\tilde{Y} = g\big(\boldsymbol{\psi}(\tilde{X}_i, \boldsymbol{V})\big)$, respectively. For example, in a Monte Carlo simulation setting with sample size $M$, the calculation of the cascade sensitivity to one input factor requires an inverse Rosenblatt transform and $M$ evaluations of $\hat{Y}$ or $\tilde{Y}$. Fortunately, using importance sampling, the distribution functions $\hat{H}, \tilde{H}$ can be computed on the same Monte Carlo sample without the need

to explicitly calculate an inverse Rosenblatt transform. It holds that, for $t \in \mathbb{R}$,

$$\hat{H}(t) = E\Big(\mathbb{1}_{\{g(\boldsymbol{\psi}(\hat{X}_i,\boldsymbol{V}))\leq y\}}\Big) = E\Big(\mathbb{1}_{\{Y\leq t\}}\frac{\hat{f}_i(X_i)}{f_i(X_i)}\Big), \tag{4.6}$$

where $\hat{f}_i$ denotes the density of $\hat{F}_i$ and $\frac{\hat{f}_i(X_i)}{f_i(X_i)}$ play the role of *importance weights*. Note that, due to independence of $\hat{X}_i$ and $\boldsymbol{V}$, the importance weights are a function of $X_i$ only. An analogous formula holds for $\tilde{H}$. Thus, starting with a Monte Carlo sample of the input vector $\boldsymbol{X}$ and the knowledge of the density $\hat{f}_i$ or $\tilde{f}_i$, the calculation of the cascade sensitivity is straightforward without the need to calculate an inverse Rosenblatt transform of $\boldsymbol{X}$. Specifically, the cascade sensitivity in Propositions 4.4.1 (4.4.3), can be estimated through the following procedure

1. Sample $M$ multivariate scenarios $\boldsymbol{x}^1 = (x_1^1, \dots x_n^1), \dots, \boldsymbol{x}^M = (x_1^M, \dots x_n^M)$ from input vector $\boldsymbol{X}$ and calculate the corresponding realisations of the output $y^1 = g(\boldsymbol{x}^1), \dots, y^M = g(\boldsymbol{x}^M)$.

2. Estimate the distribution function and density of the output $Y$, for example through the empirical distribution function and a kernel density estimator

$$H^{emp}(t) = \frac{1}{M}\sum_{j=1}^{M}\mathbb{1}_{\{y^j\leq t\}},\ t\in\mathbb{R},$$
$$h^{emp}(t) = \frac{1}{M}\sum_{j=1}^{M}\kappa(t-y^j),\ t\in\mathbb{R},$$

   for a suitable kernel $\kappa$.

3. Estimate the cascade sensitivity by

$$\mathcal{C}_i^{emp} = \frac{1}{M}\sum_{j=1}^{M}\frac{\gamma(H^{emp}(y^j))}{h^{emp}(y^j)}\left(H^{emp}(y^j) - \frac{1}{M}\sum_{k=1}^{M}\frac{\hat{f}_i(x_i^k)}{f_i(x_i^k)}\mathbb{1}_{\{y^k\leq y^j\}}\right).$$

Alternatively, representation (4.5) of the cascade sensitivity allows for a calculation without the need to estimate $h$.

## 4.5 Application to a non-linear insurance portfolio

### 4.5.1 The insurance portfolio

The marginal and the cascade sensitivity measures are illustrated on the following insurance portfolio. An insurance company has two lines of business $X_1, X_2$ that are subject to the same multiplicative factor $X_3$. On its loss, $L = X_3(X_1 + X_2)$, the insurance company has a reinsurance contract with deductible $d = 380$ and limit $l = 30$. The insurance portfolio includes another line of business $X_4$, also subject to the factor $X_3$. Thus, the total loss faced by the insurer is

$$Y = L - \min\{(L - d)_+, l\} + X_3 X_4.$$

The two lines of business, $X_1, X_2$ follow a Log-Normal($\mu = 4.98,\ \sigma^2 = 0.23^2$) distribution with mean 150 and standard deviation 35. $X_3$ is Log-Normal($\mu_3 = -0.005,\ \sigma_3^2 = 0.1^2$) distributed with mean 1 and standard deviation 0.1. The line of business $X_4$ is Gamma(100,1) distributed with mean 100 and standard deviation 10 and is independent of $X_1, X_2$ and $X_3$. The sub-vector $\boldsymbol{X}_{-4} = (X_1, X_2, X_3)$ is dependent through a Gaussian copula with correlation matrix

$$R = \begin{pmatrix} 1 & 0.3 & 0.8 \\ 0.3 & 1 & 0 \\ 0.8 & 0 & 1 \end{pmatrix}.$$

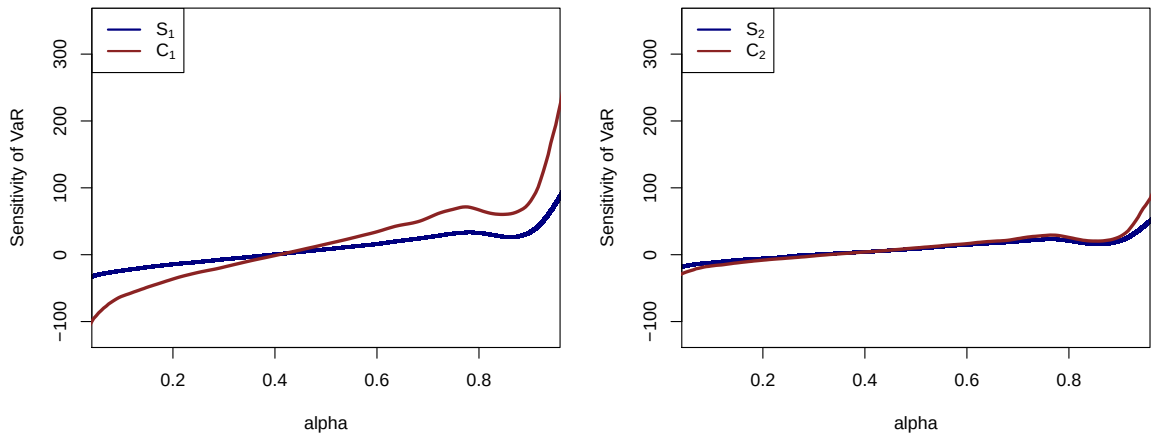


Figure 4.1: Marginal and cascade sensitivity of the risk measure $\mathrm{VaR}_\alpha$, $0.075 < \alpha < 0.925$, for input factor $X_1$ (left) and input factor $X_2$ (right), for stress $X_{i,\varepsilon} = X_i + \varepsilon(X_i - m)$, $i = 1, 2$.

All calculations in this section are based on a simulated Monte Carlo sample of size $M = 100,000$. We consider the marginal and cascade sensitivity to input factors $X_1$ and $X_2$ for the risk measures VaR and ES. Both input factors $X_1$ and $X_2$ are stressed by applying the shock $X_{i,\varepsilon} = X_i + \varepsilon(X_i - m)$, where $m = 138.5$ is the mode of either. Note that $X_1$ and $X_2$ follow the same (unimodal) distribution and the aggregation function is symmetric in $X_1, X_2$, but the dependence structure is not; $X_1$ is highly correlated to $X_3$, while $X_2$ is independent of $X_3$.

The effect of the asymmetric dependence structure of the input factors on the marginal and the cascade sensitivities respectively, is shown in Figures 4.1 and 4.2. Figure 4.1 displays the marginal and cascade sensitivity of the risk measure $\mathrm{VaR}_\alpha$, $0.075 < \alpha < 0.925$, to both input factors $X_1$ and $X_2$. The marginal and cascade sensitivities to $X_1$ and $X_2$ for the risk measure $\mathrm{ES}_\alpha$, $0.075 < \alpha < 0.925$, are plotted in Figure 4.2. For $\alpha \geq 0.5$, in Figures 4.1 and 4.2, the cascade sensitivity to input factor $X_1$, $\mathcal{C}_1$, is seen to be substantially larger than the marginal sensitivity, indicating that

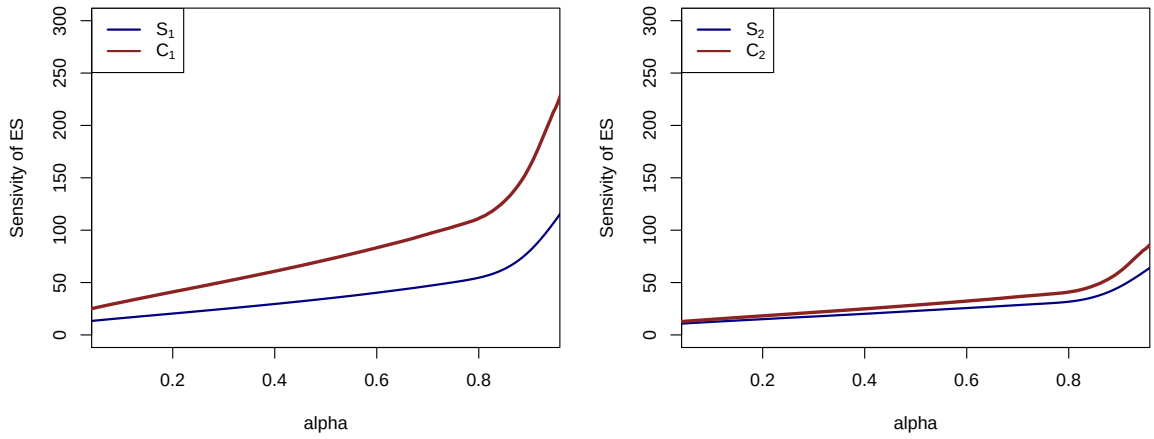


Figure 4.2: Marginal and cascade sensitivity of the risk measure $\mathrm{ES}_\alpha$, $0.075 < \alpha < 0.925$, for input factor $X_1$ (left) and input factor $X_2$ (right) for stress $X_{i,\varepsilon} = X_i + \varepsilon(X_i - m_i)$, $i = 1, 2$.

the marginal sensitivity does not reflect the indirect effects of the dependence between risks factors.

For the $\mathrm{ES}_\alpha$ risk measure, the marginal and the cascade sensitivity to input $X_1$ exceed, for all severity levels $\alpha$, the respective sensitivity measures for $X_2$. Thus the marginal and the cascade sensitivity for $\mathrm{ES}_\alpha$ produce consistent ranking of the input factors. The $\mathrm{VaR}_\alpha$ risk measure, however, does not provide a consistent ranking between the marginal and the cascade sensitivity to input factors $X_1$ and $X_2$, see also Section 4.5.2 for further discussion.

A criticism of sensitivity measures defined as partial derivatives is, that for input factors on different scales, no conclusion can be drawn regarding their relative importance (Antoniano-Villalobos et al., 2018). In the insurance portfolio example, inputs $X_1$ and $X_2$ follow the same distribution and thus are stressed in exactly the same fashion, allowing for a direct comparison of $\mathcal{S}_1, \mathcal{S}_2$ and $\mathcal{C}_1, \mathcal{C}_2$. In Section 4.6 we provide a framework, that allows to consistently stress input factors and which is independent of the distributional assumptions of the input vector.

### 4.5.2 Cascade sensitivity via importance sampling

In this section we provide the analytical formulae of the marginal and the cascade sensitivity. First, note that the gradient of the aggregation function is given by

$$\Big(X_3\mathbb{1}_{\{L<d\}\cup\{L>l+d\}},\ X_3\mathbb{1}_{\{L<d\}\cup\{L>l+d\}},\ (X_1+X_2)\mathbb{1}_{\{L<d\}\cup\{L>l+d\}}+X_4,\ X_3\Big).$$

The marginal sensitivities for the risk measures $\mathrm{VaR}_\alpha$ and $\mathrm{ES}_\alpha$ to input factors $X_i,\ i=1,2$, following Proposition 4.3.2, are thus

$$\begin{aligned}\mathcal{S}_i(\boldsymbol{X},g,\mathrm{VaR}_\alpha) &= E\Big((X_i-m)X_3\mathbb{1}_{\{L<d\}\cup\{L>l+d\}}\mathbb{1}_{\{Y=H^{-1}(\alpha)\}}\Big),\\ \mathcal{S}_i(\boldsymbol{X},g,\mathrm{ES}_\alpha) &= \frac{1}{1-\alpha}E\Big((X_i-m)X_3\mathbb{1}_{\{L<d\}\cup\{L>l+d\}}\mathbb{1}_{\{Y>H^{-1}(\alpha)\}}\Big).\end{aligned}$$

To calculate the cascade sensitivity, we apply Proposition 4.4.3, using the importance sampling routine of Section 4.4.3. The cascade sensitivities for the risk measures $\mathrm{VaR}_\alpha$ and $\mathrm{ES}_\alpha$ to input factors $X_i,\ i=1,2$, are given by

$$\mathcal{C}_i(\boldsymbol{X},g,\mathrm{VaR}_\alpha) = \frac{1}{h(H^{-1}(\alpha))}\Big[\alpha - E\Big(\frac{\tilde{f}_i(X_i)}{f_i(X_i)}\mathbb{1}_{\{Y\le H^{-1}(\alpha)\}}\Big)\Big], \tag{4.7a}$$

$$\mathcal{C}_i(\boldsymbol{X},g,\mathrm{ES}_\alpha) = \frac{1}{1-\alpha}\Big[E\Big(\frac{\tilde{f}_i(X_i)}{f_i(X_i)}\big(Y-H^{-1}(\alpha)\big)_+\Big) - E\Big(\big(Y-H^{-1}(\alpha)\big)_+\Big)\Big]. \tag{4.7b}$$

The distorted densities $\tilde{f}_i,\ i=1,2$, are equal and given by

$$\tilde{f}_i(x) = \frac{x-m}{x}\Big(1+\frac{\ln(x)-\mu}{\sigma^2}\Big)f_i(x),\ x>0. \tag{4.8}$$

Figure 4.3 displays the densities $f_1, \tilde{f}_1$ of input factor $X_1$ and its distortion $\tilde{X}_1$, respectively. As seen in the plot, the density $\tilde{f}_1$ places more emphasis on the tails of the input factor, separating the probability mass away from $m$.

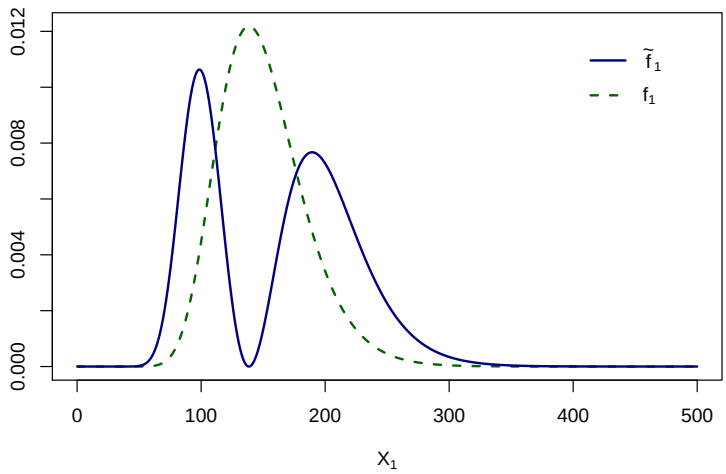


Figure 4.3: Densities $f_1, \tilde{f}_1$, defined in equation (4.8), for the input factor $X_1$ and the distortion $\tilde{X}_1$, respectively.

The left plot in Figure 4.4 shows the importance weights $\frac{\tilde{f}_1(X_1)}{f_1(X_1)}$ against the input factor $X_1$. The importance weights are zero at the mode of $X_1$ and give more weight to high and low realisations of $X_1$. The right plot of Figure 4.4 depicts the importance weights against the output $Y$, indicating the positive dependence between $X_1$ and $Y$. Note that the importance weights inflate both high and low values of the output.

The empirical distribution functions $H^{emp}$ of the output $Y$ and $\tilde{H}^{emp}$ of the distorted output $\tilde{Y} = g\big(\boldsymbol{\psi}(\tilde{X}_1, \boldsymbol{V})\big)$ are displayed in Figure 4.5. The crossing of the distribution functions is due to the importance weights accentuating both high and low realisations of $Y$. Figure 4.5 enables a visual interpretation of the cascade sensitivity, as the integrated difference of $\tilde{H}$ and $H$, weighted according to the risk measure, see equation (4.5). Thus, it illustrates why the cascade sensitivity of the $\mathrm{VaR}_\alpha$, Figure 4.1, is negative for $\alpha$ fulfilling $H(\alpha) < \tilde{H}(\alpha)$. The cascade sensitivity for $\mathrm{ES}_\alpha$, Figure 4.2, on the other hand is positive for all $0 \leq \alpha < 1$, since the difference of the distribution functions is integrated over $\{y > H^{-1}(\alpha)\}$.

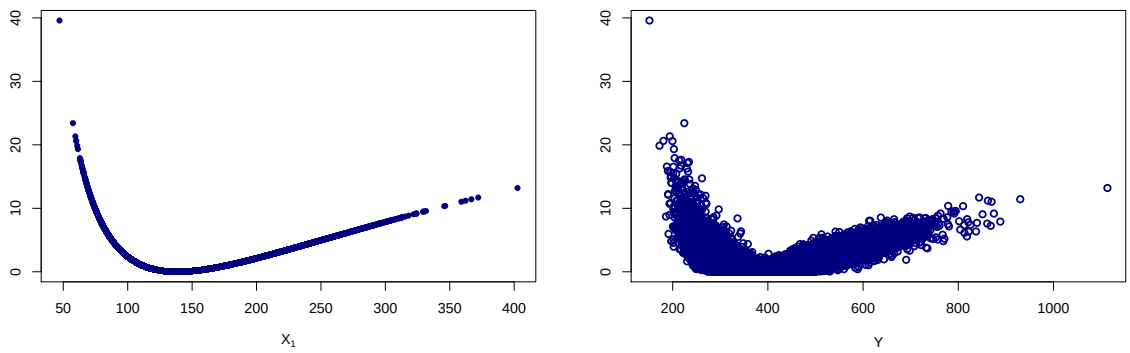


Figure 4.4: Importance weights $\frac{\tilde{f}_1(X_1)}{f_1(X_1)}$ against input factor $X_1$ (left) and output $Y$ (right), used for the calculation of cascade sensitivity by importance sampling.

### 4.5.3 Cascade sensitivity via the inverse Rosenblatt transform

We continue with the discussion of the insurance portfolio example to provide further insight by calculating the cascade sensitivity via an inverse Rosenblatt transform. Note that the representation of the cascade sensitivity in Proposition 4.4.3 does not require an inverse Rosenblatt transform of the input vector. However, due to the Gaussian copula structure, an inverse Rosenblatt transform is easily obtained in analytical form. Recall that $\boldsymbol{X} = (\boldsymbol{X}_{-4}, X_4)$, where $\boldsymbol{X}_{-4}$ is independent of $X_4$, has Log-Normal marginals and a Gaussian copula with correlation matrix $R$. An inverse Rosenblatt transform of the input vector can be obtained through the Cholesky decomposition of $R = DD^\top$, where

$$D = \left(d_{i,j}\right)_{1\leq i,j,\leq 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0.3 & 0.95 & 0 \\ 0.8 & -0.25 & 0.54 \end{pmatrix}.$$

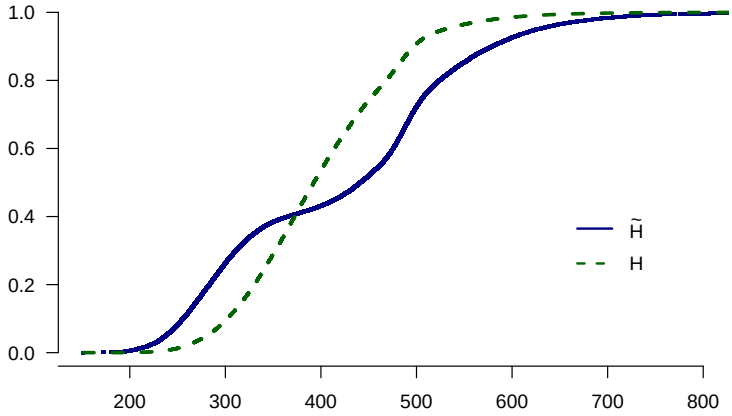


Figure 4.5: Empirical distribution function of the output $Y$, $H^{emp}$, and the empirical distribution function of the distorted output $\tilde{Y} = g(\boldsymbol{\psi}(\tilde{X}_1, \boldsymbol{V}))$, $\tilde{H}^{emp}$, calculated using importance sampling, (4.6).

Thus, we can write

$$
\begin{aligned}
X_1 &= \exp\{\mu + \sigma Z\}, \\
X_2 &= \exp\Big\{\mu + \sigma\Big[d_{2,1}Z + d_{2,2}\ \Phi^{-1}(V_1)\Big]\Big\}, \\
X_3 &= \exp\Big\{\mu_3 + \sigma_3\ \Big[d_{3,1}Z + d_{3,2}\ \Phi^{-1}(V_1) + d_{3,3}\ \Phi^{-1}(V_2)\Big]\Big\}, \\
X_4 &= F_4^{-1}(V_3),
\end{aligned}
$$

for a standard normal random variable $Z$ with distribution function $\Phi$ and a vector of independent standard uniform random variables $\boldsymbol{V} = (V_1, V_2, V_3)$. Hence, an inverse

Rosenblatt transform of $\boldsymbol{X} = \boldsymbol{\psi}(X_1, \boldsymbol{V}),\ (\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_1$, is given by

$$
\begin{aligned}
&\psi^{(1)}(X_1, \boldsymbol{V}) = X_1,\\
&\psi^{(2)}(X_1, \boldsymbol{V}) = X_1^{d_{2,1}} \exp\{\mu(1 - d_{2,1}) + \sigma d_{2,2}\ \Phi^{-1}(V_1)\},\\
&\psi^{(3)}(X_1, \boldsymbol{V}) = X_1^{d_{3,1}\frac{\sigma_3}{\sigma}} \exp\Big\{\mu_3 - \sigma_3\ \Big[\mu\frac{d_{3,1}}{\sigma} + d_{3,2}\ \Phi^{-1}(V_1) + d_{3,3}\ \Phi^{-1}(V_2)\Big]\Big\},\\
&\psi^{(4)}(X_1, \boldsymbol{V}) = F_4^{-1}(V_3).
\end{aligned}
$$

To calculate the cascade sensitivity to input factor $X_2$, we need an inverse Rosenblatt transform of $\boldsymbol{X}$ starting with input factor $X_2$. First, we reorder the input vector to $(X_2, X_1, X_3, X_4)$, exchanging input factors $X_1$ and $X_2$. The corresponding correlation matrix and the Cholesky decomposition of the reordered vector $\boldsymbol{X}_{-4}$ are given by

$$
R^* = \begin{pmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0.8 \\ 0 & 0.8 & 1 \end{pmatrix}, \quad D^* = \left(d^*_{i,j}\right)_{1\le i,j\le 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0.3 & 0.95 & 0 \\ 0 & 0.84 & 0.54 \end{pmatrix},
$$

with $R^* = D^* D^{*\top}$. Analogously to the above construction, an inverse Rosenblatt transform of $\boldsymbol{X} = \boldsymbol{\phi}(X_2, \boldsymbol{U})$, for $(\boldsymbol{\phi}, \boldsymbol{U}) \in \mathcal{R}_2$, is given by

$$
\begin{aligned}
&\phi^{(1)}(X_2, \boldsymbol{U}) = X_2^{d^*_{2,1}} \exp\{\mu(1 - d^*_{2,1}) + \sigma d^*_{2,2}\ \Phi^{-1}(U_1)\},\\
&\phi^{(2)}(X_2, \boldsymbol{U}) = X_2,\\
&\phi^{(3)}(X_2, \boldsymbol{U}) = \exp\Big\{\mu_3 + \sigma_3\Big[d^*_{3,2}\ \Phi^{-1}(U_1) + d^*_{3,3}\ \Phi^{-1}(U_2)\Big]\Big\},\\
&\phi^{(4)}(X_2, \boldsymbol{U}) = F_4^{-1}(U_3),
\end{aligned}
$$

where we write $\boldsymbol{U} = (U_1, U_2, U_3)$. Note that we used above that $d^*_{3,1} = \mathrm{Corr}(X_2, X_3) = 0$.

The partial derivatives of both $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_1$ and $(\boldsymbol{\phi}, \boldsymbol{U}) \in \mathcal{R}_2$, are displayed in Table 4.1. Note that $d_{2,1} = d^*_{2,1} = \mathrm{Corr}(X_1, X_2) = 0.3$, thus the terms reflecting the indirect effects of the dependence between the input factors are scaled according to their correlation with the stressed input factor.

Table 4.1: Derivative of the inverse Rosenblatt transforms of input vector $\boldsymbol{X}$, $(\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_1$ and $(\boldsymbol{\phi}, \boldsymbol{U}) \in \mathcal{R}_2$. Note that in $\psi_1^{(j)}$ the derivative is taken with respect to $X_1$ while in $\phi_1^{(j)}$ it is taken with respect to $X_2$.

| | $j=1$ | $j=2$ | $j=3$ | $j=4$ |
|---|---|---|---|---|
| $\psi_1^{(j)}$ | $1$ | $\mathrm{Corr}(X_1, X_2)\dfrac{X_2}{X_1}$ | $\mathrm{Corr}(X_1, X_3)\dfrac{\sigma_3}{\sigma}\dfrac{X_3}{X_1}$ | $0$ |
| $\phi_1^{(j)}$ | $\mathrm{Corr}(X_1, X_2)\dfrac{X_1}{X_2}$ | $1$ | $0$ | $0$ |

The cascade sensitivities for the $\mathrm{VaR}_\alpha$ risk measure to input factors $X_1$ and $X_2$ are given by Proposition 4.3.6:

$$\begin{aligned}
\mathcal{C}_1 &= \mathcal{S}_1 + \mathrm{Corr}(X_1, X_2) E\Big(\frac{X_2}{X_1}(X_1 - m) X_3 \mathbb{1}_{\{L<d\}\cup\{L>l+d\}} \mathbb{1}_{\{Y=H^{-1}(\alpha)\}}\Big) \\
&\quad + \frac{\sigma_3}{\sigma}\mathrm{Corr}(X_1, X_3) E\Big(\frac{X_3}{X_1}(X_1 - m)\Big((X_1 + X_2)\mathbb{1}_{\{L<d\}\cup\{L>l+d\}} + X_4\Big)\mathbb{1}_{\{Y=H^{-1}(\alpha)\}}\Big), \\
\mathcal{C}_2 &= \mathcal{S}_2 + \mathrm{Corr}(X_1, X_2) E\Big(\frac{X_1}{X_2}(X_2 - m) X_3 \mathbb{1}_{\{L<d\}\cup\{L>l+d\}} \mathbb{1}_{\{Y=H^{-1}(\alpha)\}}\Big).
\end{aligned}$$

The cascade sensitivity $\mathcal{C}_1$ includes the terms accounting for the dependence between $X_1, X_2$ and $X_1, X_3$, while $\mathcal{C}_2$ only incorporates the term reflecting dependence between $X_1, X_2$, explaining the larger (absolute) values of $\mathcal{C}_1$.

### 4.5.4 Influence of indirect dependence between input factors

We illustrate the indirect effects of the dependence structure of the input vector on the marginal and cascade sensitivity, by considering different correlations between input factors $X_1$ and $X_3$. The marginal and cascade sensitivities to $X_1$ and $X_2$ for the

risk measure $\mathrm{ES}_{0.9}$ and for $\mathrm{Corr}(X_1, X_3) = 0.0, ..., 0.8$, are calculated using (4.7) and reported in Table 4.2. In the last two columns of Table 4.2, we state $\frac{\mathcal{C}_j - \mathcal{S}_j}{\mathcal{C}_j}$, $j = 1, 2$, that is, the percentage of the cascade sensitivity that stems from the indirect effects of the dependence between the input factors.

While the absolute values of the marginal and the cascade sensitivity measures cannot be compared in a direct manner, their relative change for different correlation coefficients provides some insight. As seen in Table 4.2, both the marginal and the cascade sensitivity to input factor $X_1$ increase with the correlation between $X_1$ and $X_3$, where the extent of change is more substantial in $\mathcal{C}_1$ than in $\mathcal{S}_1$. The marginal sensitivity to input factor $X_2$ is constant, implying that $\mathcal{S}_2$ does not account for changes of the dependence between $X_1$ and $X_3$. This is in contrast to the cascade sensitivity to input factor $X_2$, which increases by 5%, thus indicating that the correlation between $X_1$ and $X_3$ plays a role in the cascade sensitivity to input factor $X_2$.

Table 4.2: Marginal and cascade sensitivity for the $\mathrm{ES}_{0.9}$ with $\mathrm{Corr}(X_1, X_3) = 0.0, ..., 0.8$.

| $\mathrm{Corr}(X_1, X_3)$ | $\mathcal{S}_1$ | $\mathcal{S}_2$ | $\mathcal{C}_1$ | $\mathcal{C}_2$ | $\frac{\mathcal{C}_1 - \mathcal{S}_1}{\mathcal{C}_1}$ | $\frac{\mathcal{C}_2 - \mathcal{S}_2}{\mathcal{C}_2}$ |
|---|---|---|---|---|---|---|
| 0.0 | 46 | 45 | 55 | 55 | 0.16 | 0.18 |
| 0.2 | 54 | 45 | 76 | 55 | 0.29 | 0.18 |
| 0.4 | 63 | 45 | 101 | 56 | 0.38 | 0.20 |
| 0.6 | 71 | 45 | 129 | 57 | 0.45 | 0.21 |
| 0.8 | 80 | 45 | 162 | 58 | 0.50 | 0.22 |

## 4.6 Application to a black box model

This section illustrates the applicability of the cascade sensitivity measure to a black box model: a setting where an analyst has no access to the full specification of either $\boldsymbol{F}$ or $g$, but only to a set of input / output simulations. Such a situation is common in insurance applications. The model represents a London Insurance Market portfolio

loss of 72 input factors, measured on the same scale, and is currently in use by a participant in the market. The analysis is based on a Monte Carlo sample of size $M = 500,000$, provided to us by the model owner, consisting of simulated observations from input vector $\mathbf{X}$ and output $Y$. We do not have access to the marginal nor the joint distribution of the input vector, indeed, the input factors are themselves output of different models.

To allow for a comparison between the cascade sensitivities to different inputs, we require a consistent way of stressing the 72 input factors. To this end, each input factor $X_j$ is transformed to a standard normal random variable $Z_j$, comonotonic to $X_j$, through

$$X_j = F_j^{-1}(\Phi(Z_j)),\ j = 1, \dots, 72. \tag{4.9}$$

Note that equation (4.9) allows to stress the input factors through distorting the standard normal random variables $Z_j,\ j = 1, \dots, 72,$. Applying an additive shock $Z_{j,\varepsilon} = Z_j(1+\varepsilon)$ to the standard normal random variables $Z_j$, for all $j = 1, \dots, 72$, stresses the input factors in a consistent fashion. The cascade sensitivity to input $X_j$, with shock $X_{j,\varepsilon} = F_j^{-1}(\Phi(Z_{j,\varepsilon}))$, is then given by Proposition 4.4.3 with

$$\tilde{Y} = g\Big(\boldsymbol{\psi}\Big(F_j^{-1}(\Phi(\tilde{Z}_j)\Big), \boldsymbol{V}\Big),\ \text{for } (\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_j,$$

where $\tilde{Z}$ follows distribution function $\tilde{\Phi}(x) = \Phi(x) - x\Phi'(x),\ x \in \mathbb{R}$. The distorted output distribution function $\tilde{H}$ of $\tilde{Y}$, and thus the cascade sensitivities, can easily be calculated via importance sampling, by the algorithm in Section 4.4.3. In particular it holds that

$$\tilde{H}(y) = E\left(\mathbb{1}_{\{Y \le y\}} \frac{\tilde{\Phi}'(Z_i)}{\Phi'(Z_i)}\right).$$

Recall that the algorithm in Section 4.4.3 does not require the explicit knowledge of the distribution of the input factors, the inverse Rosenblatt transform of $\boldsymbol{X}$ nor the

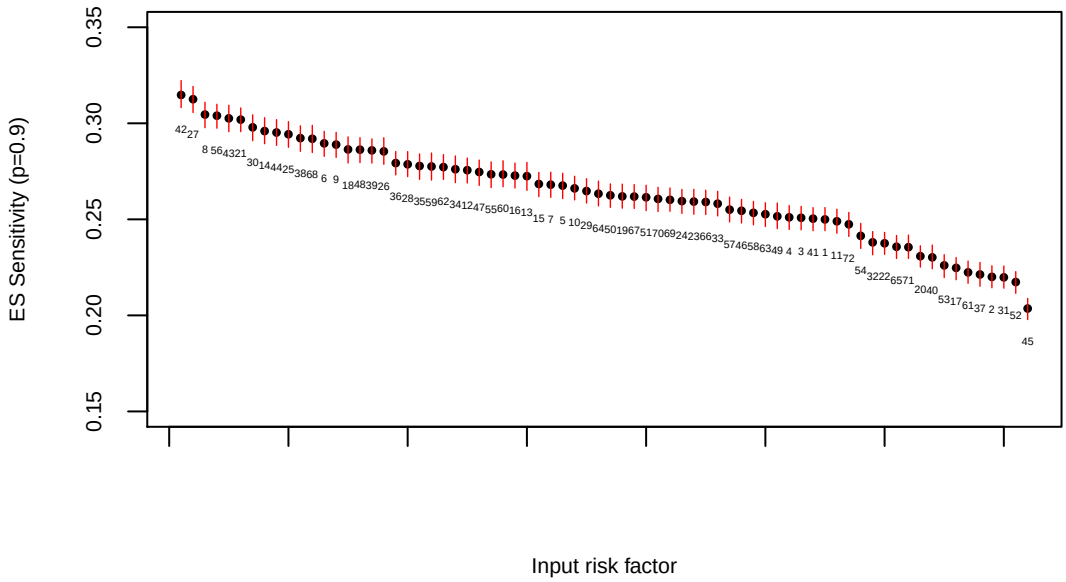


Figure 4.6: Cascade sensitivity for the $ES_{0.9}$ of the London Market portfolio, including 90% confidence intervals.

derivative of the aggregation function, hence making the cascade sensitivity framework applicable to a black box model setting. Figure 4.6 displays the cascade sensitivities for the 72 input factors for the risk measure $ES_{0.9}$ including 90% confidence intervals, calculated using bootstrap based on 500 simulations.

## 4.7 Conclusion

We propose a novel sensitivity measure, defined as a directional derivative of a distortion risk measure applied to the model output in direction to a stressed input factor. The cascade sensitivity captures not only the direct impact of the stressed input factor on the output but also the indirect effects arising via dependence with other input

factors. Through examples, we show that the dependence between input factors may substantially contribute to the cascade sensitivity of a particular input.

We provide representations of the cascade sensitivity that allow for a straightforward calculation on one Monte Carlo sample and without the knowledge of the gradient of the aggregation function. In particular, we consider stresses consisting of either an additive shock to an input factor or a perturbation, such that the stressed input factor follows a mixture distribution. These representations of the cascade sensitivity make implementation of the proposed sensitivity measure numerically efficient and, thus, attractive for applications in practice.

# Appendix 4.A Assumptions and proofs

**Assumption 4.A.1.** Let $\rho_\gamma$ be a distortion risk measure, $0 < q < 1$, and $X_{i,\varepsilon},\ \varepsilon \geq 0$, a stress. With abuse of notation, we denote the stressed input vector by $\boldsymbol{X}_{i,\varepsilon} = (X_1, \ldots, X_{i,\varepsilon}, \ldots, X_n)$ for the marginal sensitivity and $\boldsymbol{X}_{i,\varepsilon} = \boldsymbol{\psi}(X_{i,\varepsilon}, \boldsymbol{V}),\ (\boldsymbol{\psi}, \boldsymbol{V}) \in \mathcal{R}_i$, for the cascade sensitivity, respectively. We write $Y_{i,\varepsilon} = g(\boldsymbol{X}_{i,\varepsilon})$ for the stressed output and denote its distribution function by $H_{i,\varepsilon}$.

i) There exists a random variable $W$ with $E(W) < +\infty$, such that $|Y_{i,\varepsilon_2} - Y_{i,\varepsilon_1}| \leq W|\varepsilon_2 - \varepsilon_1|$ for all $\varepsilon_1, \varepsilon_2 > 0$ in a neighbourhood of 0.

ii) The derivative with respect to $\varepsilon$ of $Y_{i,\varepsilon}$ exists with probability 1 for all $\varepsilon > 0$ in a neighbourhood of 0.

iii) For all $\varepsilon > 0$ in a neighbourhood of 0, $Y_{i,\varepsilon}$ has a continuous density in a neighbourhood of $H^{-1}(q)$. The derivative of the distribution function with respect to $\varepsilon$ exists and is continuous (in both arguments) at $H^{-1}(q)$.

iv) For all $\varepsilon > 0$ in a neighbourhood of 0, the function $E\Big(\frac{\partial}{\partial\varepsilon} Y_{i,\varepsilon} \,|\, Y_{i,\varepsilon} = y\Big)$ is continuous at $y = H^{-1}(q)$.

v) It holds $P\Big(Y_{i,\varepsilon} = H_{i,\varepsilon}^{-1}(q)\Big) = 0$ for all $\varepsilon > 0$ in a neighbourhood of 0.

vi) The function $H_{i,\varepsilon}^{-1}(q)$ is differentiable for all $\varepsilon > 0$ in a neighbourhood of 0.

vii) The aggregation function $g$ is invertible in at least one argument, say the $j^{\text{th}}$, and $X_j$ has a conditional density given $\boldsymbol{X}_{-j}$.

viii) $P(H_{i,\varepsilon}(Y_{i,\varepsilon}) \in D_\gamma) = 1$, where $D_\gamma$ is the set where the weight function $\gamma$ is differentiable.

ix) $\frac{\partial}{\partial\varepsilon} H_{i,\varepsilon}^{-1}$ is bounded for all $\varepsilon > 0$ in a neighbourhood of 0.

**Lemma 4.A.2.** Let $Z$ be an integrable random variable with distribution function $F_Z$ and right-continuous density $f_Z$ whose support can be split into countably many intervals on which $f_Z$ is monotonic. Let $k\colon \mathbb{R} \to \mathbb{R}$ a non-decreasing Lipschitz continuous function with Lipschitz constant 1, that satisfies $k(x) \leq 0$ on the set where $f_Z(x)$ is non-decreasing and $k(x) \geq 0$ on the set where $f_Z(x)$ is non-increasing. Then $\tilde{F}_Z(x) = F_Z(x) - k(x)f_Z(x),\ x \in \mathbb{R}$, defines a distribution function.

*Proof.* By Lipschitz continuity of $k$, it holds $-(k(y) - k(x)) \geq -(y - x)$, for all $x < y$. Let $a, b \in \mathbb{R}$ such that $f_Z$ is non-decreasing on $[a, b]$, then it holds for all $a \leq x < y \leq b$ that

$$\begin{aligned}
\tilde{F}_Z(y) - \tilde{F}_Z(x) &= \int_x^y f_Z(u)\mathrm{d}u - k(y)f_Z(y) - (k(y) - k(x))f_Z(x) + k(y)f_Z(x) \\
&\geq \int_x^y \Big(f_Z(u) - f_Z(x)\Big)\mathrm{d}u - k(y)\Big(f_Z(y) - f_Z(x)\Big) \\
&\geq 0.
\end{aligned}$$

Similarly, for $a, b \in \mathbb{R}$ such that $f_Z$ is non-increasing on $[a, b]$, we have for all $a \leq x < y \leq b$

$$\begin{aligned}\tilde{F}_Z(y) - \tilde{F}_Z(x) &= \int_x^y f_Z(u)\mathrm{d}u - (k(y) - k(x))f_Z(y) - k(x)f_Z(y) + k(x)f_Z(x) \\ &\geq \int_x^y \Big(f_Z(u) - f_Z(y)\Big)\mathrm{d}u + k(x)\Big(f_Z(x) - f_Z(y)\Big) \\ &\geq 0.\end{aligned}$$

For $x < y$, for which $f_Z$ is not monotone on $[x, y]$, we define the partition $x = m_1 \leq \cdots \leq m_r = y$ such that $f_Z$ is monotone on $[m_j, m_{j+1}]$, for all $j = 1, \ldots, r-1$. Then, we can write $\tilde{F}_Z(y) - \tilde{F}_Z(x) = \sum_{j=1}^{r-1} \tilde{F}_Z(m_{j+1}) - \tilde{F}_Z(m_j)$, and thus $\tilde{F}_Z$ is indeed a distribution function. □

**Lemma 4.A.3.** Let $K\colon \mathbb{R} \to \mathbb{R}$ be an absolutely function with representation $K(x) = \int_b^x \kappa(s)\mathrm{d}s$ for all $x \geq b$, where $\kappa$ is a non-negative function and $b \geq -\infty$. Then, for any random variable $Z \geq b$ a.s. with $E(K(Z)) < \infty$ it holds that

$$E(K(Z)) = \int_b^{+\infty} \kappa(s)P(Z > s)\mathrm{d}s.$$

*Proof.* For any $b \geq -\infty$ we obtain using Fubini

$$\begin{aligned}E(K(Z)) &= E\Big(\int_b^Z \kappa(s)\mathrm{d}s\Big) \\ &= E\Big(\int_b^{+\infty} \kappa(s)\mathbb{1}_{\{Z \geq b\}}\mathrm{d}s\Big) \\ &= \int_{-\infty}^{+\infty} \kappa(s)P(Z > s)\mathrm{d}s.\end{aligned}$$

□

# References


Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009), 'Pair-copula constructions of multiple dependence', *Insurance: Mathematics and Economics* **44**(2), 182–198.

Acerbi, C. (2002), 'Spectral measures of risk: A coherent representation of subjective risk aversion', *Journal of Banking & Finance* **26**(7), 1505–1518.

Acerbi, C. and Tasche, D. (2002), 'Expected Shortfall: A natural coherent alternative to Value at Risk', *Economic Notes* **31**(2), 379–388.

Ahmed, S., Çakmak, U. and Shapiro, A. (2007), 'Coherent risk measures in inventory problems', *European Journal of Operational Research* **182**(1), 226–238.

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

Aliprantis, C. D. and Border, K. (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer Science & Business Media.

Antoniano-Villalobos, I., Borgonovo, E. and Siriwardena, S. (2018), 'Which parameters are important? Differential importance under uncertainty', *Risk Analysis* . (forthcoming).

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, E. (1997), 'Thinking coherently', *RISK* **10**, 68–71.

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

Asimit, A. V., Bignozzi, V., Cheung, K. C., Hu, J. and Kim, E.-S. (2017), 'Robust and pareto optimality of insurance contracts', *European Journal of Operational Research* **262**(2), 720–732.

Aven, T. and Nøkland, T. (2010), 'On the use of uncertainty importance measures in reliability and risk analysis', *Reliability Engineering & System Safety* **95**(2), 127–133.

Bäuerle, N. and Müller, A. (2006), 'Stochastic orders and risk measures: Consistency and bounds', *Insurance: Mathematics and Economics* **38**(1), 132–148.

BCBS (2012), Consultative Document May 2012. Fundamental review of the trading book, Technical report, Basel Committee on Banking Supervision, Bank for International Settlements.

BCBS (2013), Consultative Document October 2013. Fundamental review of the trading book: A revised market risk framework, Technical report, Basel Committee on Banking Supervision, Bank for International Settlements.

Beckman, R. J. and McKay, M. D. (1987), 'Monte Carlo estimation under different distributions using the same simulation', *Technometrics* **29**(2), 153–160.

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014), 'Generalized quantiles as risk measures', *Insurance: Mathematics and Economics* **54**, 41–48.

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

Bernard, C., Jiang, X. and Wang, R. (2014), 'Risk aggregation with dependence uncertainty', *Insurance: Mathematics and Economics* **54**, 93–108.

Bickel, P. J. and Freedman, D. A. (1981), 'Some asymptotic theory for the bootstrap', *The Annals of Statistics* **9**(6), 1196–1217.

Billingsley, P. (2008), *Probability and Measure*, third edn, John Wiley and Sons.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989), *Regular Variation*, Cambridge University Press.

Blanchet, J., Lam, H., Tang, Q. and Yuan, Z. (2017), Applied robust performance analysis for actuarial applications, Technical report, The Society of Actuaries (SOA).

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

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

Borgonovo, E. and Cillo, A. (2017), 'Deciding with thresholds: Importance measures and value of information', *Risk Analysis* **37**(10), 1828–1848.

Borgonovo, E., Hazen, G. B. and Plischke, E. (2016), 'A common rationale for global sensitivity measures and their estimation', *Risk Analysis* **36**(10), 1871–1895.

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

Boyd, S. and Vandenberghe, L. (2004), *Convex optimization*, Cambridge University Press.

Breuer, T. and Csiszár, I. (2013), 'Systematic stress tests with entropic plausibility constraints', *Journal of Banking & Finance* **37**(5), 1552–1559.

Breuer, T., Jandačka, M., Mencía, J. and Summer, M. (2012), 'A systematic approach to multi-period stress testing of portfolio credit risk', *Journal of Banking & Finance* **36**(2), 332–340.

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

Cao, X.-R. and Wan, X. (2017), 'Sensitivity analysis of nonlinear behavior with distorted probability', *Mathematical Finance* **27**(1), 115–150.

Cheridito, P. and Li, T. (2009), 'Risk measures on Orlicz hearts', *Mathematical Finance* **19**(2), 189–214.

Cont, R., Deguest, R. and Scandolo, G. (2010), 'Robustness and sensitivity analysis of risk measurement procedures', *Quantitative Finance* **10**(6), 593–606.

Cooke, R. M. and van Noortwijk, J. M. (1999), 'Local probabilistic sensitivity measures for comparing FORM and Monte Carlo calculations illustrated with dike ring reliability calculations', *Computer Physics Communications* **117**(1-2), 86–98.

Cover, T. M. and Thomas, J. A. (2012), *Elements of information theory*, John Wiley & Sons.

Csiszár, I. (1975), 'I-divergence geometry of probability distributions and minimization problems', *The Annals of Probability* **3**(1), 146–158.

Dana, R.-A. and Scarsini, M. (2007), 'Optimal risk sharing with background risk', *Journal of Economic Theory* **133**(1), 152–176.

Delbaen, F. (2002), Coherent risk measures on general probability spaces, *in* 'Advances in Finance and Stochastics', Springer, pp. 1–37.

Denneberg, D. (1990), 'Premium calculation: Why standard deviation should be replaced by absolute deviation', *ASTIN Bulletin: The Journal of the IAA* **20**(2), 181–190.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2006), *Actuarial theory for dependent risks: Measures, orders and models*, John Wiley & Sons.

Diestel, J. (1991), 'Uniform integrability: An introduction'.

Dobrushin, R. (1970), 'Prescribing a system of random variables by conditional distributions', *Theory of Probability & its Applications* **15**(3), 458–486.

Dudley, R. M. (2002), *Real Analysis and Probability*, Cambridge University Press.

Durrett, R. (2013), *Probability: Theory and Examples*, re-edited edn, Cambridge University Press.

EIOPA (2009), Directive 2009/138/ec of the European Parliament and of the Council, Technical report, European Insurance and Occupational Pensions Authority.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (2013), *Modelling Extremal Events for Insurance and Finance*, Springer Science & Business Media. 10th printing.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013), 'Model uncertainty and var aggregation', *Journal of Banking & Finance* **37**(8), 2750–2764.

Embrechts, P., Wang, B. and Wang, R. (2015), 'Aggregation-robustness and model uncertainty of regulatory risk measures', *Finance and stochastics* **19**(4), 763–790.

Esary, J. D., Proschan, F. and Walkup, D. W. (1967), 'Association of random variables, with applications', *The Annals of Mathematical Statistics* **38**(5), 1466–1474.

Feinstein, Z., Rudloff, B. and Weber, S. (2017), 'Measures of systemic risk', *SIAM Journal on Financial Mathematics* **8**(1), 672–708.

Filipović, D. and Svindland, G. (2012), 'The canonical model space for law-invariant convex risk measures is $L^1$', *Mathematical Finance* **22**(3), 585–589.

Föllmer, H. and Schied, A. (2002), 'Convex measures of risk and trading constraints', *Finance and stochastics* **6**(4), 429–447.

Föllmer, H. and Schied, A. (2011), *Stochastic Finance: An Introduction in Discrete Time*, 3rd edn, Walter de Gruyte.

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

Frittelli, M. and Rosazza Gianin, E. (2002), 'Putting order in risk measures', *Journal of Banking & Finance* **26**(7), 1473–1486.

Gerber, H. U. (1974), 'On additive premium calculation principles', *ASTIN Bulletin: The Journal of the IAA* **7**(3), 215–222.

Gibbs, A. L. and Su, F. E. (2002), 'On choosing and bounding probability metrics', *International Statistical Review* **70**(3), 419–435.

Givens, C. R. and Rae, S. M. (1984), 'A class of Wasserstein metrics for probability distributions', *The Michigan Mathematical Journal* **31**(2), 231–240.

Glasserman, P. (1991), *Gradient estimation via perturbation analysis*, Vol. 116, Springer Science & Business Media.

Glasserman, P., Kang, C. and Kang, W. (2015), 'Stress scenario selection by empirical likelihood', *Quantitative Finance* **15**(1), 25–41.

Glasserman, P. and Liu, Z. (2010), 'Sensitivity estimates from characteristic functions', *Operations Research* **58**(6), 1611–1623.

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

Gotoh, J.-Y. and Takano, Y. (2007), 'Newsvendor solutions via conditional value-at-risk minimization', *European Journal of Operational Research* **179**(1), 80–96.

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

Gourieroux, C., Liu, W. et al. (2006), Sensitivity analysis of distortion risk measures, Technical report.

Gustafson, P. et al. (1996), 'Local sensitivity of posterior expectations', *The Annals of Statistics* **24**(1), 174–195.

Haezendonck, J. and Goovaerts, M. (1982), 'A new premium calculation principle based on Orlicz norms', *Insurance: Mathematics and Economics* **1**(1), 41–53.

Hampel, F. R. (1971), 'A general qualitative definition of robustness', *The Annals of Mathematical Statistics* **42**(6), 1887–1896.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (2011), *Robust statistics: The approach based on influence functions*, Vol. 196, John Wiley & Sons.

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

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2017), *Copula: Multivariate Dependence with Copulas*. R package version 0.999-18.
**URL:** *https://CRAN.R-project.org/package=copula*

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

Hong, L. J. and Liu, G. (2009), 'Simulating sensitivities of conditional value at risk', *Management Science* **55**(2), 281–293.

Huber, P. and Ronchetti, E. M. (2009), *Robust Statistics*, 2nd edn, John Wiley & Sons.

IAIS (2014), Consultation Document December 2014. Risk-based global insurance capital standard, Technical report, International Association of Insurance Supervisors.

Jarrow, R. A. and Turnbull, S. M. (2000), *Derivative Securities*, 2 edn, South-Western College Publishing.

Kachapova, F. and Kachapov, I. (2012), 'A measure of monotonicity of two random variables', *Journal of Mathematics and Statistics* **8**(2), 221–228.

Kaina, M. and Rüschendorf, L. (2009), 'On convex risk measures on $L^p$-spaces', *Mathematical methods of operations research* **69**(3), 475–495.

Karlin, S. and Rinott, Y. (1980), 'Classes of orderings of measures and related correlation inequalities. I. multivariate totally positive distributions', *Journal of Multivariate Analysis* **10**(4), 467–498.

Kiesel, R., Rühlicke, R., Stahl, G. and Zheng, J. (2016), 'The Wasserstein metric and robustness in risk management', *Risks* **4**(3), 32.

Krätschmer, V., Schied, A. and Zähle, H. (2014), 'Comparative and qualitative robustness for law-invariant risk measures', *Finance and stochastics* **18**(2), 271–295.

Krätschmer, V., Schied, A. and Zähle, H. (2017), 'Domains of weak continuity of statistical functionals with a view toward robust statistics', *Journal of Multivariate Analysis* **158**, 1–19.

Kullback, S. (1959), *Information theory and statistics*, John Wiley & Sons.

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

Kusuoka, S. (2001), On law invariant coherent risk measures, *in* 'Advances in mathematical economics', Springer, pp. 83–95.

Lehmann, E. L. (1966), 'Some concepts of dependence', *The Annals of Mathematical Statistics* **37**(5), 1137–1153.

Lloyd's (2016), Internal model validation: Supplementary notes to the Solvency II model validation guidance (April 2014), Technical report, Lloyd's.

Ludkovski, M. and Risk, J. (2017), Sequential design and spatial modeling for portfolio tail risk measurement.
**URL:** *https://arxiv.org/abs/1710.05204*

Mai, J.-F., Schenk, S. and Scherer, M. (2015), 'Analyzing model robustness via a distortion of the stochastic root: A dirichlet prior approach', *Statistics & Risk Modeling* **32**(3-4), 177–195.

Makarov, G. D. (1982), 'Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed', *Theory of Probability & its Applications* **26**(4), 803–806.

Mao, T. and Wang, R. (2015), 'On aggregation sets and lower-convex sets', *Journal of Multivariate Analysis* **138**, 170–181.

Mara, T. A. and Tarantola, S. (2012), 'Variance-based sensitivity indices for models with dependent inputs', *Reliability Engineering & System Safety* **107**, 115–121.

McNeil, A. J. and Smith, A. D. (2012), 'Multivariate stress scenarios and solvency', *Insurance: Mathematics and Economics* **50**(3), 299–308.

Müller, A. and Scarsini, M. (2001), 'Stochastic comparison of random vectors with a common copula', *Mathematics of Operations Research* **26**(4), 723–740.

Müller, A. and Stoyan, D. (2002), *Comparison methods for stochastic models and risks*, John Wiley & Sons.

Nelder, J. A. and Wedderburn, R. W. M. (1973), 'Generalized linear models', *Journal of the Royal Statistical Society. Series A (General)* **135**(3), 370–384.

Newey, W. K. and Powell, J. L. (1987), 'Asymmetric least squares estimation and testing', *Econometrica: Journal of the Econometric Society* pp. 819–847.

Oakley, J. E. and O'Hagan, A. (2004), 'Probabilistic sensitivity analysis of complex models: A Bayesian approach', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(3), 751–769.

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

Pesenti, S. M., Millossovich, P. and Tsanakas, A. (2016), 'Robustness regions for measures of risk aggregation', *Dependence Modeling* **4**(1), 348–367.

Pesenti, S. M., Millossovich, P. and Tsanakas, A. (2018*a*), Cascade sensitivity measures. Working paper.
**URL:** *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3270839*

Pesenti, S. M., Millossovich, P. and Tsanakas, A. (2018*b*), 'Robustness regions for measures of risk aggregation', *European Journal of Operational Research* **274**(2), 654–670.

Plischke, E., Borgonovo, E. and Smith, C. L. (2013), 'Global sensitivity measures from given data', *European Journal of Operational Research* **226**(3), 536–550.

Rockafellar, R. T. and Uryasev, S. (2002), 'Conditional value-at-risk for general loss distributions', *Journal of Banking & Finance* **26**(7), 1443–1471.

Rosenblatt, M. (1952), 'Remarks on a multivariate transformation', *The Annals of Mathematical Statistics* **23**(3), 470–472.

Rubinstein, R. Y. and Melamed, B. (1998), *Modern simulation and modeling*, Vol. 7, John Wiley & Sons.

Rüschendorf, L. (1982), 'Random variables with maximum sums', *Advances in Applied Probability* **14**(3), 623–632.

Rüschendorf, L. (1983), 'Solution of a statistical optimization problem by rearrangement methods', *Metrika* **30**(1), 55–61.

Rüschendorf, L. and de Valk, V. (1993), 'On regression representations of stochastic processes', *Stochastic Processes and their Applications* **46**(2), 183–198.

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

Saltelli, A., Chan, K. and Scott, E. M. (2000), *Sensitivity analysis*, John Wiley & Sons.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008), *Global sensitivity analysis: The primer*, John Wiley & Sons.

Saltelli, A. and Tarantola, S. (2002), 'On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal', *Journal of the American Statistical Association* **97**(459), 702–709.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004), *Sensitivity analysis in practice: A guide to assessing scientific models*, John Wiley & Sons.

Samuelson, P. A. (1941), 'The stability of equilibrium: comparative statics and dynamics', *Econometrica: Journal of the Econometric Society* pp. 97–120.

Sandström, A. (2016), *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*, CRC Press.

Schepsmeier, U. (2015), 'Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review', *Journal of Multivariate Analysis* **138**, 34–52.

SCOR (2008), From principle-based risk management to solvency requirements, 2nd edition, Technical report, SCOR Switzerland AG.

Shapiro, A., Dentcheva, D. and Ruszczyński, A. (2009), *Lectures on Stochastic Programming: Modeling and Theory*, 2nd edn, SIAM.

Shorack, G. R. and Wellner, J. A. (2009), *Empirical Processes with Application to Statistics*, SIAM.

Sobol, I. M. (1993), 'Sensitivity estimates for nonlinear mathematical models', *Mathematical Modelling and Computational Experiments* **1**(4), 407–414.

Spear, R. C., Grieb, T. M. and Shang, N. (1994), 'Parameter uncertainty and interaction in complex environmental models', *Water Resources Research* **30**(11), 3159–3169.

Strassen, V. (1965), 'The existence of probability measures with given marginals', *The Annals of Mathematical Statistics* **36**(2), 423–439.

Swiss Solvency Test (2006), FINMA SST, Technical report.

Tapiero, C. S. (2005), 'Value at risk and inventory control', *European Journal of Operational Research* **163**(3), 769–775.

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

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

Vallender, S. (1974), 'Calculation of the Wasserstein distance between probability distributions on the line', *Theory of Probability & its Applications* **18**(4), 784–786.

Wagner, H. M. (1995), 'Global sensitivity analysis', *Operations Research* **43**(6), 948–969.

Wakker, P. (1994), 'Separating marginal utility and probabilistic risk aversion', *Theory and Decision* **36**(1), 1–44.

Wang, B. and Wang, R. (2011), 'The complete mixability and convex minimization problems with monotone marginal densities', *Journal of Multivariate Analysis* **102**(10), 1344–1360.

Wang, B. and Wang, R. (2016), 'Joint mixability', *Mathematics of Operations Research* **41**(3), 808–826.

Wang, R., Peng, L. and Yang, J. (2013), 'Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities', *Finance and stochastics* **17**(2), 395–417.

Wang, S. (1996), 'Premium calculation by transforming the layer premium density', *ASTIN Bulletin: The Journal of the IAA* **26**(I), 71–92.

Wang, X., Liu, Q., Hou, Y. and Peng, L. (2018), 'Nonparametric inference for sensitivity of haezendonck–goovaerts risk measure', *Scandinavian Actuarial Journal* pp. 1–20.

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

Wüthrich, M. V. (2016), 'Non-Life Insurance: Mathematics & Statistics', *Available at SSRN: 2319328* .

Wüthrich, M. V. and Merz, M. (2013), *Financial modeling, actuarial valuation and solvency in insurance*, Springer.

Xu, C. and Gertner, G. Z. (2008*a*), 'A general first-order global sensitivity analysis method', *Reliability Engineering & System Safety* **93**(7), 1060–1071.

Xu, C. and Gertner, G. Z. (2008*b*), 'Uncertainty and sensitivity analysis for models with correlated parameters', *Reliability Engineering & System Safety* **93**(10), 1563–1573.